MECHANICAL ENGINEERING / PHYSICS / PRESERVATION OF THE ARCHITECTURAL HERITAGE / STRUCTURAL, SEISMIC AND GEOTECHNICAL ENGINEERING / URBAN PLANNING, DESIGN AND POLICY / AEROSPACE ENGINEERING / ARCHITECTURE, BUILT ENVIRONMENT AND CONSTRUCTION ENGINEERING / ARCHITECTURAL, URBAN AND INTERIOR DESIGN / BIOENGINEERING /DATA ANALYTICS AND DECISION SCIENCES / DESIGN / ELECTRICAL ENGINEERING / ENERGY AND NUCLEAR SCIENCE AND TECHNOLOGY / ENVIRONMENTAL AND INFRASTRUCTURE ENGINEERING / INDUSTRIAL CHEMISTRY AND CHEMICAL ENGINEERING / **INFORMATION TECHNOLOGY** / MANAGEMENT ENGINEERING / MATERIALS ENGINEERING / MATHEMATICAL MODELS AND METHODS IN ENGINEERING

PhD Yearbook | 2020

DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY

Chair:
Prof. Barbara Pernici

Introduction

The PhD program in Information Technology (IT) covers research topics in four scientific areas: Computer Science and Engineering, Electronics, Systems and Control, and Telecommunications. This broad variety of research topics is matched together by the common affinity to the ICT area, and perfectly captures the core mission of the Department of Electronics, Information and Bioengineering (DEIB) where it is based. In addition to the basic research activities in the department in the four scientific areas mentioned above, new trends of modern society and cross-areas research fields are also covered, such as robotics, machine learning, big data, intelligent data analysis, Industry 4.0, Internet of Things, ecology, environmental modelling, operations research, and transportation systems. The PhD program in IT is the largest in Politecnico in terms of number of students. Every year, about 70 new students join the program, for an overall number of students of about 200. Students must undergo a yearly evaluation of the progress in their research and course work.

Topics

Research at DEIB in the field of Information Technology is supported by 35 laboratories, and is organized in 4 main areas. *Computer Science and Engineering* (Vice-Chair: Prof. Cristina Silvano): Information systems, Database management, Information design for the web, Methods and applications for interactive multimedia, Embedded systems design and design methodologies, Dependable systems, 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-Chair: Prof. Angelo Geraci): Circuits and systems, Singlephoton detectors and applications, Radiation detectors and low noise electronics, Electronic circuit design, Electron devices. *Systems and Control* (Vice-Chair: Prof. Luigi Piroddi): Control systems, Robotics and industrial automation, Optical measurements and laser instrumentation, Dynamics of complex systems, Planning and management of environmental systems, Operations research and discrete optimization. *Telecommunications* (Vice-Chair: Prof. Matteo Cesana): 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 program features many industrial collaborations. About 50% of scholarships are funded by companies or by international research projects involving industrial partners. In the PhD school vision, the collaboration between university and companies is the ideal ground where to turn invention and scientific research into technological innovation. This collaboration also contributes to create a common terrain of friendly culture, to size research risk, and to preserve strong basic research. To monitor the activities and development of the PhD program, the PhD board cooperates with an industrial advisory board, composed by members of public and private companies, working in management, production, and applied research. The board meets once a year to identify and suggest new emerging research areas and to foster the visibility of the PhD program in the industrial world.

Educational aspects

The teaching organization and the course subjects reflect the scientific interests of DEIB faculties. The curricula include a wide choice of courses (about 20 per year), and more than 30 courses for basic soft and traversal skills offered by the PhD School of Politecnico di Milano. 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 students to make their best choices according to an internal regulation scheme.

Internationalization

Every year, several courses are delivered by visiting professors from prestigious foreign universities. Moreover, the PhD program encourages joint curricula with foreign institutions. Every year we receive many applications from foreign countries and about 20% of our selected PhD candidates have applied from outside Italy.

Conclusions

The core mission of our PhD Program is to offer an excellent PhD curriculum, through high quality courses, a truly interdisciplinary advanced education, cutting-edge research, international and industrial collaborations.

PHD BOARD OF PROFESSORS

BARBARA PERNICI	LORENZO FAGIANO	LUIGI PIRODDI
CESARE ALIPPI	GIANCARLO FERRIGNO	MATTEO PRADELLA
FRANCESCO AMIGONI	GIORGIO FERRARI	IVAN RECH
LUCA BASCETTA	SIMONE GARATTI	CARLO RIVA
GIUSEPPE BERTUCCIO	NICOLA GATTI	CRISTINA SILVANO
ANDREA CASTELLETTI	ANGELO GERACI	ALESSANDRO SOTTOCORNOLA SPINELLI
MATTEO CESANA	PAOLO MARTELLI	LETIZIA TANCA
FLORIAN DANIEL	RAFFAELA MIRANDOLA	MASSIMO TORNATORE
FABIO DERCOLE	ANDREA MONTI GUARNIERI	

PHD ADVISORY BOARD

Giorgio Ancona	Atos
Matteo Bogana	Cleafy
Mario Caironi	IIT
Paolo Cederle	Everis
Luigi Cicchese	Concept Reply
Cristina Cremonesi	The European Ambrosetti
Riccardo De Gaudenzi	European Space Agency
Giuseppe Desoli	STMicroelectronics
Alessandro Ferretti	Tre-Altamira
Giuseppe Fogliazza	MCE Srl
Bruno Garavelli	Xnext s.r.l.
Sabino Illuzzi	Prospera
Renato Lombardi	Huawei Technologies
Renato Marchi	KPMG
David Oliva Uribe	Head Industrial Doctoral School, EIT Digital
Giorgio Parladori	SM Optics srl
Francesco Prelz	INFN
Enrico Ragaini	ABB S.p.A.
Paolo Giuseppe Ravazzani	CNR
Fabrizio Renzi	IBM
Carlo Sandroni	RSE S.p.A.
Massimo Valla	TIM
Luisa Venturini	Vodafone Italy
Stefano Verzura	Huawei Technologies
Roberto Villa	IBM Italy

Prizes and awards

In 2019 the following awards have been obtained by PhD candidates:

Prof. Emilio Gatti Thesis Award issued by “Istituto Lombardo Accademia di Scienze e Lettere” - Idham Hafizh
Young Academics Award of the Alpine Convention - 1st prize - Marta Zaniolo
ECIO 2019 Best Student Poster Award - Maziyar Milanizadeh
ACM International Conference on Computing Frontiers Best Poster Award 2019 - Stefano Cherubin
2019 Best Master Thesis Award “Il Quadrato della Radio” - Mattia Brambilla
2019 IEEE Nuclear Science Symposium Best Student Paper Award - Luca Buonanno
IFAC Young Author Award at the 15th IFAC LSS 2019 - Alessio La Bella
IEEE ISCAS 2019 Best Paper Award - Luca Buonanno, Giovanni Ludovico Montagnani
BalkanCom 2019 Best Student Paper Award and BalkanCom 2019 Best Paper Award - Francesco Linsalata
2019 CAiSE Ph.D. Award - Giovanni Meroni
Intellectual Property Award for “ICT, Intelligenza Artificiale, IoT, Big Data, Logistica, Costruzioni” - Giacomo Pedretti
ACM Europe Council Best Paper Award - Maurizio Ferrari Dacrema
Chorafas Award - Alberto Bernardini, Dmytro Cherniak
Amazon Machine Learning Research Awards - Luca Nanni
RuCTF 2019 - Marcello Pogliani

REAL TIME TRADING OF MOBILE RESOURCES IN BEYOND-5G SYSTEMS

Özgür Umut Akgül - Supervisor: Prof. Antonio Capone

The evolution toward 5G and beyond networks brings out novel techno-economic challenges. First and most important of these problems is providing the quality of service expectations with an economically sustainable model. The exponentially increasing broadband demand along with the challenging throughput and delay constraints requires pivotal changes in the network infrastructure. However, the total cost of network upgrade further tightens an already condensed profit margin of the mobile operators and turns the network provisioning into a nonprofitable business model. In order to decrease the total costs, a possible way is to extend the infrastructure sharing agreements to include active network components as well as the spectral resources. However, as the number of shared components in the wireless network increases, the individual operators start taking higher business risks. Moreover, in order for such a static sharing agreement to be effective, the operators have to have a very good estimation of their current resource needs as well as the evolution of their needs in a relatively long period of time (e.g. a year). Therefore, the static approaches nearly always end in over-provisioning of the valuable spectral resources.
Dynamic infrastructure sharing can increase the resource efficiency as well as further decrease the costs. In order to achieve cost reduction without loss of business potential, all parties in the sharing agreement must be able to renegotiate their resource shares within very short time windows (e.g. hours or days).
Another big challenge in 5G and beyond networks is the heterogeneity of the traffic requirements. Unlike predecessor technologies, 5G is envisioned to contain a multitude of industry driven services with specific requirements and priorities. The conventional method to handle these requirements, i.e. optimizing the network resources in line with the quality of service expectations, is no longer an option due to the diverse performance indicators. A way to handle this heterogeneity is to vertically group the network resources (i.e. slicing the network) and optimize the separate groups (i.e. slices) in line with the requirements. Similar to aforementioned problem with the infrastructure sharing, static network slicing requires the mobile operators to have a very good estimation on their needs. However, being an industry driven wireless technology, 5G requires a great level of flexibility in resource allocations and network configurations in order to adapt to the evolution of the traffic conditions and the service needs. A possible way to provide adaptability to the changing conditions (both in terms of the demand and the service type) is dynamically slicing the network resources based on some slice templates. On the other hand, dynamic network slicing makes the different services to be connected in terms of the resource allocations which can easily lead service level agreement violations.
Moreover, in order to provide a robust business ecosystem, a key attribute is to guarantee resource availability. Therefore, the network capacity has to be scaled in line with the evolving needs of the network traffic. On the other hand, the heterogeneous traffic requirements and priority make it harder to compare the urgency of the capacity expansion need among two different regions. Consequently, the conventional capacity management strategies are required to be revisited in order to fit the changing dynamics in the network provisioning.
In this thesis, we have focused on the aforementioned need for a flexible and efficient network management and proposed a novel dynamic negotiation and resource sharing framework for sliced networks. In order to eliminate the over-restrictive structure of the conventional service level agreements, in our model we propose a novel approach where the operators only define their slice types, their utility expectations and their budgets in order to reach these expectations. The rest of the negotiations are automatically handled using a set of parameters that are introduced by the tenants. The minimization of the human-based negotiations allows us to repeat the inter-tenant renegotiations with a very high frequency and dynamically update the operators' resource shares in line with their objectives and the instantaneous traffic conditions. Since the determination of the resource shares per tenant while simultaneously allocating the resources is quite challenging, we have logically separated our model into two sub-problems, where in one of them we determine the resource allocations based on the predefined resource shares and in the other one, we explore the optimal shares per operator based on the observed traffic conditions in the past (i.e. the traffic mixture and the achievable rate per user). Moreover, in order to extend the network resources in line with the demand, we propose a novel pricing strategy that maps the microeconomics' law of supply and demand. The simulation results show that the proposed model increases the cost efficiency while providing an adaptive network management framework.
Next, we have extended the proposed framework to be able to handle heterogeneous traffic requirements. As a first step we have defined a novel piece-wise linear utility function which can be customized according to the service type. Although the proposed utility function is designed to be scaled according to the achieved rate of services, the delay constraint of services is integrated to the proposed scheduler. Through a large set of Monte Carlo simulations, we show how the tenants can differentiate their services. Moreover, we also explain how to control the interrelatedness between different slices by adapting the slope of the utility functions and how the tenants can exploit this interrelatedness to maximize their spectral efficiency.
Finally, we focus on the anticipatory network slicing and resource sharing. Nowadays, the research on artificial intelligence provides a large variety of prediction tools with reasonable accuracy levels. As a consequence, the researchers can anticipate the variations in the traffic conditions (i.e. both channel conditions and traffic demand), which gives them the unique opportunity to increase the network efficiency and flexibility by using predicted information to guide the resource allocations. Therefore, we have focused on how to exploit predicted data regarding the upcoming channel conditions. However, since these predictions are required to be done within very short time intervals, the accuracy of the prediction data is low. Consequently, we have proposed a novel filtering mechanism that can exploit the accurate predictions and eliminate the effects of inaccurate information. Our results firstly prove that the proposed filtering mechanism can minimize the impact of inaccurate prediction. Moreover, we show that the efficiency of the network sharing can be maximized through anticipating the upcoming channel conditions. Lastly, we have analysed the long-term impacts of anticipatory networking on the evolution of the network infrastructure. Due to the techno-economic constraints, the transition to 5G is envisioned to be distributed over time, making deployments with the rising need. However, the envisioned business ecosystem imposed by multi-tenancy as well as the large variety of traffic needs require a new approach in network capacity planning. Therefore, we proposed a novel slice-aware capacity expansion strategy that can provide efficient deployment decisions. Moreover, our investigations prove that the proposed algorithm does not require long term observations on the network status and can be used in shorter observation windows.

SAFE AND ADAPTIVE SOCIAL ROBOTS FOR CHILDREN WITH AUTISM

Ahmad Yaser Alhaddad

Supervisors: Prof. Andrea Bonarini, Dr. John-John Cabibihan

Autism spectrum disorder (ASD) is a condition that is diagnosed during early childhood and affects the neurodevelopment. Traditionally, intervention sessions for ASD have been carried out through a human therapist. However, there has been a promising potential of using social robots in interventions based on the evidence reported by many individual studies. Most of the studies conducted in robot-mediated interventions, if not all, did not address some of the safety concerns regarding the diffusion of social robots into therapy. The interaction between children with autism and social robots will introduce some safety concerns and issues.

Children with ASD often exhibit aggressive behaviors that cause a lot of troubles to their families and caregivers. The extent of such aggressive behaviors could even affect their peers and the people around them. Children diagnosed with ASD often exhibit stereotypical repetitive and stimming behaviors as a response to a sensory overload. When their body's senses are overwhelmed by too many stimuli in the environment, they might exhibit shut down, loud outbursts or aggressive behaviors in an attempt to escape from unpleasant sensations. This response could intensify to meltdowns, that is the temporary loss of behavioral control. Biting, hitting, throwing objects, and others are physical signs that are usually associated with meltdowns.

Many studies have been conducted on using robots in the therapy sessions with children with ASD. There have been many designs with various shapes, functionalities, sizes, and therapeutic objectives. Most of the existing designs, if not all, were designed initially for research purposes to target a single or a set of objectives; hence, they have been limited in terms of functionality; were not flexible to re-adjust to target different or new objectives; and lacking the feasibility to be deployed to the end-users. When it comes to design features of a robot, there are many important requirements that should be considered, such as safety, adaptability, functionality, and autonomy.

Children with autism often show aggressive behaviors that should be accounted for when designing robots for them. The robot design should be made safe enough not to harm the child and others in case of the occurrence of such behaviors. Few previous designs have paid attention to that aspect, but as a trade-off, the overall functionality of the robot has been limited. Other than the safety of the child and people around them, the robot must be designed to be robust enough to withstand any aggressive behaviors. The overall robot design should be safe and robust at the same time.

Most existing social robots that have been used in therapy were either controlled through a therapist with a hidden controller or through an operator within the same room or at a different room. The adaptability of the robot in that case was limited to the person controlling it. They might have to control every simple set of interactions manually and try their best to select the next appropriate action to be performed by the robot. The robot by itself should at least have some sort of sensing to enable more autonomy and adaptability during the sessions, especially to unwanted and aggressive interactions. The current state of a robot's autonomy that is meant for therapy has not reached a high level of complexity and it is still lagging. The development in robots that sense and respond to unwanted interactions is still needed to achieve safer interactions.

How can small social robots be made safer? In order to answer this question, this research has three main objectives as follows:

1) To quantify the potential harm due to one of the identified potential risky scenarios that might occur between a child and a social robot.
2) To investigate how a robot design can be made safer by investigating some design parameters and their influence on a selected severity index.
3) To investigate the possibility of recognizing and classifying unwanted interactions and to evaluate the influence of the emotional reaction time of a robot's response on the interactions.

The studies presented in this dissertation focus on small robotic form factors. The first study quantifies the potential harm to the head due to one of the identified risky scenarios that might occur between a child and a social robot. The results revealed that the overall harm levels based on the selected severity indices are relatively low compared to their respective thresholds. However, the investigation of harm due to the throwing of a small social robot to the head revealed that it could potentially cause tissue injuries, subconcussive or even concussive events in extreme cases. The second two studies are aimed to make small robots safer by optimizing their design. Hence, studies are conducted investigating how robot design can be made safer by investigating different design factors. The first study investigated the influence of the mass and shape on the linear acceleration of a developed dummy head. The results revealed that the two design factors considered (i.e. mass and shape) affected the resultant response significantly. The second study investigated the influence of the storage modulus and thickness of three different soft materials on the same response. The findings showed that the control factors considered are not statistically significant in attenuating the response. Finally, the last two studies attempt to make small robots more adaptable to promote safer interactions. This is carried out by embedding the recognition of unwanted physical interactions into companion robot with the appropriate timing of responses. The findings of the first study highlight the possibility of characterizing children's negative interactions with robotic toys relying on an accelerometer sensor. The second study showed that producing a late response to an action (i.e. greater than 1.0 s) could negatively affect the children's comprehension of the intended message. The work presented in this dissertation is multidisciplinary that involves the field of Mechanical Engineering and Information Technology.

The three major contributions of this research are related to safety aspects of small robotic form-factor as follows:

1) The identification of potential harm and risks due to the introduction of social robots to children with autism. The risks are intensified due to the challenging behaviors that exist within this population as compared to neurotypical children. The risks are also dependent on the form-factor of the robot involved. Several severity indices were used to quantify the potential harm due to one of the identified risky scenarios.
2) The investigation of hardware approaches to optimize small robotic design to improve safety by reducing the potential harm to the head. The influence of different design parameters on a selected severity index were considered. Two different studies were conducted that considered different parameters. The first study considered the influence of the shape and mass of a robotic design while the second study considered the potential of incorporating soft materials in mitigating the potential harm. The optimal settings for the investigated parameters were identified based on Taguchi method.
3) The embedding of unwanted interactions recognition into small companion robots with the appropriate timing of responses. The first study demonstrated the application of an embedded accelerometer inside a robotic toy to detect and classify a set of interactions based on a machine learning algorithm. The second study investigated the influence of reaction time of the emotional response of a robot to unwanted behavior on the interactions. This study identified the appropriate reaction time window that should deliver the right message to the user.

CHARACTERIZATION OF RESISTIVE SWITCHING DEVICES FOR MEMORY AND COMPUTING

Elia Ambrosi - Supervisor: Prof. Daniele Ielmini

Moore's law of downscaling has driven the evolution of the semiconductor industry for more than fifty years, but is now approaching its end mainly due to the increasing leakage of the complementary metal-oxide-semiconductor (CMOS) transistors. The operating frequency of each transistor has already reached an upper limit set by the maximum acceptable power dissipation, preventing further speed improvement at the device level to avoid an excessive temperature increase of the chip. Furthermore, conventional computing systems are based on the von Neumann architecture, where memory and central processing unit (CPU) are physically separated. This leads to an inevitable bottleneck due to the necessary data movement between the two separated units, which causes significant latency and energy consumption.

In this scenario, emerging non-volatile memory devices offer an unprecedented opportunity to tackle these problems. First of all, they generally offer good scalability beyond Flash and DRAM, and they promise to complement the memory hierarchy by filling the large Flash-DRAM performance gap. Secondly, emerging memories are typically fabricated at low temperature using CMOS-compatible materials. This allows to monolithically integrate the memory on top of CMOS logic in the back-end-of-line of a standard CMOS process, enabling high-bandwidth communication between CPU and non-volatile memory on the same chip. Most interestingly, these devices enable novel computing paradigms, such as in-memory computing and neuromorphic computing, where memory and logic functions are distributed within the same network, naturally overcoming the von Neumann bottleneck. Resistive switching memory (RRAM) is one of the most promising emerging technologies, thanks to its excellent scalability, high writing speed, good endurance, low fabrication cost and good compatibility with the back-end-of-line of the CMOS process. However, RRAM is facing serious reliability issues due to programming variability and resistance fluctuations over time, which are preventing its adoption as a commercial technology.

This Doctoral Dissertation focused on the characterization of RRAM devices for high-density crosspoint arrays, and on their application in novel in-memory computing systems. The work was carried out in the context of ERC Consolidator Grant RESCUE (REsistive-Switch CompUting bEyond CMOS) and FET Proactive HERMES (Hybrid Enhanced Regenerative Medicine Systems). First, a novel RRAM technology based on silicon oxide (SiO_x) was demonstrated. The non-volatile device, based on the Ti/SiO_x/C stack, evidenced outstanding characteristics, such as a large resistance window about a factor 10^4, excellent resistance control with reduced variability, and remarkable pulsed cycling endurance, larger than $5x10^7$ cycles. This device also showed excellent retention properties at elevated temperature, with no data loss after one hour at 260 °C, thus meeting the requirements for embedded applications. Two SiO_x-based conductive bridge (CBRAM) selectors were investigated. In particular the Ag/SiO_x/C showed bipolar volatile switching with an ON-OFF ratio larger than 10^7. However, this selector evidenced a relatively long retention time, with large cycle-to-cycle stochastic variations. An extensive comparative study between the SiO_x- and HfO_2-based non-volatile devices was carried out using the same test vehicle, at fixed geometry and electrode materials, namely Ti top electrode and C bottom electrode. The SiO_x-based device showed a significantly larger forming voltage for comparable initial leakage current. The remarkably similar switching characteristics with a large resistance window, and the good retention performance up to 260 °C for 1 hour suggested a common switching mechanism where Ti impurities are injected from the top electrode into the switching layer. On the other hand, the HfO_2 device was more prone to negative set failure, whereas in the SiO_x it was not evidenced up to a very large negative voltage. To enable RRAM high-density storage applications it is essential to minimize the power consumption of the single device. A reliability study of the SiO_x device was then carried out to enable low current operation. A complex tradeoff was evidenced, where the high-resistance state (HRS) should be increased to accommodate the increased low-resistance state (LRS) variability at low compliance current, yet leading to an increased set voltage which causes overshoot effect. Operation at a relatively low compliance current of 20 µA with no overshoot and an acceptable programming variability was demonstrated. To operate at even lower compliance current, such tradeoff should be optimized on a cycle-to-cycle basis through program-and-verify techniques, to allow a sufficient read margin while minimizing the overshoot effect. Then, a compact Arrhenius model was developed to capture the evolution of the HRS resistance and set voltage after reset, as a function of the reset voltage amplitude and duration. The impact of time-dependent resistance fluctuations was studied as a function of the initial resistance, finding a constant relative distribution broadening for different initial values, with no resistance drift. These results support the need for a large resistance window to increase the robustness against resistance fluctuations. The feasibility of a novel neural network-based in-memory digital computing concept was proved. Multiple HfO_2-based RRAM devices were arranged into a circuit to perform logic computing. Several logic functions could be implemented with the same circuit topology by changing the applied voltages and the load resistor. Two-input NOR and NAND logic gates were experimentally demonstrated using 3 RRAM devices (2 input and 1 output) and a load resistor in a single-step operation. XOR was demonstrated with 3 devices in 2 steps, by sequentially cascading NIMP and C-NIMP operations. Also, a 1-bit full adder was implemented and experimentally demonstrated with 5 RRAMs and one load resistor, in 2 steps. To controllably and continuously tune the RRAM conductance for analogue computing applications, two incremental programming schemes were investigated, namely the incremental-current set and the incremental-voltage reset. The feasibility of a new in-memory analogue computing concept was experimentally proved. Here, a reconfigurable crosspoint of RRAM devices was connected into a feedback circuit to solve linear algebra problems, such as systems of linear equations and matrix eigenvectors, in just one computing step. In summary, this Doctoral Dissertation focused on the characterization of RRAM devices, including device performance and reliability, and on their application to in-memory analogue and digital computing systems. These results pave the way towards future RRAM applications in high-density crosspoint memory arrays for storage and computing beyond von Neumann.

PRIVACY-PRESERVING SERVICE DELIVERY IN INTERNET

Davide Andreoletti - Supervisor: Prof. Massimo Tornatore

Initially designed as a global network mainly aimed at connecting geographically-distributed computers, the Internet has rapidly become a complex infrastructure through which end-users are provided with a vast number of indispensable services. As a result of this transformation, today's Internet is a composition of many entities, which cooperate to perform the delivery of a service in a mutually-profitable manner. To better figure out which are the main entities in Internet and to understand their interdependency, let us think of a pyramidal structure where the basis is represented by the Internet Service Providers (ISPs), which own and manage the infrastructures that transport the Internet traffic. This traffic carries the services offered by the entities referred to as Over-The-Top (OTTs). Two remarkable examples of OTTs that are nowadays extremely popular are the Video Content Providers (CPs), which own and manage catalogues of video contents that users can retrieve, and the Online Social Network (OSN) providers, which offer users digital platforms where they can connect to each other and publish the most varied contents. On top of them, we find service providers that reach their users exploiting the platforms offered by giant OTTs (e.g., a Location-Based Service that performs advertising on an OSN).

In this thesis, we observe that the characteristics of the today's Internet open the door to the implementation of more complex cooperative schemes with respect to the basic ones that we have briefly mentioned. In particular, we notice that:

- Such entities generally have different footprints (e.g., an ISP offers Internet connectivity within a limited area, while a CP distributes its contents on a global basis) and
- they possess different information about the final users (e.g., if encryption schemes are in place, only the CPs know users' preferences)
- The wide use of virtualization strategies (which allow to decouple a service from the underlying physical devices) is making the boundaries in Internet increasingly blurred (e.g., a service can be offered within an area not covered by its provider).

An example of service that particularly benefits from an increased cooperation (namely, between ISPs and CPs) is video content distribution. In fact, if the CPs can serve their videos from network
positions closer to the users (e.g., inside the ISP's network), the offered Quality of Experience (QoE) is enhanced, by virtue of a reduction of the retrieval latency and the congestion probability. For example, in one of the works described in this thesis, we address the problem of optimally deploying Virtual Servers (VSs) inside the network of the ISP, from which the CP can better serve Live Videos (LVs) to its viewers. This is a clear example of an optimization process that has to be executed jointly by multiple entities, as each of them possesses a portion of the data needed to perform the optimal deployment. Specifically, only the ISP knows the precise users' position (since it covers the last segment of traffic delivery) and, under the use of encryption schemes (which is a common practice nowadays), only the CP knows the LVs that users request. However, these data are deemed privacy-sensitive by users and considered business-critical assets by the owner entities, which may therefore not be willing (or entitled) to freely share them with each other.

To address this issue, in this thesis we develop methodologies aimed to improve the effectiveness of service delivery without sacrificing privacy. In other words, we propose *privacy-preserving* data sharing solutions approaches that make the entities involved in the delivery of a service only able to extract from these data the information that they need to improve the service, but not to violate privacy. To this end, we consider several services provided over the Internet and we formally define the privacy requirements of the involved entities (e.g., CPs, ISPs and final users). As the privacy-preserving strategies often reduce the knowledge that it is possible to extract from data, it may happen that privacy can only be guaranteed at the cost of services' effectiveness degradation. In this research, we perform numerical evaluations of this phenomenon (which is better known as *privacy-utility* trade-off) in several scenarios.

In this research, the service that we mostly consider is video content distribution, which poses nowadays the strongest pressure on ISPs' infrastructures. Therefore, ISPs are constantly looking for innovative and reliable strategies to handle the impressive amount of traffic generated by the CPs. As explained in the aforementioned example, the ISP is in the favourable position to be much more than a simple traffic transportation carrier. For example, caching strategies represent a consolidated solution to both reduce network resource occupation in ISPs' networks and increase the QoE offered by the CPs. In fact, by performing caching an ISP stores a portion of the CPs' catalogues in servers located within its area (i.e., the *caches*) and serve contents directly from there. The result is a reduction of network traffic, retrieval latency and congestion probability. To be effective, caching requires that the most requested (i.e., popular) contents are delivered from inside the ISP's network. In a context of all-encrypted web, however, the ISPs are not aware of the contents traversing their network and, therefore, they are not able to assess their popularity without implementing advanced forms of cooperation with the CPs. In relation to this, in this research we mainly consider the privacy issues resulting from such cooperation. For example, we develop strategies to guarantee that an ISP can perform effective caching without discovering the popularity of CP's contents, or we propose techniques to enable a CP to serve a LV from a source close to its viewers, without knowing their position. In particular, the latter objective is achieved by employing an existing secure multiple-party computation protocol to perform a private set intersection.

Then, we consider the Network Neutrality (NN) issues arising from the implementation of cooperative caching strategies. Specifically, we explore the problem of subdividing a limited ISPs' cache storage among a set of CPs in a way that is both efficient (i.e., that minimizes the network resource occupation) and fair towards the CPs that exploit the caching system. We propose our definition of NN-compliant caching and design a protocol (based on the Shamir Secret Sharing scheme) to enforce it in a privacy-preserving manner (e.g., CPs are not required to expose the information about their contents' popularity).

Our attention is then shifted to the protection of users' location privacy in OSNs. Specifically, we develop a machine-learning-based privacy control tool that Twitter users can employ to (i) estimate how accurately their locations can be inferred from publicly-available data and (ii) to understand the factors that mainly affect their vulnerability to this inference. Then, we also propose data perturbation techniques and provide a qualitative evaluation of the trade-off between users' privacy and effectiveness of a Location Based Service (LBS) that exploit the OSN platform to reach its users.

Finally, we focus on the privacy-preserving Virtual Network Embedding (VNE) problem over a multi-ISPs infrastructure. In the considered problem, a customer is willing to find the most cost-effective deployment of a set of virtual functions over a wide-network composed of several ISPs, which keep the information about their network infrastructure private (e.g., the cost of traversing a link is not exposed). In particular, we propose a novel privacy-preserving Reinforcement Learning algorithm, which is implemented under the Shamir Secret Sharing scheme.

PRINTED PATCH ANTENNAS FOR MODERN WIRELESS COMMUNICATION SYSTEMS

Hamid Aslani

The increasing demand for higher bandwidth and higher speed wireless communication motivates the exploration of modern wireless communication. Ultra-wide band (UWB) technology is one of the most promising solutions for future communication systems due to the high data rate and excellent immunity to multi-path interference. Also, The IEEE 802.11ad and IEEE 802.11ay operating on 60 GHz mmWave are the two most expected wireless local area network (WLAN) technologies for ultra-high-speed communications. The 802.11ad standard (WiGig) provides throughput speeds of multi-Gb/s covering tens of meters by offering a wide beamforming channel in 60GHz ISM band channel.

In the proposed work, we will focus on both the above mentioned technologies. As a first approach,a novel wide band microstrip patch antenna (MPA) configurations that can be used for UWB applications with enhanced performance is designed and dicussed. The impedance bandwidth of the proposed antennas has been enhanced by using techniques such as patch sloting, as an improved technique. An impedance bandwidth between 5 -11 GHz has been achieved. The antenna performance has been characterized in terms of the reflection coefficient, peak gain and radiation characteristics. We also present an efficient microstrip patch antenna (MPA) with superstrate Technique. The antenna configuration can be used for UWB applications. Based on the problem of interference in modern wireless communication systems. the proposed work has been extented in order to design an efficient UWB antennas with band rejection characteristics at WLAN band of 5.2 GHz. The Simulations were performed using different EM software such as Ansys HFSS which uses the Finite Element Method (FEM) and Finite-Difference Time-Domain method (FDTD). We follow the strategy as the first step by designing patch antenna operating in Ultra-Wide Band (UWB) frequency range and the next step, the UWB antenna is embedded with anisotropic filter structure (SRR) which has been demonstrated and its band rejection properties has been shown.

A numerical investigation and analysis of a high-gain printed antenna arrays is performed for mm-waves wireless applications. The structure presented based on microstrip arrays antennas, the proposed antenna is designed for 60GHz high gain backhaul transceiver systems. An array structure has been also proposed to cover 360 degree. In order to analyze and validate the electromagnetic performance of these antennas, a numerical analysis was performed by ANSYS HFSS, which employs the Finite Element Method (FEM).

MODELS, CODE GENERATION, AND ABSTRACTION A TRIPLE APPROACH TO ENHANCE ROBOT SOFTWARE DEVELOPMENT

Gianluca Bardaro - Supervisor: Prof. Matteo Matteucci

In recent years, robotic applications have surged and the popularity of robotics has increased both in academia and industry. Robotics is evolving from its mechatronics roots, more focused on the development of the ideal hardware platform, to explore the advanced functionalities offered by complex applications. We are entering a new software age for robotics. Unfortunately, the tools available to developers are not on par with the expectations. Nowadays, developing an application for a robot is more similar to craftsmanship than engineering. An all-around robotics expert with a combined knowledge about the application, the capabilities of the platform, and the underlying middleware is necessary to guide the design and development process. Our objective is to provide a collection of methodologies, techniques and tools to support all the actors involved in the development process of a robotic system. The contribution is threefold, with each part targeted to a specific development role.
The system designer needs tools to outline the architecture of the system and describe the high-level interactions and requirements of components. This can be achieved using a modelling language to describe components and their inner workings in an agnostic way with respect to the underlying middleware. For our approach, we exploited existing technologies: the Architecture Analysis and Design Language (AADL), as the main modelling language, and Abstract Syntax Notation One (ASN.1) or JSON, as data modelling languages. To achieve a general representation of robotic architectures we relayed on the fact that the most middleware adopts a component-connector paradigm. We then focused on creating a more specialized description to model architectures implemented using the Robot Operating System (ROS). The generalized approach covers the basic elements of a robotic middleware: components (i.e., executable elements), ports (i.e., communications interfaces) and connections (i.e., communication channels). The ROS-based specialized description goes more in details by providing models for messages and the internal structure and functionalities of ROS components. The direct advantages for the system designer are multiple: an architectural overview of the system, early detection of design flaws, system consistency checks, and early documentation. Additionally, the system designer can rely on a library of already existing templates, this simplifies the design of the system and makes it more robust. Lastly, by basing this work on AADL it is possible to exploit all the existing analysis tools, such as latency estimation, computational load, hardware allocation and fault propagation.
The component developer, together with the domain expert, should focus only on the implementation of the internal logic and not on the structure of the component itself since this is the role of the designer. To do so, they need an environment that abstracts from the middleware-related code and provides a contained development space. Building on top of the model created by the system designer, we provide the ideal development environment by delegating to an automatic code generator most of the essential implementation, and by defining a bounded reference component that can be used as a starting point. The target implementation is based on a reference ROS node specifically engineered to minimise the amount of boilerplate code and to provide additional features that are usually managed by the component developer, few examples are internal life cycle of the node, well-defined initialization procedure, encapsulation of parameters and internal state, clear separation between the middleware and implementation. The latter is particularly important for the role of the domain expert. Since they are expert of a specific domain and carrier of specialized and valuable knowledge, they often do not, and ideally, should not implement the component directly, they should have access to a suitable interface. In our proposed model and automatic code generation approach a domain expert can implement a functionality independently and then embed it in the model, the automatic code generator will include it in the final implementation. The application developer implements high-level applications only loosely related to the underlying architecture of the robot. In practice, it should exist an abstraction layer between the low-level functionalities provided by the components and the applications. Robot configurations vary greatly; however, it is possible to abstract most of the capabilities independently from the system. Using these abstract robot capabilities, the application developer can implement application compatible with multiple robots with minimal modifications. We define them as low or medium-level functionalities (e. g., directional movement or navigation) that can be used to interact with the robot. The capabilities are defined manually by analysing different types of robots (i.e., mobile platforms, drones and manipulators), but the active capabilities on a running system are extracted automatically by analysing the ROS graph. On top of the concept of capabilities, we developed an abstraction layer to decouple the application from the underlying middleware. In our approach, we implemented a bridge between the capabilities and ROS systems. To do so, we developed a dynamic ROS component creating a bidirectional communication with an external system. We provide dynamically defined APIs a developer can use to interact with the robot through capabilities. All these methodologies, techniques and tools are all part of a continuous design and development process. The system designer uses the modelling tools and templates to define the architecture of the system. He can embed directly the reference to the source code developed by the domain expert, and, using properties, even enrich the component with their evoked capability. Through automatic code generation, most of the source code is already available with minimal effort, at this point the component developer can finalise the implementation, for example by adding special interfaces with the hardware components or specific initialization and shutdown procedures. The result is a robust system where all the components are known, well designed and well implemented, this is the suitable starting point for an application developer to exploit safely the abstraction layer defined using robot capabilities.

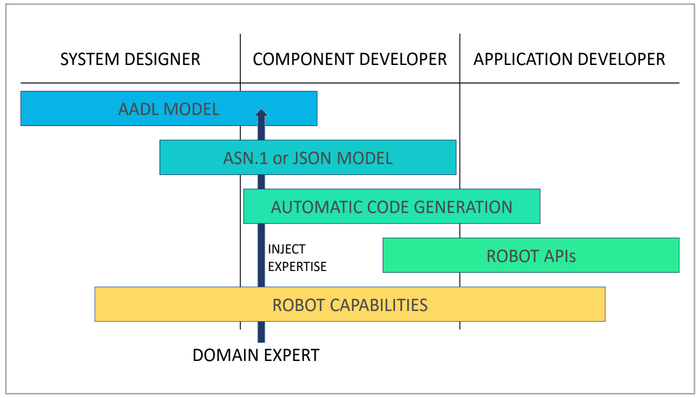

Fig. 1

AUTONOMOUS DRIVING AT THE LIMITS OF HANDLING

Marco Baur - Supervisor: Prof. Luca Bascetta

To be safer than their human counterparts, automatic pilots must be able to control the car up to its limits of handling. Therefore, this PhD thesis
aims at overcoming some of the limiting assumptions under which most of the path tracking controllers which can be found in the literature, along with Advanced Driver Assistance Systems (ADAS), have been developed:

- small vehicle sideslip angle, an assumption not satisfied when the car drifts due to, for example, a sudden decrease of rear wheel friction coefficient, which may be the result of a wet or icy road patch;
- linear relation between tyre lateral force and slip angle, a hypothesis only valid for small values of the tyre slip angle. Therefore, automatic drivers leveraging on it cannot take advantage of all grip made available by tyre-ground interaction. This fact is restrictive when the car is travelling on a low grip terrain, as for instance a wet road, and when a suddenly appeared obstacle must be avoided to prevent a collision.

These two assumptions hold only in a normal driving condition, i.e., far from the limits of vehicle handling. However, we strongly believe that an automatic driver, to be really safer than his human counterpart, has to overcome these assumptions, being able to exploit all the grip made available by tyre-ground interaction and to hold car control even in the presence of a large sideslip angle.

To study the problem of autonomous driving in the presence of large vehicle sideslip angles, this work focuses on drifting manoeuvres. Drifting is an aggressive manoeuvre, which brings the car close to its limits of handling, and which is characterized by a large vehicle sideslip angle, counter-steer and throttle action, for the control of vehicle lateral dynamics. The interest for this manoeuvre is justified by the fact that it is not only used by rally drivers on slippery terrains, but sometimes it is the only viable option to avoid an obstacle in an emergency condition. In addition, it has been recognized as a driving technique expanding vehicle mobility at low speeds.
The autonomous execution of a drifting manoeuvre represents a complex control problem, due the multiple number of inputs and outputs, the saturation constraints on the control inputs, and the nonlinearities of the underlying dynamics. This work leverages on Linear Quadratic (LQ) control strategy to design a drifting stabilization controller. Though several works have already used this control approach for drifting stabilization, the following novel aspects characterize this thesis:

- differently from the other works which are present in literature, we used the same control input available to a human driver, namely front tyre steering angle and longitudinal force developed by rear tyres. The latter of these is related to torque delivered by motor, which, in turn, is linked to throttle command;
- for the first time, to the best of authors' knowledge, LQ approach has been extended to perform, simultaneously, drifting stabilization and tracking of a circular reference path;
- last but not least, LQ controllers have been tested by means of extensive experimental campaigns, aimed at assessing their performance, robustness and shortcomings. These experiments have highlighted the robustness of the LQ control approach in stabilizing drifting and in tracking a circular reference path.

To perform experimental tests, a dedicated setup was built for the purpose of this PhD research project. It consists of a Radio Controlled 1:10 scaled car, which has been made autonomous by the addition of sensors and computational units. A picture of the experimental platform is shown in Figure 1.

To be able to design model-based controllers, a model of steering actuator dynamics and of vehicle lateral and longitudinal dynamics were identified, and their parameters were estimated. Validation experiments highlighted that single-track vehicle dynamic model, which is typically used for the design of ADAS and real car path tracking controllers, is able to reproduce vehicle lateral dynamics, up to the limits of handling. To further confirm that a scaled vehicle, in place of a real car, can be used to assess performances, robustness and shortcomings of path tracking controllers, while avoiding the danger and the cost of accidental collisions, the dynamic similitude between the experimental platform and a real vehicle has been verified.

Apart from drifting, a control-oriented model, suitable to the development of path tracking controllers able to take advantage of all the grip made available by tyre-ground interaction, has been studied. This model, which is called Affine in the Force Input (AFI) model, takes front tyre lateral force as control variable, in place of front tyre steering angle. This implies that:

- nonlinearities of front tyre lateral force and front tyre slip angle relationship are extracted out from vehicle lateral dynamic model, yielding a linear dynamical system;
- constraints on maximum and minimum front tyre lateral force, which are dictated by tyre-ground friction coefficient and tyre normal load, are easily expressed as lower and upper bounds on the control input.

Therefore, AFI model is suited to the design of linear path tracking controllers, which are able to take advantage of the maximum lateral force which can be developed by tyres. In particular, AFI naturally calls for the adoption of a linear Model Predictive Control (MPC) approach, for the design of path tracking controllers.
AFI benefits come at the expense of yaw rate oscillations, which arise when vehicle is travelling at high speed and front tyre lateral force is the control variable. These oscillations, also referred to as fish tail phenomenon, severely hamper passenger comfort by giving to him the impression of an impending spinning and, hence, must be avoided. A novel aspect of this work lies in the analysis and explanation of the cause of these yaw rate fluctuations, which have never been addressed so far. It has been discovered that, by controlling front tyre lateral force, a hidden pole placement control law, which decreases yaw rate damping at high speed, is applied.
Then, it has been shown, by means of simulations, that undamped yaw rate dynamics not only affects system open loop response, but also the performance which can be achieved with an MPC path tracking controller. Therefore, a solution increasing the damping of yaw rate poles at high speed was proposed. A pole placement control law was designed and added to MPC path tracking controller. Simulations showed that these nested loops control architecture is able to track a reference path, while using all the grip made available by tyre-ground interaction, without yaw rate oscillations when vehicle is travelling at high speed.

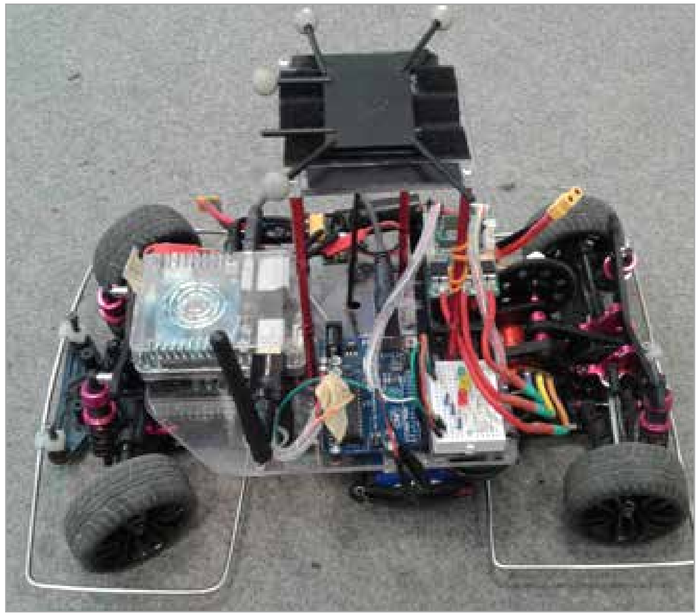

Fig. 1 - The autonomous scaled car used to experimentally test autonomous drifting controllers

ADVANCING JOINT DESIGN AND OPERATION OF WATER RESOURCES SYSTEMS UNDER UNCERTAINTY

Federica Bertoni - Advisor: Prof. Andrea Castelletti

Co-advisor: Dott. Matteo Giuliani, Prof. Patrick M. Reed

Globally, many countries are actively seeking to maximize the hydropower potential of major river basins, yielding proposals for constructing approximately 3,700 major dams in the near future. The planning of new water reservoir systems raises several major challenges that must be conjunctively accounted for within the system design phase, namely (i) potentially conflicting and heterogeneous objectives; (ii) interdependency between dam size and operations; (iii) future uncertainties in the main external drivers (e.g., climate, human demands); and (iv) vast amount of information that is becoming increasingly available to system planners at different temporal and spatial scales. Such issues must be jointly addressed through novel, integrated approaches in order to design efficient yet sustainable infrastructures able to satisfy multiple water needs and perform well under a wide range of external future changes.

Building on these research challenges, the main goal of this thesis is to advance the current planning and operation of water reservoir systems, focusing on the coupling of dam sizing and operation design in order to thoroughly capture their interdependencies also with respect to uncertainty in the main external drivers. In addition, the role of exogenous information (e.g., streamflow forecasts) in dam design is investigated to further analyze how dam design is shaped by information feedbacks.

We contribute novel methodological approaches as the primary outcome of our research, which have been developed by extending and integrating existing optimization techniques traditionally applied to the water management field in order to additionally account for the planning dimension of the problem, cover all the challenges of current planning and operation of water resource systems, and eventually provide supporting tools to water system planners intended to design water reservoir systems in complex, highly uncertain decision making contexts. In particular, this research has been developed in three main phases associated to the following methodological contributions:

(i) Integrated dam design via Reinforcement Learning (RL); (ii) Robust framework for dam design; and (iii) Informed dam design by forecasts.

The first outcome is a novel Reinforcement Learning (RL)-based approach to integrate dam sizing and operation design, while significantly containing computational costs with respect to alternative state-of-the-art methods. Our approach first optimizes a single operating policy parametric in the dam size and, then, searches for the best reservoir size operated using this policy. The parametric policy is computed through a novel batch-mode RL algorithm, called Planning Fitted Q-Iteration (pFQI). The proposed RL approach is tested on a numerical case study, where the water infrastructure must be sized and operated to meet downstream users' water demand while minimizing construction costs. Since real case studies can present a vast array of complexities ranging from the definition of the system boundaries to include all the relevant drivers and the unfeasibility of conducting various experiments, to the lack of information regarding the decision makers' objective functions and the actual economic value of planned infrastructures and associated operations, the choice of a synthetic case study allows to address these issues while allowing to perform experiments in a controlled environment and therefore assess the actual potential of novel methodological contributions. Results show that our RL approach is able to identify more efficient system configurations with respect to traditional sizing approaches that neglect the optimal operation design phase. However, when compared with other integrated approaches, we prove the pFQI algorithm to be computationally more efficient.

Secondly, robustness of dam designs with respect to uncertainties in the main external drivers is addressed, by developing a robust dam design framework that jointly considers sizing and operations while explicitly accounting for key human and hydro-climatic uncertainties. Bridging the Multi-Objective Robust Decision Making and Evolutionary Multi-Objective Direct Policy Search methods, our dam design framework jointly optimizes planning (i.e., dam size) and management (i.e., reservoir release policy) in a single optimization process. We demonstrate the potential of this integrated dam design framework through an ex-post design analysis of the Kariba dam in the Zambezi river basin. The Kariba dam was built in 1960 and sized via standard design methods that assumed a pre-defined operating rule to maximize hydropower production. A unique contribution of this work is to utilize ex-post analysis of an actual large dam to validate our approach over an existing water infrastructure designed via classic sizing methods, by assessing whether the dam has been over-dimensioned and/or under-performs under uncertain future changes. While providing insights to a high impact reservoir system, our ex-post analysis more broadly is a rigorous means of evaluating our framework relative to standard design techniques. Our results show that careful exploration of the coupled planning/operation search space yields designs that significantly outperform the existing Kariba system. Moreover, we demonstrate that our framework leads to a significant reduction in capital costs (i.e., smaller reservoir sizes) while simultaneously improving system robustness with respect to changing hydro-climatology and human irrigation demands.

In the end, we explore the added benefits of including valuable information, and in particular streamflow forecasts, during the optimal dam design phase to identify more efficient system configurations. Building on the robust dam design framework mentioned above to solve coupled dam sizing and operation design problems, in this third contribution we want to assess whether searching for more flexible operating policies informed by streamflow forecasts allows to design smaller reservoir sizes with respect to solutions that do not rely on forecast information. This is achieved by first selecting the most informative lead times of perfect streamflow forecasts to be then included within the dam design phase to inform reservoir operations. After estimating the corresponding perfect forecasts value, we analyze its sensitivity with respect to realistic streamflow forecasts characterized by different biases synthetically generated from the perfect forecast set. We demonstrate the potential of our contribution through an ex-post design analysis of the Kariba dam in the Zambezi river basin. Results show that we are able to successfully identify the most valuable perfect streamflow forecasts to improve the Kariba system design, covering about 20% of the space for improvement estimated under exact knowledge of the future and reducing capital costs by 20% while still attaining a satisfactory performance. When informed with realistic streamflow forecasts characterized by different systematic errors, the Kariba dam system has proven to be more sensitive to a forecasts over-estimation, leading to an 8% maximum loss in the corresponding forecast value.

FREQUENCY SYNTHESIZERS BASED ON PLLS FOR CELLULAR RADIO APPLICATIONS

Luca Bertulessi - Supervisor: Prof. Salvatore Levantino

New 5G communication standards have to cover different frequency ranges to face the increasing demand of high data rate wireless access and the growing number of devices simultaneously connected to the network.
The legacy mobile radio frequency spectrum, located below 6GHz, is not able to satisfy these new requirements alone due to the limited channel bandwidth. Therefore, 5G radio devices need to exploit also a new spectrum segment above 30GHz in the so-called mm-wave range.
The phase noise requirements for the sub-6GHz and for mm-Wave frequency ranges are different.
Due to the narrow channel spacing and the crowded spectrum segment, sub-6GHz devices have to meet stringent phase noise masks while in a mm-Wave device, due the high output frequency, low jitter is mandatory to guarantee the target bit error rate (BER).
Moreover, mm-Wave design usually exploits BiCMOS technology to guarantee low noise performance at such high frequency compromising the integration of complex digital signal processing algorithms.

The target of this PhD thesis is to identify which phase locked loop (PLL) and oscillator architectures are the best candidates to reduce the output phase noise and satisfy the 5G stringent specifications in both the frequency ranges, using a 65nm full CMOS technology.

At first step, the performance of Type-I and Type-II phase locked loop in terms of output jitter and locking transient are compared.
The two schemes, with the same noise sources, are both able to reach the same integrated output jitter.
Type-I has a fast locking transient with respect the Type-II due its larger loop bandwidth.
However, Type-I filters less the DCO flicker noise with respect to the Type-II topology and moreover the Type-I locking range is limited by the phase detector and the DCO dynamic range.
Type-II shows instead zero phase error at the steady-state frequency. This property allows to implement a small dynamic range phase detector such as the Bang-Bang phase detector that reduces the power consumption and the area occupation with respect to the standard charge-pump PLL.

For sub-6GHz PLL the analysis faces the trade-off between phase noise and the power consumption.
To deal with some realistic and stringent spot phase noise, in the absence of detailed 5G specifications, the ETSI GSM 05.05 standard was therefore considered.
Demanding phase noise specifications force to increase the efficiency of noise critical blocks.
Energy efficiency of CMOS LC oscillators is discussed, and the best topology to meet the out-of-band phase noise mask without increasing too much the power consumption is identified.
The adoption of a digital PLL architecture based on a single bit phase detector (bang-bang) improves the power consumption with respect to the power-hungry multi-bit time to digital converter.
On the other hand, the trade-off between bandwidth and locking transient is exacerbated due to the limited bang-bang dynamic range. To overcome this issue, a novel technique based on nested DCO control loops is presented, thus speeding-up locking in bang-bang digital PLLs.
The implemented sub-6GHz fractional-N synthesizer has an output frequency from 3.59GHz to 4.05GHz with an integrated output jitter of 182fs and a FoM of -247.5dB.
Measured output phase noise is compliant with the stringent ETSI GSM 05.05 phase noise requirements and thanks to the implemented locking technique the loop can perform a frequency step of 364MHz, to within 10 MHz from the final frequency, in only 5.6us. These figures advance the state of the art in terms of power-jitter FoM and locking time of a digital PLL.

The challenging mm-wave targets were instead defined to push performance, thus reducing the existing gap with respect to the sub-6GHz synthesizer. A new sub-sampling bang-bang phase detector topology has been investigated. A low-power divider-by-six injection-locked prescaler has been adopted in the feedback path to reduce the overall power consumption. The resulting implemented synthesizer can operate in the mm-Wave range between 30.4 and 34.2 GHz with an integrated RMS jitter below 180 and 197.6 fs for the integer-N and fractional-N channels, respectively. The fractional spurs, measured at the 5-GHz prescaler output, are below -54 dBc, even considering near-integer channels. To prove the robustness of the divider-by-six prescaler, the architecture is also suitable to generate sawtooth chirps around 33.4GHz with peak-to-peak amplitudes up to 1.14GHz. The -238.6dB jitter-power figure of merit for fractional-N channels advances the state of the art in terms of power-jitter FoM for CMOS mm-Wave frequency synthesizer.

A RANDOMIZED MODEL STRUCTURE SELECTOR FOR COMPLEX DYNAMICAL SYSTEMS

Federico Bianchi - Supervisor: Prof. Luigi Piroddi

This thesis addresses the problem of choosing suitable model structures for dynamical systems when the data-driven model learning is pursued with parametric methods. The *model structure selection* (MSS) problem is known to be challenging due to its combinatorial nature which requires in principle to exhaustively search for the model terms to be included into the model within a space that might be large. Accordingly, many strategies have been proposed with the aim of exploring in a smart way the model structure space, ranging from greedy incremental policies, regularization based techniques, evolutionary methods, and probabilistic approaches. In this thesis, the MSS problem has been investigated with reference to the identification of nonlinear systems when data are distributed among agents and that of switched nonlinear systems, and the estimation of the process noise statistics in Kalman filtering. Each scenario has its own peculiarities that impact on the MSS problem. The proposed solutions are all based on a probabilistic reformulation of the selection problem whereby a probability distribution is defined over the model structure space. Accordingly, we proposed adequate sample-and-evaluate procedures that allow to finitely tune the parameters of the introduced probability distribution in order to concentrate its probability mass on a target structure.

Distributed Nonlinear Model Identification - In this thesis, we focus on the Nonlinear AutoRegressive with eXogenous input (NARX) class. Within this model class, the nonlinear mapping between data is often represented by means of a functional expansion involving lagged inputs and outputs. A popular choice for the form of the functional expansion is the polynomial one, which yields a linear-in-the-parameters model structure that is particularly convenient for parameter estimation purposes. On the downside, the model complexity of the NARX class grows rapidly with the model order and nonlinearity degree, and this motivates the interest in the problem of MSS for such model class. The problem of identifying a model of a nonlinear system from input/output observations is typically formulated as an optimization problem over all available data that are collected by a central unit, in the same operating conditions. However, the massive diffusion of networked systems is changing this paradigm: data are collected separately by multiple agents and cannot be made available to some central unit due, e.g., to band limitations or privacy constraints. Therefore, we address this novel set-up and consider the case in which multiple agents are cooperatively aiming at identifying a model for a system, by local computations based on private data sets. This problem is particularly challenging because the combinatorial nature of the MSS problem hampers the application of classical distributed schemes. Here, we propose a method that overcomes this limit by adopting a probabilistic reformulation of the model structure selection problem and seeking the consensus among agents on both the model structure and the parameter estimates at the same time.

Identification of Hybrid Nonlinear Systems - There are many physical processes whose behavior is characterized by different continuous dynamics (modes) among which the system can switch according to some discrete dynamics. For example, in electrical circuits, continuous phenomena, such as the charging of capacitors, are interrupted by switches or diodes. In a thermal control process a thermostat is used to control the temperature by switching on or off heating or cooling devices. More in general, complex systems exhibit different continuous dynamics as individual components are switched on or off. In such cases, a single dynamical model, even if nonlinear, is often not sufficient to capture the real dynamics of the system. Hybrid dynamical systems (HSs), whose behavior can be described by the interaction of time- and event-driven dynamics, provide a unified framework for the representation of such cases. Most research regarding the identification of hybrid systems (HSI) has focused on switched affine (SA) and piecewise affine (PWA) models due to their universal approximation properties and their simple interpretation. Indeed, they provide the simplest extensions of continuous systems that can handle hybrid phenomena. The optimization problem induced by the identification task is of a mixed-integer type, since it involves the identification of discrete variables (representing the mapping of the samples to the modes and the model structure associated to each mode), as well as continuous ones (the parameters of the models describing the continuous dynamics associated to the various system conditions). Many approaches have been proposed over the last two decades for the case of affine dynamics. Surprisingly fewer works have tackled the case of nonlinear continuous dynamics associated to the modes, in spite of its importance in modeling complex applications. Indeed, if no a priori information on the number of modes is available, one can in principle identify an arbitrarily high number of local linear models (and switchings among them) in order to achieve a good model accuracy. However, this prevents the identification of the real dynamics of the hybrid system and hinders its physical interpretation. It also greatly aggravates the combinatorial complexity of the optimization problem, due to the increasing number of switchings. In this thesis, we consider the identification of switched nonlinear systems in input-output form, namely Switched Nonlinear ARX (SNARX), in the case of unknown model structures. We propose a black-box iterative identification method, where each iteration is characterized by two stages. In the first stage the identification problem is addressed assuming that mode switchings can occur only at predefined time instants, while in the second one the candidate mode switching locations are refined. This strategy allows to significantly reduce the combinatorial complexity of the problem, thus allowing an efficient solution of the optimization problem using a randomized method.

Structure Selection of the Process Noise Covariance Matrix in Kalman Filter Applications - A third problem in which MSS is crucial is the estimation of the process noise covariance matrix in state estimation problems in the Kalman filter setting. Kalman filtering for linear systems is known to provide the minimum variance estimation error, under the assumption that the model dynamics is known. While many system identification tools are available for computing the system matrices from experimental data, estimating the statistics of the output and process noises is still an open problem which significantly impacts the filter performance. In fact, although some methods based on maximum likelihood and correlation approaches have been proposed in the literature, it turns out that the existing techniques are either too computationally expensive or not accurate enough. Above all, in many papers the process and output covariance matrices are provided as a prior knowledge or their estimation is   eclassified into an empirical tuning problem in which diagonal parameterizations are often assumed, for simplicity. Our study indicates that this assumption does not always provide the best compromise between computational complexity and tracking accuracy. This evidence encouraged us to further investigate this problem leading to an algorithm for the selection of the structure of the process noise covariance matrix, which actually is an adaptation of the same optimization algorithm originally employed for MSS in nonlinear identification applications.

SPACE-TIME PROCESSING FOR SOUND FIELD RECONSTRUCTION (SENSE)

Federico Borra - Supervisor: Prof. Augusto Sarti

The technological development, together with the growing interest in applications in the virtual/augmented reality realm, have given rise, in the recent years, to a series of challenging problems in different areas of research. As regards to the acoustic signal processing community, the sound field navigation problem has gained attraction for both researchers and companies. Virtual navigation of a sound field enables a listener to explore a recorded acoustic scene and, ideally, experience an accurate perception. In other words, a listener can freely move in an acoustic scene and, independently from its position and the positions of the microphones that have recorded the scene, listen as if she/he were actually there. In order to enable the listener to experience the immersion, applications should allow for six-degree-of-freedom, where users can both translate her/his position and rotate her/his head. This gives rise to a challenging problem that concerns the recording of the sound field at arbitrary positions. In this thesis we use either sound field reconstruction or virtual miking to indicate the process of estimating the sound field at positions that can potentially differ from the ones where has been recorded and we propose both parametric and non-parametric approaches to tackle such problem. Generally, parametric approaches are less demanding from the hardware requirement standpoint, but they usually rely on simplified model of the sound field. Instead, non-parametric approaches require an higher number of microphones and are based on the solutions of the wave equation and, typically, do not rely on any a-priori assumption.

The first goal that we pursue is to improve parametric sound field reconstruction methods by improving both the underlying models and the required parameter-estimation step. In fact, in most of the state-of-the-art methods, sources are considered to have an omnidirectional directivity pattern, i.e., they emit with equal power in all directions. However, in real-world scenarios, sources are typically not characterized by an omnidirectional pattern. One need only think, for example, of a conversation between two persons. According to the position of the listener with respect to the speaker the perception of the received sound differs. Since in sound field navigation application one can freely move in the acoustic scene, the source directivity must be taken into account. Moreover, the majority of the proposed method require the knowledge of the positions of the sources for each time-frequency bin that is a challenging problem. The methodology adopted in this thesis to tackle such problems is as follows: at first we identify suitable parametric models that keep into account the directivity of the sources and then we devise analysis techniques for the estimation of the model parameters based on the solution of sparsity-constrained optimization problem. This methodology enables to consider sources whose directivity pattern is not omnidirectional and to avoid an explicit estimation of the source location for each time-frequency bin.

The second goal that we pursue is to improve non-parametric reconstruction methods by trying to reduce the requirements in terms of both number of microphones and computational cost. The non-parametric methods proposed in the literature reduce the hardware requirements by either using higher-order microphones at the expense of their individual complexity or by assuming a sparse representation that comes at the expense of an higher computational cost and an higher number of hyperparameters that are usually difficult to tune. Such methods consider the sound field in a given point as due to the superposition of two contributions: the exterior field, defined as the sound field outside a (spherical) region that encompasses all the sources that generate it, and the interior field, defined as the sound field inside a (spherical) region that does not contain any source. We consider a scenario in which the acoustic scene of interest is surrounded by an array of higher-order microphones with no independent sources outside. Despite less general, this is a common scenario. In fact, it is sufficient to think of an application in which one wants to navigate an acoustic field generated by a speaker inside a room and recorded by a set of microphone surrounding it. In this case, the interior sound field is only due to effect of reverberation and therefore can be considered as a function of the exterior sound field and the environment. Exploiting the dependence of the interior field from the exterior field, we propose a methodology to tackle the non-parametric sound field reconstruction problem and relaxing the hardware and computational requirements. In particular, we probe the environment response to acoustic events by measuring the room impulse response (RIR) for a set of points and we show how the RIR measurements can be used to find an explicit expression that links the interior and the exterior field.

Lastly, the third goal concerns the study of different acoustic signal representations in order to perform analysis tasks such as sources localization and geometry inference. The analysis of the acoustic signals is at the root of the sound field reconstruction problem. In particular, the location of the sources is a required information for both the proposed parametric and non-parametric approaches. Also the information about the geometry of the environment can be exploited in both paradigms. As a matter of fact, in a scenario in which the interior field is a function of the exterior field, an accurate knowledge of the environment together with an estimate of the signals emitted by the sources and their directivity (or an estimate of the exterior sound field coefficients) are sufficient to estimate the interior field (reverberant component). As far as the source localization problem is concerned, the goal that we pursue is the extension of a sound field representation approach that maps the acoustic signals acquired by a linear array of microphones onto a domain known as ray space in order to accommodate different array geometries. The mapping operation can be interpreted as a two-stage approach: the first consists in a series of beamforming operations on sub-arrays, while the second in the mapping of the beamforming outputs onto the ray space. Here acoustic primitives, such as point-sources, appear as linear patterns and thus the localization of the sources can be accomplished using pattern-analysis techniques. The methodology adopted is aimed at improving both the mapping stages. On one hand we propose a new parametrization of the rays to accommodate 3D geometries. On the other hand, we propose an improvement of the beamforming phase by adopting a system that consists of a microphone array where each sensor is in turn a first-order differential microphone. In regards to the geometry inference problem, instead, we develop a novel representation of RIR measurements acquired using a linear array of loudspeakers/microphones and a single microphone/loudspeaker. Exploiting the sparse nature of the early part of the RIRs, Elastic Net regularization is applied to obtain a 2D polar-coordinate map, on which the direct path and early reflections appear as distinct peaks, described by their propagation distance and direction of arrival. Assuming a separable room geometry with four side-walls perpendicular to the floor and ceiling, and imposing pre-defined geometrical constraints on the walls, the 2D-map is segmented into six regions, each corresponding to a particular wall. The salient peaks within each region are selected as candidates for the first-order wall reflections, and a set of potential room geometries is formed by considering all possible combinations of the associated peaks. The room geometry is finally inferred using a cost function evaluated on the higher-order reflections computed via beam tracing.

CHARACTERIZATION AND MODELING OF SPIN-TRANSFER TORQUE (STT) MAGNETIC MEMORY FOR COMPUTING APPLICATIONS

Roberto Carboni - Supervisor: Prof. Daniele Ielmini

With the ubiquitous diffusion of mobile computing and Internet of Things (IoT), the amount of data exchanged and processed over the internet is increasing every day. This data-centric scenario requires secure data communication and storage and new computing primitives with low power consumption and reduced area in order to be embedded in resource constrained IoT devices. Moreover, in various emerging applications, such as autonomous driver-less vehicle navigation, the latency associated with cloud computing is usually unacceptable, thus requiring complex computations to be performed at the edge of the network, i.e. on the IoT device itself. For the past 50 years, the microelectronics industry has evolved according to Moore's Law, which describes the exponential growth of the number of transistors on integrated circuits (IC), thus enabling a steady increase in computing capabilities. Such scaling path is nowadays coming to an end due to the increasing heat dissipation and the growing impact of charge/dopants discretization on device performances. In addition, von Neumann architecture-based digital processors are hindered by the performance gap between the central processing unit (CPU) and memory, which makes them generally inefficient in terms of both energy and latency particularly in data-centric applications such as machine learning.

To face such challenges, emerging memory technologies, which are sometimes related to the memristor concept, such as resistive switching random access memory (RRAM), phase change memory (PCM), ferroelectric memory (FERAM) and spin-transfer torque magnetic memory (STT-MRAM) have recently gained significant interest thanks to their non-volatility, area scalability, low current and fast operation, and compatibility with the CMOS process. Among various emerging memory technologies, STT-MRAM is a promising candidate for next-generation high-density and embedded non-volatile memory. This boost of applicative interest in STT-based memory is due to the electronic switching which enables high speed, negligible resistance variation, and high endurance. However, cycle-to-cycle switching variability and reliability issues in terms of cycling-induced dielectric breakdown are extremely critical for STT-based memory circuits.

This doctoral dissertation addressed these issues by means of detailed electrical characterization and modeling of STT magnetic memory. Moreover, emerging security and computing applications of STT-MRAM, such as neuromorphic and stochastic computing, were demonstrated thanks to the exploitation of stochastic switching. The research was performed in the framework of the European Research Council (ERC)-funded project RESCUE (Resistive-Switch CompUting bEyond CMOS) under the supervision of Prof. Daniele Ielmini.

A compact model for conduction in magnetic tunnel junctions (MTJs) is firstly introduced in order to carefully reproduce the voltage-dependent tunnel magnetoresistance (TMR) of the devices. The model is verified against experimental current-voltage (I-V) and resistance-voltage (R-V) MTJ characteristics in both DC and pulsed-AC regimes. Random thermal fluctuations induce cycle-to-cycle statistical switching variations, which are harmful for memory circuits. Thus, stochastic switching was thoroughly characterized by means of write error rate (WER) study as a function of the applied voltage and pulse width. Data demonstrated an anomalous thermal regime of switching below 200 ns, which is not taken into account by the Néel-Brown thermal model. A new compact model for stochastic switching allowed for the correct prediction of WER data on applied voltage and pulse duration. The anomalous thermal regime of switching is explained by the non-linear lowering of the energy barrier associated with the perpendicular magnetic anisotropy (PMA). The model is extensively verified also against experimental distributions of switching time-delay and switching voltages. The compact model demonstrated accurate results while keeping a simple mathematical formulation comprising only 3 equations and 5 free parameters, which enables straightforward portability to other MTJ technologies and easy integration in commercial circuital simulators.

Even though critical for memory applications, stochastic switching in STT-MRAM can be exploited towards the realization of an ultra-compact entropy source at the nanoscale. To demonstrate this concept, a novel true-random number generator (TRNG) based on the differential read of stochastic switching was introduced and its performances tested against the NIST statistical tests. Moreover, a stochastic spiking neuron showing Poissonian spiking activity with voltage-controlled frequency was demonstrated. The compact model for stochastic switching was employed for the simulation of computing applications with spiking signals, such as an analogue multiplier and spiking neural network (SNN).

Another critical issue for STT-MRAM is the cycling-induced degradation of the MgO tunnel barrier, which is connected to the switching current flowing though the MTJ and eventually leads to endurance failure. To better understand this phenomenon, STT-MRAM cycling endurance was measured as a function of the applied voltage, pulse polarity, pulse duration and device area. A novel semi-empirical endurance model based on defect generation, activation and diffusion in MgO is then introduced and extensively verified against the experimental data. The model allows for the prediction of STT-MRAM cycling-induced failure under various programming conditions.

A solution to the cycling-induced failure is to avoid a writing phase involving a high current through the MgO. This can be achieved by the emerging spin-orbit torque MRAM (SOT-MRAM) concept, where the switching of an MTJ structure is obtained thanks to the SOT effect. To demonstrate this concept, heavy-metal/MTJ heterostructures were integrated in the Politecnico di Milano cleanroom (PoliFab) and have been electrically characterized with a custom-built electromagnet-based probe station. Finally, the experimental observation of SOT-induced switching paves the way towards future research works involving a thorough characterization of the stochastic switching behavior and computing applications of the fabricated SOT-MRAM devices.

ALLOWING A REAL COLLABORATION BETWEEN HUMANS AND ROBOTS

Andrea Casalino - Supervisor: Prof. Paolo Rocco

The future production plants will see more and more the presence of robots, performing a large variety of actions, from simple moving tasks to complex manipulations. In the past, the most active research lines were mainly devoted to developing algorithms for motion or force control, while the current trend is trying to provide robots with some more sophisticated cognitive capabilities. Robots are no longer conceived as something that will completely replace the human workers since they will become valuable assistants, The future robots will have a certain level of autonomy, in order to prevent humans from doing something wrong or anticipate the human needs. This kind of tasks can be accomplished by robots only with an increased cognitive capability.

Collaborate with a human mate is a non-trivial task for a robotic device. Indeed, robots should be able to behave as humans naturally do, interpreting the other worker's actions and forecasting future ones. The interpretation of human behaviour can be tackled by exploiting advanced sensors, providing the robots a huge amount of data to process. Finding a way to handle such data in a fast and reliable way is becoming crucial.

The main goal of this work was to develop algorithms and strategies able to allow an efficient collaboration between humans and robots, with particular attention to industrial contexts, where these agents have to alternate and synchronize for performing structured tasks, as for instance assemble some components.

This work presents four main contributions.

Firstly, it proposes algorithms and methodologies for inferring the current action that a human operator inside a robotic cell is undertaking, from the simplest ones as for instance those for reaching tools or objects, to the more complex ones as for instance performing a screwing. The reaching tool problem was addressed exploiting Gaussian Mixtures for building a model of the human intention, mainly considering the motion of the operator's wrists. At the same time, dynamic Markovian Random Fields were exploited to detect and the starting and the ending of human actions, reasoning on the motion performed by the human in the past.

At a second stage, higher-order Markov models were used to predict the human actions in the near-far future, according to the detected sequence of actions performed in the past.

The predictions of the human intentions were then exploited to develop innovative scheduling algorithms for planning the operations assigned to the robots in the plant, with the aim of assisting in an optimal way human mates, by minimizing the inactivity times. This is far from the most common paradigm adopted, which is a centralized approach that dispatches instructions to all agents, treating the human as no more than a high cognitive robot, refer also to Figure 1.

On the opposite, in this work the collaboration is conceived as human-centric: it is the human in charge of driving the evolution of the assembly phases, while robots must adapt to the human needs.

All the developed scheduling approaches consider a particular class of Timed Petri Nets, specifically derived for describing collaborative tasks.

Finally, innovative approaches for the motion control of collaborative robots were developed. Although this topic was already well studied, new algorithms were developed applying similar approaches to the scheduling techniques above discussed. Indeed, machine learning techniques can be adopted to predict the human motion during time, and then adapt the robot behaviour, this time in term of applying corrective actions for the motion of the robots. Gaussian processes were adopted for predicting the human motion, providing a probabilistic result. Then, stochastic optimization techniques were used to compute robotic trajectories compliant with latter prediction.

The combination of the above methodologies was able to significantly improve human-robot co-assemblies, refer to Figure 2, minimizing the inactivity times of both the human and the robots.

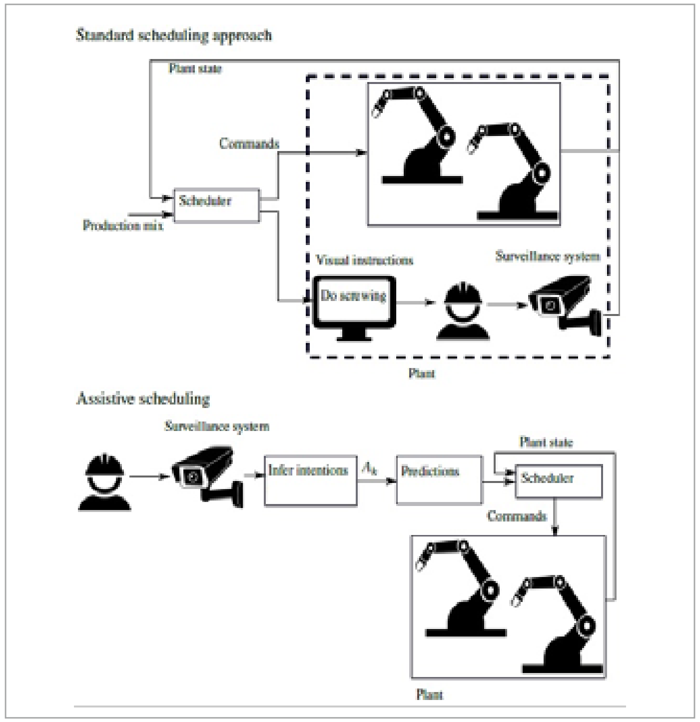

Fig. 1 - Differences between a standard (centralized) scheduling approach, on the top, and an assistive scheduling, on the bottom.

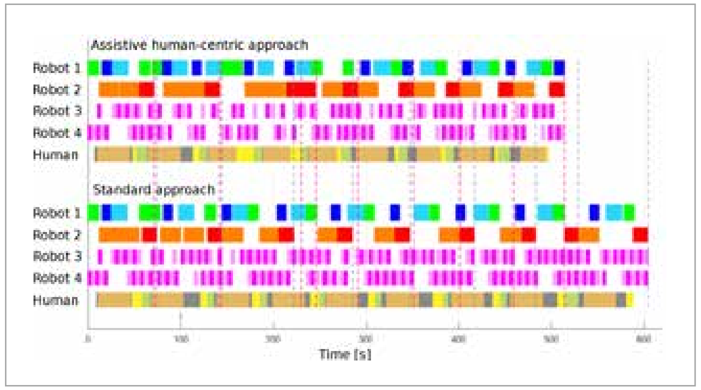

Fig.2 - Sequence of events occurred for a particular co-assemblies: one applying the standard centralized approach (bottom) and one the assistive approach developed in this work (top). The temporal duration of the activities performed by the agents is proportional to the length of the corresponding coloured bar. Gray bars represent the human waiting activity. The dashed vertical lines refer to time instants when a new finite product is available: the blue ones referring to the centralized approach, the red ones to the assistive one. As can be seen, the productivity of the plant for the approach at the top is significantly higher.

COORDINATION AND CORRELATION IN MULTI-PLAYER SEQUENTIAL GAMES

Andrea Celli - Supervisor: Prof. Nicola Gatti

The computational study of game-theoretic solution concepts is fundamental for describing the optimal behavior of rational agents interacting in a strategic setting, and to predict the most likely outcome of a game. Equilibrium computation techniques have been applied to numerous real-world problems. Among other applications, they are the key building block of the best poker-playing AI agents, and they have been applied to physical and cybersecurity problems.

This thesis focuses on extensive-form games, which can model sequential interactions among players, imperfect information, and outcome uncertainty. In this setting, a vast body of literature focuses on the computation of Nash equilibria in two-player, zero-sum games, where recent results demonstrated that it is possible to compute strong solutions in theory and practice. Some works weaken the two-player, zero-sum assumptions by considering two-player general-sum games, or by studying multi-player games with particular structures, such as compact games. Recent results on multi-player poker showed that it is possible to exploit techniques developed for the two-player, zero-sum setting in games with more than two players. However, these results are specific to the poker domain, and they lack theoretical guarantees on general multi-player extensive-form games.

The two-player, zero-sum model is somewhat restrictive for most practical applications, as many real-world scenarios are not zero-sum and involve more than two players. Moreover, especially in general-sum games, the adoption of a Nash equilibrium may present some difficulties when used as a prescriptive tool. Indeed, when multiple Nash equilibria coexist, the model prevents players from synchronizing their strategies, since communication between players is prohibited. In real-world scenarios, where some form of communication among players is usually possible, different solution concepts are required as communication allows players to coordinate their behaviors. This thesis focuses precisely on problems where communication is crucial to reach coordination.

When the players' interaction takes place sequentially, they can communicate at different stages of the game. In this work, we mainly focus on settings where players have an opportunity to discuss and agree on tactics before the game starts, but will be unable to communicate during the game (i.e., preplay communication). Consider, as an illustration, the case of a poker game where multiple players are colluding against an identified target player. Colluders can agree on shared tactics before the beginning of the game but are not allowed any explicit communication while playing. In other settings, players might be forced to cooperate by the nature of the interaction itself. This is the case, for instance, in Bridge. Preplay players' coordination introduces new challenges with respect to the case in which agents make decisions individually, as understanding how to coordinate before the beginning of the game requires reasoning over the entire game tree. It is easy to see that this causes an exponential blowup in the agents' action spaces and, therefore, even relatively small game instances are usually deemed intractable in this setting. When modeling preplay communication, it is instructive to introduce an additional agent, called the mediator, that does not take part in the game but may send signals (usually actions recommendations) to other players just before the beginning of the game. We explore different forms of preplay coordination in sequential games. The scenarios we consider can be classified through the following questions: i) who is receiving the mediator's recommendations? ii) do players have similar goals? iii) is the mediator self-interested? Moreover, does the mediator have more information on the state of the game than other players?

First, we study games comprising a team of agents, where a team is defined as a set of agents sharing the same objectives. Team members coordinate their actions against an opponent. Even without communication during the game, the planning phase gives the team members an advantage. For instance, the team members could skew their strategies to use specific actions to signal about their state (for example, in card-playing games, the current hands they are holding). In other words, by having agreed on each member's planned reaction under any possible circumstance of the game, information can be silently propagated in the clear, by simply observing public information. In this part of the thesis, we define the appropriate solution concepts for this setting, characterize them from a computational perspective, and provide several algorithms to compute them. Interestingly, we show that preplay communication is often enough to reach near-optimal performances (i.e., team members' benefits from communicating during the game may be negligible).

Then, we consider the case in which agents who are receiving the mediator's recommendations may not share the same objectives. In this case, the mediator has to incentivize each player to follow moves recommendations. Therefore, the appropriate solution concepts for this setting belong to the family of correlated equilibria. We show that the problem of computing and approximating a standard correlated equilibrium is mostly intractable in the extensive-form domain. On the other hand, in some settings, we devise efficient polynomial-time algorithms to compute coarse correlated equilibria. In this part of the thesis, we assume the mediator to be benevolent, i.e., she aims at maximizing the expected social welfare of the game.

Finally, we study the case in which the mediator is self-interested, and may exploit asymmetries in the availability of information to design a signaling scheme, in order to persuade players to select favorable actions. In this setting, the mediator is looking for a way to coordinate the individual behavior of each player in order to reach a preferred outcome of the game. In order to describe this setting, we provide an extension of the Bayesian persuasion framework via the definition of a new notion of persuasion, which we call ex ante persuasion. We show that, in some settings, an optimal ex ante persuasive signaling scheme can be computed in polynomial time, and we characterize the features that make the problem NP-hard.

The thesis explores the effects of preplay communication in the three aforementioned scenarios from a practical and theoretical perspective. We study the advantages of coordinating strategies before the beginning of the game, the complexity of such a problem, and devise algorithms that are theoretically sound as well as applicable in practice.

COMPILER-ASSISTED DYNAMIC PRECISION TUNING

Stefano Cherubin - Advisor: Prof. Giovanni Agosta

Given the current technology, approximating real numbers with finite-precision is unavoidable. Determining which finite-precision representation to exploit for each variable in the program is a difficult task. To face this problem, several precision mix solutions have been proposed so far in the state-of-the-art. However, the best precision mix configuration may vary at runtime along with input data.
In this thesis we aim at suggesting two effective approaches to solve the precision tuning problem.
Precision tuning can be performed either statically or dynamically. Static precision tuning produces a single mixed precision version that runs for every possible input data. Dynamic precision tuning adapts the mixed precision version to varying runtime conditions such as resource availability and input data.

Static mixed precision tuning is performed once for every application and it is considered part of the system design. Indeed, such precision tuning tools can be part of wider hardware/software co-design environments. The base component of every mixed precision tool is a set of data type representations. The IEEE-754 standard offers a selection of widely supported floating point representations. However, fixed point representations are usually implemented either as signed or as unsigned integers representations. Once a floating point or fixed point representation has been consolidated, there is the need to consistently change the data type in the application. This code conversion can be performed on the source code of the application, on the binary machine code, or at an intermediate level within the compiler. Different variables in the code may require to be converted to different data types. The analysis of the application that has to be tuned usually involves a profiling phase to empirically measure the range of possible values that each variable can assume at runtime. This profiling phase needs to run a sufficiently large input set to cover all the relevant branches in the application. The complexity of this task scales up exponentially with the number of variables that have to be tracked, and with the number of control-flow statements in the application.
Thus, this analysis requires a large amount of time for real-world applications.

Our proposed solution takes advantage of the programmers' application domain knowledge of the nature of the processed values. In particular, we rely on source code annotations written by the programmer (we consider as input a valid C/C++ source file) to know which variables should be converted to fixed-point. Then, we perform a value range propagation analysis to propagate the value range information from annotated variables along data-dependence chains, thus inferring the value range for each variable involved in the computation. The output of this analysis is a fully annotated code having the dynamic range of each variable annotated in their declaration.
To perform the value range propagation analysis, we re-purposed the GeCoS framework, which was originally designed as an hardware/software codesign environment.
In particular, the ID.Fix plugin for GeCoS is the component that tracks the dynamic range of the annotated variables via an automatic instrumentation of the code. We modified the data type allocation to enforce the use of a data width multiple of the word size for the specific architecture.
From the value ranges, it is then possible to compute the number of bits needed for the representation of the integer part of the fixed-point representation. The width of the fractional part is then obtained as the difference between the architectural constraint on the total bit size and the size of the integer part.
The GeCoS source-to-source compiler takes then care of replacing the annotated floating-point variables with their fixed-point equivalent. It also adds to the original source code the utility functions to perform data type conversions from floating-point to fixed-point and vice versa.
The output of this stage is a new version of the kernel source code exploiting fixed-point computation instead of floating-point computation. The fixed-point code can then be compiled using any compiler compliant with the C++11 standard. We develop a C++ library that defines a template type FixedPoint<integer_bits,fractional_bits> with operators properly defined to make its use convenient.

Dynamic mixed precision tuning is a recurring task that gets invoked multiple times while the application is running. In this case, the precision tuning is considered as a possible code transformation that performs continuous program optimization. While the selection of the set of supported floating and fixed point number representations is not different from the static precision tuning, in this case the code conversion tool either supports the dynamic generation of a mixed precision version, or it should rely on ahead-of-time compilation techniques. Thus, the structure of dynamic precision tuning introduces a Dynamic Integration component that is not present in the static case. Another difference with respect to the static precision tuning lies in the Tuning Policy: this software component decides which mixed precision version should be used at runtime and when to dynamically generate a new mixed precision version.
The profiling phase of the analysis of the application is performed with a reduced input test set, or it is replaced by heuristics or static analysis to further reduce the overhead.
Similarly to the analysis, the verification phase is also considered as part of the overhead resulting from the generation of a new mixed precision version. Therefore, different verification techniques are recommended with respect to the static precision tuning.

Dynamic precision tuning entails the reconfiguration at runtime of the main application itself. The reconfiguration happens every time the runtime conditions significantly vary. Tuning the threshold on the runtime conditions triggering the reconfiguration is a task that depends on the reconfiguration time and on the benefits that the mixed precision version brings.
We obviously want to maximize the benefits of the mixed precision version. However, dynamic precision tuning adds reconfiguration time in the equation, and this overhead needs to be minimized. Indeed, we prefer to exploit a suboptimal mixed precision version whenever its setup is significantly faster with respect to the optimal mixed precision version. The need for a quick reconfiguration time changes the structure of the precision tuning toolchain originally employed for static precision tuning. In particular, the profiling phase should be replaced by a static analysis, the source-to-source compilation stage should be replaced by a compiler-level transformation, and the validation can be performed by a static estimation of the error bounds rather than by a reference-based error verification.
In the dynamic precision tuning structure we also introduce a dynamic integration component, which has no corresponding equivalent in the static precision tuning structure.
This component is the joint point between the precision tuning process and the continuous program optimization paradigm. The main application should be able to reconfigure itself multiple times via the dynamic integration system.
We therefore propose a solution based on the libVersioningCompiler dynamic compilation library, on the TAFFO Tuning Assistant for Floating point to Fixed point Optimization, and on application-specific tuning policies.

We evaluate the static and the dynamic precision tuning solutions on a set of high performance computing and approximate computing benchmarks.
We show how the dynamic precision tuning toolchain can be used also for static precision tuning.
Our second toolchain is capable of achieving good results in terms of performance gain while maintaining acceptable precision loss threshold.

TIME-GATED SINGLE PHOTON COUNTING CMOS CIRCUITS

Enrico Conca - Supervisor: Prof. Franco Zappa

Silicon SPADs (Single Photon Avalanche Diodes) offer high sensitivity and sub-ns temporal resolution to detect visible and near-infrared photons, coupled with low noise, room temperature operation, and high ruggedness. CMOS SPADs add the benefit of enabling the integration on the same chip of readout and processing electronics, which allow the deployment of high-performance, low-cost and compact systems. Moreover, SPADs provide the ability of time-filtering incoming photons, by simply enabling detection only during well-defined time slots. This time-gated detection can be performed by acting on the detector's reverse bias and OFF-to-ON transitions shorter than 1 ns can be achieved. However, up to now such "fast" gating was limited to just single-pixels or small SPAD arrays, thus preventing its widespread adoption.
Main goal of my thesis was to develop large CMOS SPAD arrays with sub-ns time-gating capabilities, to be employed in imaging diagnostics and Non-Line-Of-Sight 3D ranging, as well as in many other advanced applications such as quantum photon correlation measurements.
The first outcome of my research is a large-area fast-gated digital Silicon PhotoMultiplier (dSiPM), developed within the framework of the H2020 "SOLUS: Smart OpticaL and UltraSound diagnostic of breast cancer" research project, coordinated by the Physics Dept. of Politecnico di Milano. The project goal was to develop a novel multi-modal imaging system combining ultrasonography and Time-Domain Near Infrared Spectroscopy (TD-NIRS), to improve the specificity in breast lesion classification with special focus on improving the discrimination of borderline lesions, between benign and malignant cases (BI-RADS 3 vs. 4a), that presently undergo screening evaluation with high false-positive rate. Thanks to the SPAD enabling technology, this project shrunk the detection system from a rack-mounted unit down to a detection-head of few cm^3. This was made possible by combining detector and processing electronics in a single CMOS chip and introducing the gating capability. I developed the chip in a 0.35 µm CMOS technology, chosen for the low detector noise, and I devised a novel frontend electronics that drastically reduces power consumption with respect to previous implementations of fast-gated frontend circuits. This allowed to integrate more than 3,000 fast-gated SPADs, for a combined photo-sensitive area in excess of 8 mm^2, together with a Time-to-Digital Converter (TDC) with 78 ps resolution and on-chip histogram builder in a chip sized 6 × 7 mm^2 that interfaces directly with a microcontroller through a simple SPI serial interface for programming and data readout. The chip characterization and validation lead to a temporal resolution better than 500 ps (FWHM) and >50 dB dynamic range for fast-gated measurements, meeting the specifications required for the SOLUS project. The chip was successfully integrated in a 3 cm^3 "optode", i.e. a standalone TD-NIRS detection channel that includes 8 pulsed laser diodes and the detector. The complete diagnostic instrumentation is now being finalized at the time of writing this abstract.
A second topic of this research regards the design of a high-performance fast-gated SPAD frontend in a 0.16 µm BCD technology, designed within the DARPA-funded "REVEAL" project, aiming at the development of a new state-of-art instrument for Non-Line-Of-Sight 3D ranging. This emerging area of research requires cutting-edge temporal resolution and time-gating performance, which at the beginning of the program was only attainable by single-pixel detectors and external timing electronics, resulting in extremely bulky and expensive systems. One of the goals of the project was to make the system more portable so to be exploited in actual case-studies, by designing a new time-resolved SPAD camera with tailored performance. The circuit I designed achieves gate window transitions faster than 0.5 ns through a low-threshold readout. It has been integrated in a 16 × 16 SPAD array with built-in picosecond TDCs that will enable video-rate high-resolution NLOS scene reconstruction for the first time.
Finally, I developed three general-purpose time-gated SPAD imagers in a 0.35 µm CMOS technology, all based on a refined smart pixel architecture to significantly improve resolution, range, linearity and real-case usability of high-end time-resolved Single-Photon imagers. The chips have 8 × 8, 128 × 1 and 32 × 32 pixels, respectively, and achieve very high measurement duty-cycle and < 500 ps (FWHM) temporal resolution with best-in-class noise performance. A general-purpose camera based on these imagers and a commercial FPGA board was developed and the unique capabilities allowed its use in challenging entangled photons measurements within the "Q-MIC: Quantum-enhanced on-chip interference microscopy" H2020 project, which aims at developing a super-sensitive microscope.

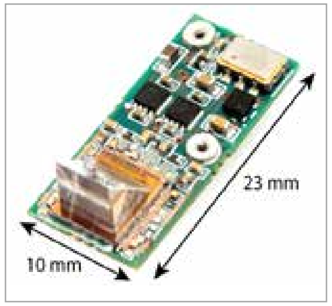

Fig. 1 - Assembled detector board of the SOLUS optode. A prism is glued on top of the detector IC to redirect the light onto the SPAD array.

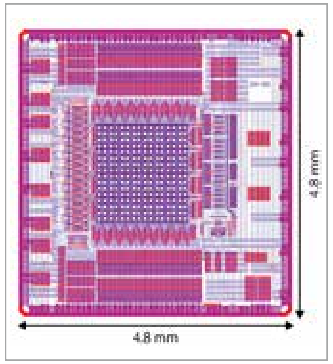

Fig. 2 - Layout of the 16 × 16 SPAD array for Non-Line-Of-Sight imaging.

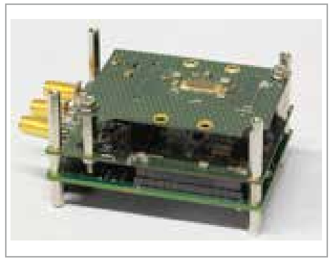

Fig. 3 - General purpose SPAD camera assembly (without housing). The topmost board hosts the detector, the middle board provides all power and interfaces, while on the bottom one is an FPGA board.

CONTEXT-AWARE RESOURCE MANAGEMENT AND NETWORK DESIGN IN 5G MILLIMETER WAVE ACCESS NETWORKS

Francesco Devoti - Supervisor: Prof. Antonio Capone

The exploitation of mm-wave technologies is one of the key enablers for 5G mobile radio access networks, potentially providing several-GHz bandwidths. However, their introduction to cellular networks poses several challenges. The harsh propagation conditions limit the mm-wave access availability and make necessary the exploitation of high-gain antenna systems in order to overcome the high path loss and limited power. Consequently, highly directional transmissions must be used with a significant increase in system complexity. A further issue with mm-waves is related to their weak penetration ability, which makes every obstacle opaque to the mm-wave propagation, and thus a potential cause of link blockage. Therefore, to fully unleash mm-wave great potential and meet the stringent 5G requirements, we need novel network design strategies as well as new ways of dealing with legacy network functionalities to provide a fast and reliable mm-wave access. For guaranteeing a reliable signaling channel, legacy access technologies and a control and user plane (C-/U-plane) functional split need to be employed. This provides the opportunity to collect context information from the users that can support network operations management. We leverage the context information related to user positions to improve the directional cell discovery process, which is one of the most critical network operations in mm-wave access as it can cause non-negligible latency if not properly managed. We investigate the fundamental trade-offs of this process and the effects of the context information accuracy on the overall system performance. We also cope with obstacle obstructions in the cell area and propose an approach based on a geo-located context database where past system history is stored to guide future searches. Moreover, we investigate the coordination problem of multiple mm-wave base stations that jointly process user access requests. Analytic models and numerical results are provided to validate proposed strategies.
The availability of a rich and accurate context is fundamental to effectively drive network operations and overcome mm-wave propagation weaknesses. Nevertheless, the mm-wave sensibility to the propagation environment can be exploited to enrich the context.
To prove this concept, we present a passive human detection system that passively monitors indoor environments leveraging beamforming alignment procedures on already deployed indoor mm-wave communication systems, detecting and locating persons with high accuracy. We implemented our system in commercial off-the-shelf devices and deployed in an office environment for validation purposes.
A widely adopted technique to guarantee mm-wave service reliability is to establish multiple connections from mobile to different base stations. Smart base-station selection must be made to minimize simultaneous blockage probability and maximize multi-connectivity effectiveness. However, the cell selection process is constrained by the network topology. The traditional approach to provide multi-connectivity is based on k-coverage planning. However, it is not guaranteeing reliable connection alternatives. Therefore, we propose a novel mm-wave access network planning framework specifically aimed at producing blockage robust network layout and ensuring the required QoS. Our framework provides the desired k-coverage while reducing the simultaneous blockage probability. The results show that our approach provides much better connection alternatives than traditional k-coverage against obstacle blockages.
To improve the coverage and throughput of the mm-wave access network, the Integrated Access and Backhauling (IAB) paradigm is under standardization. IAB allows to dynamically use the large bandwidth available at mm-wave by tightly scheduling access and backhaul links. Within this framework, we investigate the operational problems of mm-wave multi-hop backhaul networks and propose a MILP model to address the joint optimization of both traffic routing and transmissions scheduling, according to the IAB paradigm. The model captures several technological aspects proper of mm-wave hardware and includes power allocation strategies and rate adaptation. The model is followed by a thorough numerical evaluation where the performance of the multi-hop IAB network is compared against the one of a single base station, showing the potential of IAB architecture.

RESOURCE ALLOCATION FOR MOBILE NETWORKS

Osama Mohamed Mostafa Elghary - Supervisor: Prof. Luca Reggiani

The radio resource management plays a vital role in the mobile communication systems; its objective is to achieve the quality of service (QoS) requirements of a set of applications run by the system users. Through resource allocation, we can fulfill the data rate, delay and energy consumption requirements for the users. A smart allocation of resources allows us to perform this goal in an efficient and effective way.

The focus of this thesis concerns scheduling and power allocation. Power allocation affects the received SINR of the users, through a better signal level and the generated interference. By means of the scheduling we can distribute the radio resources amongst the users in an efficient way, achieving given performance targets with the advantage of approaches as multi-user diversity. Throughout this thesis, we explore different types of resource allocation techniques and observe their impact on different performance metrics.

At the beginning of the thesis we study the impact of users' location and the effect of the users' rate limit on the performance, either through scheduling or through power allocation. In the Hybrid Power Control, characterized by different parameters for cell center and edge users, we have seen that we can have a reduction in transmission power, and an improvement in the average rate at the cell edge for addressing the fairness provided by the allocation system.

Secondly, we have present a study on the modifications that can be inserted in the MAPEL, a known global optimization technique for power allocation, in order to reduce the computational complexity; it is shown that the optimization time can be reduced of a factor close to 20, while keeping the accuracy of the results very high. Moreover, it has been included a feature concerning the respect of the minimum rate and total power constraints, allowing it to be used more realistically in various cellular applications, such as the narrow band internet of things (NB-IoT) application, which is presented in the thesis.

With respect to the NB-IoT application, we study the new resource unit configurations, analyze the optimal problem formulation and design a sub-optimal, computationally feasible allocator that can be used as a benchmark close to the optimal solution, considering the users' achieved rate and the reduction of the number of dropped packets. Furthermore, we see that by incorporating the modified MAPEL in the algorithm, the throughput can be approximately doubled, and the power consumption reduced by 3 dB. Moreover, we study the repetition triggering techniques and we decrease the number of dropped packets from 29 % to 1.6 %.

The last part of the thesis has been dedicated to two resource allocation problems for D2D communication. In a decentralized communication of D2D devices without base station support, we have verified the limits of the analytical approximation of the total interference for D2D communications based on the Fenton-Wilkinson method and we have shown, by means of simulative results, the potential advantage of a very simple power allocation technique to be used when the transmissions are broadcasted by some reference nodes, like in case of proximity services or peer discovery search.

SELF-MANAGEMENT OF GEOGRAPHICALLY DISTRIBUTED INFRASTRUCTURES AND SERVICES

Danilo Filgueira Mendonça - Supervisor: Prof. Luciano Baresi

Introduction

The advent of *mobile computing* and the *Internet of Things* (IoT) poses new challenges to the cloud data center model. Data-intensive applications hosted by devices at the network edge are already pushing the boundaries of how much data needs to be transported and processed by cloud data centers. Current estimations point out the continuous increase in this trend. According to Cisco, the total amount of data created will reach 847 zettabytes (1021) per year by 2021. The number and diversity of applications with compute-intensive features are also increasing, which partially nullify the gains in computing power and battery life in modern devices. These applications need to offload computation to more powerful servers in the cloud (a.k.a. *Mobile Cloud Computing*). However, applications with strict latency and throughput requirements raise further questions about the feasibility of the data center model in satisfying all these needs.

The paradigm of *edge computing* aims to tackle the challenges above by filling the gap between cloud data centers and data *prosumers* at the network edge. Edge computing has been associated with different terminologies, architectures, and technologies, all of which share the common goal of enabling new types and scales of latency-sensitive and data-intensive applications. Some of the prominent use cases targeted by edge computing are consumer-driven. In this category are disruptive applications such as Mobile Augmented Reality, Cognitive Assistance, Connected and Autonomous Vehicles, Industry 4.0. Network operators and third-party are also expected to harness edge computing to deliver new types of services.

In this thesis, we distinguish two main use case categories: *latency-sensitive computation offloading* and *data pre-processing*. As the name suggests, the former considers latency as a first-class requirement. Its primary goal is to augment the capabilities of mobile and IoT devices. In contrast, the second scenario is mainly concerned with the anticipation of the processing of voluminous data near its source, often IoT devices.

The satisfaction of edge computing use cases created research opportunities in multiple areas. Many authors dedicated attention to the deployment of edge resources, often within cellular infrastructures. Other authors focused on the decision of if and when to offload tasks to edge servers, while others tackled management aspects such as the dynamic deployment and scaling of services. The dynamic placement and migration of services according to the mobility of users also received considerable attention from the optimization community.

Problem Statement

From the software architecture perspective, the particularities of the decentralized infrastructure model envisioned had been often ignored. The same is true for the service model to be adopted by edge providers. For instance, some authors adopt typical cloud service models in which virtual machines hosting monolithic applications are provisioned on demand. The dense distribution of servers entails additional challenges regarding current service models and architectures. Thus, a more precise and adequate definition of the service model to be used by operators, and the architectural style of consumer applications is needed.

The paradigm of *serverless computing* has been conceived to allow developers to focus on the core aspects of their application, while a third-party provider handles the management of the required infrastructure. The serverless paradigm has evolved in different forms and more prominently as computing services, a.k.a. *Functions-as-a-Service*. In the context of edge computing, the previous model has critical advantages over previous cloud service models. First, functions are lightweight by design and run in stateless containers. Thus, they can be scaled and managed more efficiently. Also importantly, the delegation of infrastructure management to the provider prevents conflicting decisions and facilitates the configuration of individual servers and the deployment of software components across a potentially large number of servers.

Despite its potential, serverless computing is still young if compared to other consolidated paradigms and models. Its adoption in the fulfillment of our target use cases needs further assessment: the benefits and limitations of modeling, implementing and deploying application logic as functions must be evaluated; in a similar direction, key performance indicators such as resource allocation efficiency, latency overhead and scalability of existing serverless technology must be assessed and optimized to fulfill the requirements at hand.

The management of geo-distributed infrastructures and services also entails multiple challenges. In a typical cloud computing deployment, users define the policies that will govern the orchestration of resources. Such a high autonomy level is possible, thanks to the virtually unlimited resources that characterize nowadays' cloud data centers. The decentralized nature of edge computing adds a spatial dimension to the resource orchestration problem. Not only the allocation of resources needs to be dynamically decided, but also the placement of application components onto the mesh of geo-distributed servers. User-centric orchestration decisions are likely to conflict with each other unless arbitrated by the provider.

Contributions

We investigated and compared different service models and virtualization technologies for the provisioning of computing services by edge infrastructures. We also evaluated key aspects regarding the architectural style of consumer applications. Based on our findings, we proposed a *Serverless Architecture for Multi-Access Edge Computing*. The proposed architecture was evaluated both qualitatively and quantitatively through relevant application scenarios and experiments. Once the benefits and limitations of the proposed architecture had been demonstrated, we expanded the landscape of our contributions with a broader range of deployment configurations resulting from the combination of mobile, edge, and cloud computing. The resulting *Mobile-Edge-Cloud Continuum* extends the Serverless MEC Architecture.

To manage the life-cycle of serverless functions deployed to the heterogeneous platforms comprising the Continuum, we proposed *A3-E framework*. A3-E exploits both serverless and autonomic computing to allow serverless functions to be opportunistically and autonomously fetched, deployed and exposed as services by the distinct edge and cloud providers, or consumed locally. A3-E prototypes were implemented and evaluated. Results demonstrated significant gains in terms of response time and battery consumption, which are critical requirements for the application scenarios targeted by this thesis.

The opportunistic nature of A3-E yields an efficient allocation of resources. Nonetheless, A3-E does not tackle the horizontal collaboration among edge nodes from a single edge provider. We address this limitation with the *PAPS framework*. PAPS is based on a multi-level, decentralized self-management approach that partitions the larger scale edge topology into delay-aware communities and allocates resources among and within communities. The PAPS framework combines optimal allocation and placement at the community level with a fast node-level container scaling to render the overall self-management process both efficient and effective. Our framework also comprises a simulator platform, which enables the evaluation of different scenarios. The PAPS framework was evaluated with a large scale edge topology hosting up to 100 serverless functions. Results showed the criticality of the node-level self-management to achieve a more efficient allocation of resources and to prevent SLA violations.

ARTIFICIAL INTELLIGENCE AND AUGMENTED REALITY FOR ENTERTAINMENT APPLICATIONS

Darian Frajberg - Supervisor: Prof. Piero Fraternali

Outdoor Mobile Augmented Reality (MAR) applications promote a new way of marketing the touristic offer of a territory, by overlaying useful information directly on top of the user's camera view. These applications aim at exploiting the sensor readings (GPS position, phone orientation and possibly also the camera) of the device to understand what the user is looking at and enrich the view with contextual information to enable knowledge acquisition and exploration. Furthermore, Deep Learning (DL) has recently exhibited superior performance in a wide variety of Computer Vision (CV) tasks and can lead to novel AR solutions. The development of outdoor mobile AR apps poses several technical challenges: 1) they must understand the current context and activity of the user; 2) they must find the appropriate information pertinent to the user's current activity and view; 3) they must overlay the retrieved information onto the device screen in a way that is adequate to the user's experience. Besides, also non functional requirements are relevant: the application must work in real-time on devices with limited computational power and strict energy consumption constraints, and functioning should ideally be guaranteed also in absence of Internet connectivity. Taken together, all such requirements make the development of outdoor mobile AR apps a non trivial task.

The goal of this thesis was to exploit the commoditization of Artificial Intelligence (AI) methods and the forthcoming wave of low-cost mass market AR devices, to propose methods, architectures and components to support the creation and evaluation of solutions for outdoor AR applications efficiently executable on low-power portable devices.

We explored the feasibility of exploiting the recent progress in the AI, CV and AR fields in order to provide enhanced outdoor AR entertainment/informative solutions. We concluded that such integration can lead to the construction of high quality, successful and engaging applications, suitable for limited portable devices by means of better understanding the surrounding visual context captured by the camera with a marker-less approach executed on-board the devices.

We addressed the arising difficulties and validated the proposed solutions through the use case of PeakLens (Figure 1 and Figure 2), a real-world outdoor MAR app that identifies mountain peaks and augments the view with their corresponding information. It is available for Android and has so far reached over 500k installs. This application can be conveniently used for both improving the awareness about the mountain environment and for crowdsourcing environmental tasks, such as data collection.

We introduced the use of DL models for mountain skyline extraction to deal with accurate image-to-terrain geolocalization and camera orientation estimation in natural mountain environments. Such models were trained with a large set of annotated images taken in uncontrolled conditions, and are capable of identifying obstacles interrupting the skyline with a patch-based approach. Obstacles identification represents an essential element to take into account for a correct alignment w.r.t. the terrain.

DEVICE	TF LITE (MS)	POLIMIDL (MS)
Asus Zenfone 2	807.67	179.33 (-77.80%)
Google Pixel	95.00	68.00 (-50.84%)
LG G5 SE	138.33	35.33 (-62.81%)
LG Nexus 5X	193.00	80.33 (-58.38%)
Motorola Nexus 6	225.67	66.00 (-70.75%)
One Plus 6T	68.67	22.67 (-66.99%)
	Average	**(-64.59%)**

Tab.1 - Inference time experimental results with PL Optimized model on a set of devices.

We presented a framework for the development of outdoor MAR applications to cope with several difficulties, such as unreliable connection, uncertain positioning, occlusions and real-time requirements; and defined a characterization for the dimensions of heterogeneous meta-data to augment the view. The resulting applications can feature an intelligent CV module to exploit the camera view and thus, compensate potential projection errors caused by other typical noisy sensors (e.g. GPS, accelerometer, magnetometer).

We specifically tackled the problem of efficiently supporting the optimization and deployment of DL models on-board devices with limited resources and heterogeneous architectures. We have also presented PolimiDL, an open source publicly available framework for the acceleration of DL inference on mobile and embedded systems, which has proved competitive w.r.t. TensorFlow Lite, achieving better performance on a set of small models (Table 1).

In contrast to traditional AR systems purely based on location and/or motion sensors, the proposed approach requires non-trivial, realistic and automated testing, which is also covered in this work. We presented a capture and replay framework for the automated testing of mobile applications that depend on noisy multiple correlated sensor streams, thus becoming suitable to support the development and maintenance of MAR applications under controlled lab conditions.

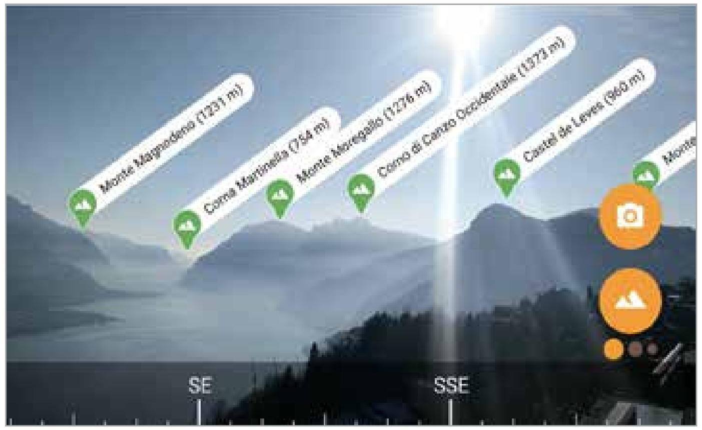

Fig. 1 - Screenshot of PeakLens app.

Fig. 2 - Examples of images augmented with PeakLens app.

ACCIDENT PREVENTION, DETECTION, AND EARLY RESPONSE: USING MACHINE LEARNING AND DATA ANALYSIS TO IMPROVE DRIVER AND PEDESTRIAN SAFETY

Simone Gelmini - Advisor: Prof. Sergio Matteo Savaresi

The benefits of personal mobility are unquestionable. However, with its high number of fatalities every year, road traffic accidents are still one of the top causes of death worldwide, with more fatalities than, *e.g.*, tuberculosis, HIV/AIDS, and malnutrition. Despite enormous efforts by the scientific community and industry to make vehicles safer, traditional technological development has proved unable to cope with this issue entirely. Studies show that human factors, rather than vehicle performance, are contributing to the numerous vehicle fatalities. More than ever, drivers are distracted by external factors and electronic devices, reducing their capability to counteract unexpected situations and increasing the likelihood of a crash. Additionally, inexperienced drivers tend to underestimate the risks they often take while driving, overestimating the vehicle capabilities to react in some limit conditions. Drivers are usually unaware of their driving behaviors, in particular how distraction and overconfidence play on their safety.

In case of an accident, emergency calls are often delayed when nobody witnesses the accident, increasing the risk of serious injury or fatality if the injured person(s) are incapable of calling for help. This is particularly critical because studies show that in 70% of all road accidents, the emergency call is made by somebody not involved in the accident. The automatic dial of an emergency call – *eCall* – in response to a road accident is a feature that is gaining interest in the intelligent vehicle community. It indirectly increases the driving safety of road vehicles, but presents the main technical challenge of developing an algorithm accurate enough to trigger the emergency call only when needed, limiting the social and economic costs of unnecessary calls. Its primary goal is to jointly minimize the access time of the emergency services and, for insurance telematics purposes, to accurately store the event dynamics for further investigation on the accident's responsibilities

Inspired by standard medical practices, this thesis presents a set of methods that aim to reduce the number of accidents, both for motorists and pedestrians, and mitigate the injuries. Overall, the proposed strategy can be broken down into three main phases: *prevention*, *prompt diagnosis*, and *immediate therapy*. First, we discuss an algorithm that promotes safe driving practices by enabling driver awareness, focusing on the driver's average driving behavior. In this context, we designed a four-dimensional driving-style assessment for safety-related purposes, promoting driving awareness. The main contribution is adding, to more common proxies of risky-driving, the dimension of smartphone usage. Phone use while driving is performed through appropriate processing of smartphone-based inertial sensors. Analyzing the motion of the smartphone for assessing its use preserves drivers' privacy more than traditional methods, which are subject to some privacy-related issues, *i.e.*, phone fingerprinting through the analysis of system API calls.

Then, in the unfortunate occurrence of a crash, we propose a method for automatically detecting it, grading its severity, and calling for rescue in case the person is injured or needs immediate assistance. More particularly, we investigated the automatic detection of motorcycle accidents. In the case of two-wheeled vehicles, this task is even more challenging due to their complex dynamics, where the measured signals are correlated in a non-common and non-trivial way. We propose two methods which aim to detect crash-like situations, searching for anomalies in the data time series. The presented strategies are less sensitive to driving-style than traditional algorithms, reducing the number of incorrect detections.

Following the same approach, we present a similar strategy for pedestrians, in which we detect when the subject loses consciousness, becoming incapable of calling for rescue. This condition make the rescue process particularly challenging, especially when the location of lone pedestrians is unknown or not easily accessible. In this circumstance, time for rescue increases, exposing the vulnerable person to several secondary health-related issues. Additionally, for some specific anomaly events, we provide a methodology that deploys an airbag cushion placed inside a garment, aiming to mitigate the injuries severity. Due to the significant mass difference and the lack of protective structures, pedestrians are some of the most vulnerable road users, especially during collisions against motor vehicles. For this reason, the proposed strategy is one of the fastest ways to provide immediate therapy during an accident.

The algorithm design was challenged also by a major technological challenge. Every year, more than fifty million vehicles are used in Italy. According to the registration year, the range of vehicles spreads from the most recent to some old ones registered back in the 60s. The sensing and computational setups are very different according to the different generations. Furthermore, there has always been a lack of standardization among different makers, resulting in a difficult (or even impossible) common strategy given such heterogeneous layouts. However, thanks to the increasing availability of smart devices, we have chosen to leverage machine learning and data analysis techniques to tackle the aforementioned challenges.

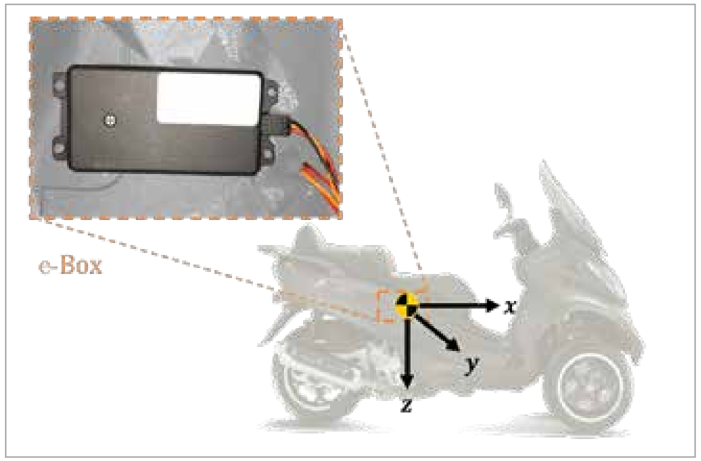

Fig. 1 - Test vehicle equipped with an e-Box

Special attention has been devoted to designing a flexible, methodical sounding algorithmic structure, whose outcome could be easily interpreted and the overall system adapted to different problems. The proposed framework has also been designed considering the limited resources of telematics e-Boxes (Fig. 1), smart devices already used by a large number of drivers due to widespread use and promotion by automotive insurance companies. All the proposed methods have been validated against data collected during dedicated experimental campaigns, involving both regular drivers, riders, and pedestrians, but also professional riders on high-speed track tests and professional stuntmen for characterizing the most dangerous dynamics.

HIGH-RESOLUTION DIRECT-WRITTEN FIELD-EFFECT TRANSISTORS FOR HIGH FREQUENCY APPLICATIONS

Michele Giorgio - Supervisor: Prof. Marco Sampietro

Organic electronics is gaining more and more attention from scientific public and from industries because of his unique properties such as flexibility, transparency, and solution-processability. High throughput manufacturing methods such as roll-to-roll coating and inkjet printing allow organic electronics to be suitable for new and diverse applications and to be an attractive viable way to make low-cost electronics. Before employing this technology for the production of complex circuits, the performance of the elementary block, i.e. the transistor, has to be optimized. The willingness to fabricate transistors capable of high-frequency operation is spurred by possible applications like high-resolution flexible displays or devices able to communicate via wireless. So far, a record frequency of transition of 27 MHz is achieved for transistors with lithographic contacts and evaporated semiconductor, while of 20 MHz for transistors fabricated without using masks in the production flow. Such record frequencies of operation were not measured directly but they were extrapolated due to bandwidth limitation, typically in the order of 1 MHz, present in the measurement setups used. Thanks to recent improvements in polymers charge carrier mobility, further enhancement of devices maximum operational frequency is in sight, and therefore it is not possible to rely further on such measurement methods for characterizing the upcoming organic high-frequency devices.

In this thesis work, we resorted to Scattering parameters (S-parameters) for reliable determination of the frequency of transition without extrapolations. We report n-type FETs based on a solution-processed polymer semiconductor where the critical features have been realized by a large-area compatible direct-writing technique, allowing to obtain record frequencies of transition of 19 MHz in the case of n-type and 24 MHz for p-type polymer transistors. This is the first report of solution-processed organic FETs characterized with S-Parameters. We then made a step forward in the improvement of the OFETs maximum operational frequency by reducing the gate to source and gate to drain geometrical overlap to some hundreds of nanometers by fabricating also the gate contact by laser sintering. In this way we succeeded in the demonstration of the fastest n- and p-type organic transistors ever realized, showing transition frequency of 70 MHz and 100 MHz, respectively. Finally, in order to fabricate high-performance organic circuitry, we propose a method for self-assembled monolayers (SAMs) formation by using inkjet printing techniques, allowing to grow different SAMs on the desired contacts on the same substrate. We demonstrated both n and p unipolarized devices starting from a single ambipolar semiconductor uniformly bar-coated on the substrate, with mobilities higher than 1 cm^2/Vs in both cases.

OPTIMAL DESIGN OF OFF-GRID WATER-ENERGY SYSTEMS IN SMALL MEDITERRANEAN ISLANDS

Federico Giudici - Supervisor: Prof. Andrea Castelletti

Co-supervisor: Elisabetta Garofalo

Small Mediterranean islands represent a paradigmatic example of remote, off-grid systems facing a large number of sustainability issues, mainly due to their distance from the mainland, the lack of accessible water sources, and the high seasonal variability of both water and electricity demand. Energy security is generally reliant on carbon intensive diesel generators, which are usually oversized to meet peak summer electricity demand driven by high touristic fluxes. Potable water is often produced by energy intensive desalination technologies, which strongly impact on the electricity system, increasing air pollution and greenhouse gas emissions.

In order to improve the economic and environmental sustainability of small islands, the design of hybrid energy systems, combining traditional power generation with renewable energy sources and storage technologies, represents a viable and promising solution. However, traditional methods for designing such systems usually neglect important aspects and challenges. Major challenges include (i) the optimal control of the electricity system as well as its interconnection with other energy vectors (e.g., gas, heat) and domains (e.g., water system) for fully exploiting renewable power, (ii) the interdependency between system planning and its operation, (iii) the presence of multiple, potentially conflicting, objectives reflecting economic, environmental and other sustainability aspects, (iv) deep uncertainty in climate, technological and socio-economic conditions that may affect the system performance over a medium-to-long term horizon.

Driven by these challenges, this thesis contributes novel methodologies for supporting energy systems transition towards decarbonization, helping decision makers to identify viable solutions at different temporal scales in light of plausible future conditions that might unfold. In particular, we develop a set of modelling and optimization tools for optimizing both the design and the operations of off-grid hybrid energy systems, also considering the uncertainty related to future changes in the main external drivers. The proposed methodologies allow us to (i) investigate the benefits of explicitly considering the interdependency between system design and operation with respect to multiple economical, environmental and efficiency objectives, (ii) assess the vulnerability of hybrid energy systems to the future uncertainty in the main external drivers, and (iii) design solutions that are robust with respect to this uncertainty.

The first deliverable of this research is a novel multi-objective, dynamic approach for conjunctively optimizing design and operation of water-energy systems by focusing on the interconnection between electricity generation and water supply through the optimal control of desalination plant. Secondly, we propose a methodological framework to assess the vulnerability of hybrid energy systems with respect to changes in the main climate drivers (i.e., solar radiation, wind speed, temperature). More precisely, we evaluate how historical variability and future uncertainty in these climate variables affect the performance of highly renewable hybrid energy systems, designed under average historical conditions, in terms of different sustainability indicators. Finally, we focus on the challenge of directly including deep uncertainty in future climate drivers within the system design phase. Since the performance of hybrid energy systems in small Mediterranean islands strictly depends on multiple, deeply uncertain co-varying drivers, a very large number of future scenarios should be considered and, consequently, included within the optimization process for generating robust solutions, leading to very high, or even intractable, computational time for solving the problem. To address this issue, we develop ROSS (Robust Optimal Scenario Selection), a novel algorithm that uses active learning for adaptively selecting the smallest scenario subset to be included into a robust optimization process.

We test our novel approaches on the real case study of the Italian Ustica island, which represents a paradigmatic example of off-grid Mediterranean island. Main thesis outcomes show that considering the interdependency between system design and operation by dynamically modelling the nexus between water production and electricity generation allows to significantly improve system performance by reducing the structural interventions, the investment costs and the environmental impacts. In addition, results suggest that wind speed represents the climate variable that mainly influences the performance of hybrid energy systems, which will likely degrade on a medium-to-long term horizon. Finally, our novel ROSS algorithm allows to obtain robust hybrid energy system designs reducing computational requirements between 23% and 84% compared with traditional robust optimization methods, depending on the complexity of the robustness metrics considered. Moreover, it is able to identify very small regions of the scenario space containing the most informative scenarios highlighting the main system vulnerabilities.

OPTIMIZATION-BASED CONTROL OF MICROGRIDS FOR ANCILLARY SERVICES PROVISION AND ISLANDED OPERATION

Alessio La Bella - Supervisor: Prof. Riccardo Scattolini

As many research studies and industrial projects witness, a key solution to enhance the diffusion of renewable energy sources and facilitate their management consists in the spread of autonomously controlled small-scale grids, called microgrids. Microgrids can be identified as self-contained clusters of dispatchable micro-generators, storage devices, renewable sources and loads, operating as a single controllable system. A high-level schematic of a generic microgrid is depicted in Figure 1. A microgrid can operate either connected to the main utility or in islanded mode. This doctoral research aimed to design dedicated control architectures for these two microgrid operating modes, characterized by different issues and opportunities. The research has been carried out in collaboration with research centre RSE (Ricerca sul Sistema Energetico) and it is structured in two main parts. The first part focuses on the management of grid-connected microgrids providing ancillary services, i.e. supporting power services provided to the system operators in order to maintain the safe operation of the overall electrical system. Indeed, it should be considered that given the spread of non-deterministic distributed generation resources, like Renewable Energy Sources (RESs), and new loads, like electric vehicles' charging stations, it has become more and more crucial to ensure the real-time balance between generation and demand. The power balance is in fact of vital importance in electric systems, not only to guarantee the secure power supply, but also because the network frequency and voltages may seriously deviate if power imbalances are not promptly restored, leading to instability and black-out events. Because of this, the electrical system is today required to become more flexible, continuously and actively adapting generation and demand patterns of each distributed energy resource. In this context, microgrids are considered as a great opportunity to enhance this flexibility characteristic, managing their multiple and heterogeneous energy sources to provide external ancillary services to the electrical system, so that to ensure the real-time balance and the proper operation of the system. Nevertheless, ancillary services, traditionally provided by big generation plants, are characterized by large minimum power and capability requirements in order to have a significant impact on the system. Single microgrids, having commonly available a limited generation capability, can hardly meet these requirements, having also to satisfy their internal loads. Given these issues, we designed an optimization-based control framework where different interconnected microgrids can be managed as part of a unique electrical aggregation and managed by an external entity, named aggregator supervisor, potentially capable to provide external ancillary services. In this context, a parallel activity concerned the study and analysis of the European regulations on the energy markets and on the provision of ancillary services, in order to design solutions compatible with the actual electrical framework. Therefore, different algorithms have been designed for different kinds of ancillary services that aggregation of microgrids can provide, and these are described in detail in the first five chapters of the doctoral thesis. In particular, the main focus is on the provision of balancing power services from microgrids, which consists in the online adaptation of the output power profiles in order to ensure the balance between the power generation and absorption in the overall electrical system. It should be underlined that a centralized management of the microgrid aggregation is not feasible, since it would imply that the aggregator supervisor has complete access to the microgrids internal information such as generators' characteristics, renewable sources' production and loads' consumption profiles. Since microgrids are usually private facilities, this could be not a desirable aspect, leading some elements to not join aggregators in order to keep the full control and knowledge of their own resources. On the other hand, centralized approaches are not scalable, and they may lead to computational issues as the aggregator size grows. Therefore, one of the main contributions of the first part of the thesis is the design of several optimization-based control frameworks for the microgrid aggregators using distributed and hierarchical control approaches. This has the great advantage of ensuring the scalability of the approach, preserving at the same time microgrids internal information and the full control of local resources. All the designed approaches have been tested through extensive numerical simulations considering real network benchmarks provided by IEEE, showing their effectiveness in fostering the integration of microgrids, as well as their beneficial effects for the electrical system. Then, the second part of the thesis focuses on the microgrid islanded operation. This condition involves different technical issues that must be addressed. In fact, without the support of the main grid, the continuous balance between generation and absorption must be internally ensured, which can be critical due to the presence of non-deterministic renewable energy sources and loads. Moreover, the islanded condition requires the regulation of the internal frequency and voltages, since this task is no more carried out by the main utility. This aspect becomes more critical since microgrids are characterized by a low electrical inertia, given the presence of electronic power converters, and the electrical frequency and voltages are extremely sensitive to power imbalances. Therefore, the islanded operating mode requires the Microgrid Central Controller to both ensure the optimal management of resources and the low-level regulation of frequency and voltages. Since these tasks involve different time scales and system modelling, hierarchical control architectures are proposed in the second part of the thesis. These multi-layer control architectures are designed and proposed both for microgrids with standard Alternating-Current (AC) networks and for microgrids with Direct-Current (DC) networks. Indeed, DC microgrids, due to their ability to interface naturally with most renewables sources, batteries and many electronic loads (e.g. electric vehicles, LED systems) have gained a lot of attraction in recent years and they are expected to become a future standard. In both cases, Model Predictive Control approaches are adopted for the Energy Management System layer, since these techniques allows to include unit's capability constraints, economic objectives and forecasts of RESs and loads. Then, fast decentralized controllers, properly interfaced with the high-level Energy Management System, are entitled of the low-level regulation of MG frequency and voltages. Also in this case, extensive simulation tests have been carried out considering realistic electrical networks, showing the advantages of the proposed approaches in ensuring both the prompt electrical regulation and the efficient high-level energy management.

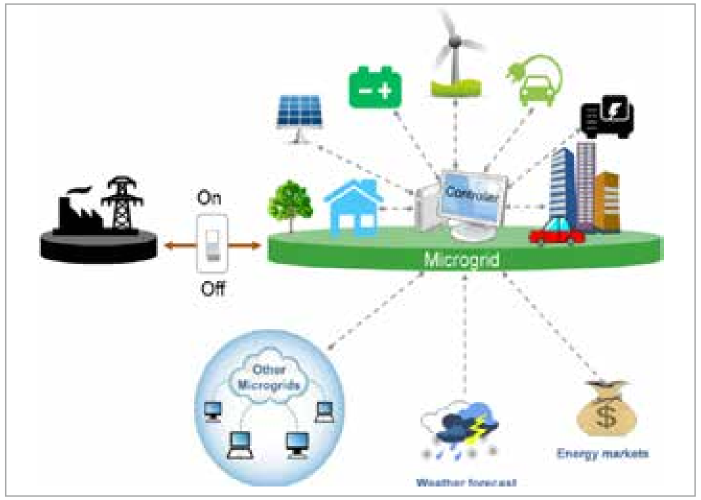

Fig. 1 - High-level schematic of a microgrid

A TOTAL ATTENUATION TIME SERIES SYNTHESIZER FOR NON-GEOSTATIONARY SATELLITE LINKS

Luciano Machado Tomaz - Supervisor: Prof. Carlo Capsoni

The development of space communications is increasingly dependent on the direct knowledge (satellite experiments) and on the output of ad-hoc developed tools (prediction methods) that enable to understand and quantify the main impairments affecting wave propagation. In recent times, the high frequency bands (Ku and Ka) are extensively used in commercial and military applications while the Q/V bands are expected to be used in the next future. It follows that the knowledge and the study of the propagation problems at such very broad range have become a critical issue. Moreover, the Low Earth Orbit (LEO) and Middle Earth Orbit (MEO) satellites, widely used in remote sensing and Earth Observation (EO) applications, are nowadays becoming more and more interesting also for television broadcasting and direct-to-user communications due to the fact of being closer to the Earth surface. These satellite configurations have shorter orbital periods with respect to GEO and lower signal propagation delays. Also, since the signal propagation paths are shorter, the requested transmission power is also smaller, which allows building small and low-cost satellites (although in a larger number).
While in the geostationary satellites (GEO) applications the position (geometrical parameters) of the link remains static and only the weather presents variability, in non-GEO (NGEO) satellites applications, in addition to the meteorological changes, it is necessary to consider that the satellite in the short orbital period is seen from the ground station at variable values of azimuth and elevation thus making the modelling task very challenging.
The thesis presents a model of a Total Attenuation Time Series Synthesizer for NGEO satellite links. The development of time series synthesizers allows the simulation of the time varying condition of the channel and represents the today most advanced propagation prediction tool. In fact, a time series synthesizer is the basic instrument for the development and testing of the Propagation Impairment Mitigation Techniques (PIMTs), which represent the best approach for an effective design of a satellite communication system operating at microwave frequencies.
The philosophical approach at the basis of the total attenuation time series synthesizer for non-geostationary satellites we are going to present in this PhD work is the development of a model that maintains as much as possible a physical sound basis. The modelling procedure relied on data obtained by simulations using field maps generated with MultiEXCELL (rain), SMOV (vapor), and SMOC (clouds) models, as well as on the long-term ITALSAT database. The use of these databases of measurements (even in a modified form) provides a realistic approach to the modelling with a strong physical basis.
First, we presented a procedure for scaling the long-term statistics of rain attenuation down to very low elevation angles. And Since in the NGEO satellite systems, together with the elevation changes, there are also azimuthal changes, we proposed a model to account for the azimuthal variations. The rain attenuation time series synthesizer is then completed with the generation of an azimuth window that considers the effects of these variations during each passage of the satellite.
We present also a study on the feasibility of the cosecant scaling applied to the water vapor attenuation and, since in comparison to the rain structures, the water vapor results well uniform, we decided to adopt only the geometrical scaling for water vapor. For the cloud's contribution, in turn, we derived a scaling factor for scaling up the elevation with a procedure analogous to the one used to rain.
Since there are no direct and statistically meaningful measurements for NGEO satellite links and aiming to test our model, we performed simulations of a LEO satellite by using MultiEXCELL maps.

The sites to which the simulations were performed are Spino d'Adda and Miami, assuming as minimum elevation 5º. In order to reduce the enormous amount of time required for a complete simulation we used the data of one passage of the satellite. The simulations were performed at the carrier frequency of 26.5 GHz that is indicated to be the one for Earth Observation Data Down Link (EODDL).
We have chosen one geosynchronous orbit of the satellite MetOp-A in order to test the model. Figure 1 below shows the CCDF of the rain attenuation time series component obtained with the proposed attenuation time series synthesizers for the selected satellite at Spino d'Adda. The CCDF is normalized to 1, so the zero samples present in the time series were discarded. The result shows a good agreement with MultiEXCELL simulations. The small differences are due to accumulation of the errors present in each step of the model.

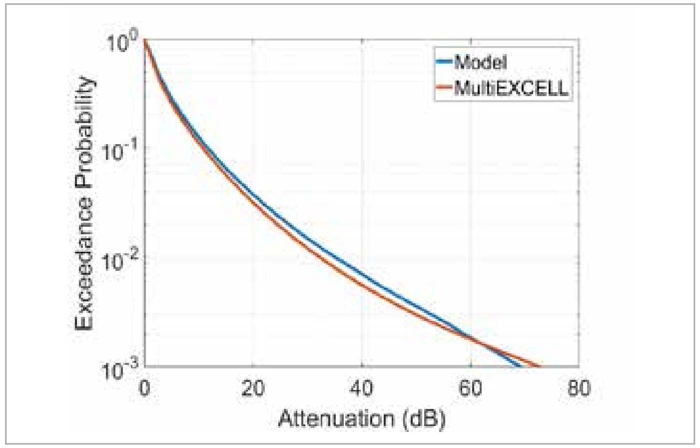

Fig. 1 - Comparison between the CCDF of rain attenuation obtained with the NGEO ATTS and the one obtained with MultiEXCELL simulations (Spino d'Adda).

FORENSICS AND COUNTER-FORENSICS METHODS FOR SOURCE DEVICE IDENTIFICATION

Sara Mandelli - Supervisor: Prof. Stefano Tubaro

Images and videos of various nature flood the web in an unbridled fashion everyday, overwhelming our social network profiles. This wildfire spreading of visual content published online can be seen as the direct consequence of a new communication paradigm, founded on immediate and effortlessly knowledge. The concept of “social” networks itself has been revolutionized during the last few years. Nowadays, they represent not only platforms for connecting people around the world but actual websites for quick information, marketing purposes and politics propaganda.
As a matter of fact, visual communication is by far the most powerful and rapid instrument to convey a message in terms of data intelligibility. Every age, group and education can access an enormous amount of data, absorb information and share new content in few seconds. This phenomenon inevitably injects potential dangers into the communication process: whenever illegal or counterfeited data are uploaded on Internet, the longer we wait, the harder is to verify the content authenticity and avoid its uncontrolled diffusion.
In this vein, performing forensics investigations on multimedia content answers the need of smart solutions for assessing data authenticity and integrity. Tampered with and neural network generated data, as well as unknown provenance and illicit material constantly fill days of forensics analysts, just to mention some examples.
In this thesis, we tackle a few of these forensic challenges, specifically focusing on source device identification problems on images and videos. Indeed, determining the origin of visual data proves extremely helpful to expose copyright violations or fight distribution of illegal content (e.g., child exploitation clips and terroristic threats). This is actually a very fine request, because it is required to estimate the precise source device, not just the corresponding camera model or vendor.
If attributing images to their correct device is a relatively known problem in forensics community, statistical methods may suffer from high computational times whenever it is required to identify the source among a huge device collection, for instance when dealing with images uploaded on social networks where there is a large pool of provenance devices to be analyzed. Moreover, when working on social networks, often only cropped and resized images are available. This greatly drains the performances of state-of-the-art methods, as these usually require investigations over a large pixel portion for returning acceptable identification accuracy. Up to now, data-driven approaches based on convolutional neural networks (CNNs) have limited their studies to camera model identification. As a matter of fact, these approaches present very accurate performances when discriminating from one model to another, but cannot identify the specific device which shot an image. Following these premises, we propose a novel and fast approach based on CNNs to identify the source camera of a query image. Specifically, we investigate two scenarios: (i) given a query image and a pool of known devices, detect which device shot the image; (ii) given a query image and a single device, detect if the device shot the image. We train a CNN able to distinguish the subtle camera traces left on images and match them with the correct source camera. Our method proves to be faster than statistical methods in case a large amount of potential provenance devices is investigated. Moreover, it requires less query image content to obtain comparable attribution accuracies.
Extending image source identification strategies to videos is far from being straightforward, and several peculiar issues must be addressed. First of all, videos are typically acquired at lower resolution and stronger compression than images. Further issues have been rising for the last five years, when mid-price smartphones have begun including video stabilization mechanisms, following the trend started by the most expensive ones. Such mechanisms strongly hinder the performances of standard forensics techniques for device identification since they introduce subtle misalignment between the frames. In light of this, we thoroughly investigate the stabilization mechanism and propose multiple strategies for dealing with source device identification on stabilized sequences. Specifically, we propose: (i) two strategies to extract the characteristic fingerprint of a device, starting from either a set of images or stabilized video sequences; (ii) three diverse strategies to match a stabilized video sequence with a given camera. The proposed methodology is tested on videos coming from a set of different smartphones, taken from a publicly available dataset. The conducted experiments also provide an interesting insight on the effect of modern smartphones video stabilization algorithms on specific video frames.
Finally, and just as important, forensics analyst can evaluate the effectiveness of the adopted strategies by tackling problems from a counter-forensics perspective. Indeed, studying the boundaries of image anonymization can enable analysts to be aware of the robustness of camera attribution methods in the presence of malicious attacks. In addition, when privacy is a concern, it would be desirable to anonymize photos, unlinking them from their specific device. Photojournalists in states at war, human right defenders and activists are a few examples in which preserving the privacy of data owner may be required. In view of this, we draw two solutions for the source device anonymization task on images, specifically: (i) a strategy that deletes some pixels from the image and reconstruct them by means of inpainting described as an inverse regularized problem; (ii) a data-driven strategy exploiting a CNN which edits images in a visually imperceptible way. Both the two proposed methods aim at hindering the device attribution problem, removing the traces of the source device from the image under analysis.
To face all these challenges, we rely on subtle sensor traces left by each camera on its acquired content. These characteristic footprints take the form of a random noise and are unique per device, thus enabling to directly trace data back to their origins. Specifically, we exploit a unique device fingerprint known as photo-response non uniformity (PRNU), without limiting our investigations to statistical methods but opening new horizons towards automatic and data-driven approaches.
The performed experimental campaigns are always carried on images and videos from well-known disclosed datasets. The achieved results assess our strategies as competitive solutions, in terms of accuracy and computational complexity accomplishments.

LEADERSHIP GAMES: MULTIPLE FOLLOWERS, MULTIPLE LEADERS, AND PERFECTION

Alberto Marchesi - Supervisor: Prof. Nicola Gatti

Over the last years *algorithmic game theory* has received growing interest in AI, as it allows to tackle complex real-world scenarios involving multiple artificial agents engaged in a competitive interaction. These settings call for rational agents endowed with the capability of reasoning *strategically*, *i.e.*, taking into account not only how their actions affect the external environment, but also their impact on the behavior of other agents. This is achieved by exploiting ideas from game theory, and, in particular, *equilibrium* concepts that prescribe the agents how to behave strategically. Thus, the challenge faced by the researchers working in algorithmic game theory is to design scalable computational tools that enable the adoption of such equilibrium notions in real-world problems.

In this thesis, we study the computational properties of a specific game-theoretic model known as the *Stackelberg paradigm*. In a Stackelberg game, there are some players who act as *leaders* with the ability to commit to a strategy beforehand, whereas the other players are *followers* who decide how to play after observing the commitment. Recently, Stackelberg games and the corresponding Stackelberg equilibria have received considerable attention from the algorithmic game theory community, since they have been successfully applied in many real-world settings, such as, *e.g.*, in the security domain, toll-setting problems, and network routing. Nevertheless, the majority of the computational works on Stackelberg games study the case in which there is one leader and one follower, focusing, in most of the cases, on instances enjoying very specific structures, such as security games. A comprehensive study of general Stackelberg games with (possibly) multiple leaders and followers is still lacking. In this work, we make substantial steps towards filling this gap, following three different lines of research.

In the first part of the thesis, we address the largely unexplored problem of computing Stackelberg equilibria in games with a *single leader* and *multiple followers*, focusing on the case in which the latter play a *Nash equilibrium* after observing the leader's commitment. We analyze different classes of games. First, we study general *normal-form* Stackelberg games, where we tackle two versions of the problem: the *optimistic* version in which the followers break ties in favor of the leader, and the *pessimistic* one where they break ties against her. While the former can be easily solved, the latter begets non-trivial computational challenges. In particular, for its solution, we propose a *branch-and-bound* algorithm based on advanced optimization techniques. Then, we switch the attention to games with a compact representation, focusing on two classes of games, namely Stackelberg *polymatrix* and *congestion* games. Our main contribution is to identify a simple class of congestion games with many followers, in which each player can only select a single a resource, where the problem of computing a Stackelberg equilibrium is solvable in polynomial time, both in the optimistic and pessimistic setting. This represents the first example of games where finding Stackelberg equilibria is easy even with an arbitrary number of players.

In the second part of the thesis, we study Stackelberg games with *multiple leaders*, proposing a new way to apply the Stackelberg paradigm in such settings. Our idea is to let the leaders decide whether they want to participate in the commitment or defect from it by becoming followers. This is orchestrated by a suitably defined agreement protocol, which allows us to introduce interesting properties for the commitments. Specifically, the protocol is managed by an external mediator who proposes (correlated-strategy) commitments to the leaders and then asks each of them, in turn, whether he wants to stick to the commitment or not (becoming a follower). The first property we define predicates of the *stability* of the agreement, ensuring that no leader has an incentive to defect from the commitment. Instead, the second property guarantees the Pareto *efficiency* of the solution. We also introduce a framework to compute Stackelberg equilibria in such setting, showing that, surprisingly, equilibria with some interesting properties can be computed in polynomial time in some classes of games with a compact representation.

In the third and last part of the thesis, we focus on Stackelberg games with a sequential structure, addressing, for the first time in such setting, the problem of equilibrium refinement. This problem has been widely investigated for the Nash equilibrium, as it is well-known that refinements can amend some of its weaknesses, such as sub-optimality off the equilibrium path. In this work, we show that such issues also arise in Stackelberg settings, and, thus, we introduce and study Stackelberg equilibrium refinements based on the idea of *trembling-hand perfection* so as to solve them. We first provide a characterization of trembling-hand Stackelberg equilibria in general, showing that their set is closed under refinement. Then, we focus on a particular type of refinement called *quasi perfection*, introducing an axiomatic definition of quasi-perfect Stackelberg equilibrium and an algorithm to compute one equilibrium.

HETEROGENEOUS DATA COLLECTION AND ANALYSIS FOR WELLNESS ASSESSMENT

Andrea Masciandri - Advisor: Prof. Fabio Salice

The critical phenomenon of aging population is increasingly reshaping the demographic structure of modern societies. The challenge posed to worldwide Social Health Systems is to extend the period from the moment when a person, getting older, passes from independent living to the need of caregiving services. This thesis proposes a method that involves ICT technology and defines a mechanism of synergistic cooperation among technical, social, and medical competences as a key factor to support health-care, leading to a decrease in hospitalization demands and healthcare costs. Thanks to a dedicated hardware platform, the system is capable to collect data related to the Quality of Life of patients both in their home and in the healthcare facilities; these data, together with other sources such as weather reports and questionnaires, are preprocessed and transmitted to a central server to be analyzed. Finally, in order to provide doctors with relevant information that can be part of their domain knowledge, it is necessary to group and process data in accordance with known wellness models.

Indoor localization The localization of the patients is the first important information that has to be collected in order to infer their wellbeing. We designed a coarse grained localization system based on iBeacon technology constituted by disposable bracelets, fixed antennas, and a mobile application used by professionals and doctors to monitor all the patients. This system has been validated directly in a facility dedicated to the care of Alzheimer's disease.

Monitoring solutions Other source of information can be derived from environmental sensors without affecting the privacy of the monitored subjects: motion and thermal sensors can reveal the presence of a person in its activation area, contact sensors can reveal doors and windows openings, pressure sensor can reveal the presence of a person on a sofa, etc. The use of these systems was limited to the domestic environment with a single inhabitant, in fact at the moment there is no valid solution to distinguish the activation of the sensors in the multi-user case without introducing obtrusive sensors.

Activities of Daily Living Recognition We introduced a new method for Activities of Daily Living recognition in Smart Home environments that provides relevant informations to interpret the patient behaviour. The proposed approach - Semi-supervised Method for Activities of daily living REcognition (SMARE) - it is composed by multiple subsequent steps: the first step consists in removing sensor data artifacts, which can affect the output result (e.g., false activations of sensors). Then, the Segmentation layer detects sequences of Sensor State Transitions related to Epilogue-Prologue ACTs; within this periods, an algorithm identifies the Splitting Activities Border, obtaining the so called "occupation periods", i.e. the time interval to which the some activities are attributed. Those periods are subsequently analyzed through a semantic layer which returns a set of activities and the respective probabilities of true detection. If the semantic layer fails in the recognition of an activity, a semi-supervised approach is used to learn the new scenario.

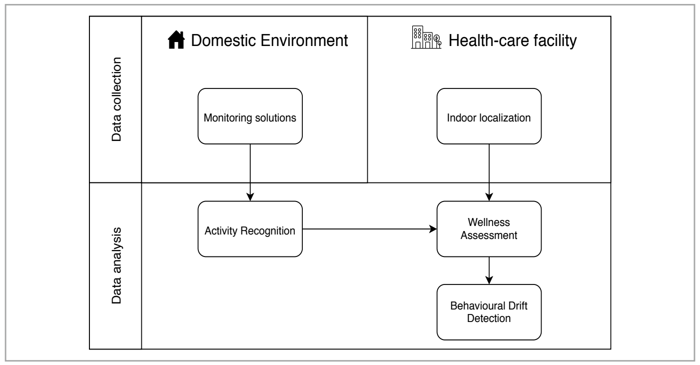

Fig. 1 - The main contributions of this thesis are on data collection and analysis for both domestic environments and healthcare facilities.

Wellness Assessment Wellness Assessment refers to the evaluation of physical, mental, and social well-being. Based on the available localization data, a set of indicators have been designed to support professionals and doctors in assessing the patient's well-being. Starting from the study of wellness models that are currently used by the medical community, we have developed some indicators that can be constantly calculated and analyzed, providing additional information for framing an individual's state of well-being. Those indexes can be divided in Physical Activity indexes, Social Activity Indexes, and Psychological indexes.

Behavioral Drift Detection Patient's activities as well as Wellness Indexes constitute a picture of the state of patients in a given instant. However, collecting daily "pictures", it is possible to create a "movie" for each patient: a further process of temporal analysis can extract hidden information from these data, highlighting the occurrence of anomalies that may occur over a long time period.

The experimentation of the proposed systems was carried out in a domestic environment and in a healthcare facility that deals with Alzheimer's patients. These installations have allowed us in the first place to experiment with emerging IoT technologies and protocols; we also had the opportunity to collect large amounts of data relating to the daily life of the elderly in their respective residences, creating datasets that are in great demand in the scientific field for the validation of new algorithms. In particular, much effort has been made for the experimentation and the partial validation of the proposed system at "Il Paese ritrovato": the results of this study gathered the favorable opinion of doctors and professionals working in the structure. "Il Paese Ritrovato" is the first village in Italy exclusively designed by the La Meridiana cooperative as a pioneering care facility for people affected by dementia, located in the city of Monza. As a gated model village, "Il Paese Ritrovato" is a self-contained community where patients are free to move and interact with each other in wide open spaces, and where structures have been specifically designed to meet the patient's needs, freeing them to the burden of feeling slaves of the disease.

The proposed technological system has the potential to increment the quality of life of patients, to mitigate the stress and anxiety suffered by caregivers, to improve clinical knowledge on the disease, to enhance safety, and to promote the patients autonomy, all outcomes that are particularly relevant for the future of elderly care.

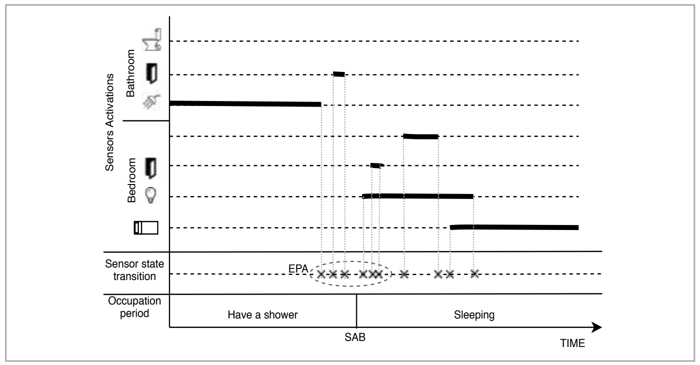

Fig. 2 - The top part of the images reports Sensors Activation: the thick line represents activated sensors. Every time a sensor is switching its state, a Sensor State Transition is fired. Epilogue-Prologue ACTs are time periods characterized by Sensor State Transitions that happen in the neighborhood of Splitting Activities Borders.

TECHNIQUES FOR LOW-JITTER AND LOW-AREA OCCUPATION FRACTIONAL-N FREQUENCY SYNTHESIS

Mario Mercandelli – Supervisor: Prof. Salvatore Levantino

In the last decades, wireless connected mobile devices became pervasive in the consumer market, thanks to the constantly increasing datarate capability and power efficiency offered by modern transceivers. To sustain the forecasted demand for higher communication bandwidth, the current research trend is to adopt complex modulation schemes at higher carrier frequencies. Moreover, to enable the expected ubiquitous diffusion of devices capable of wireless connection, the transceiver cost has to be kept low. This poses significant challenges on the frequency synthesizer implementing the transceiver local oscillator (LO), which has to operate under stringent low output jitter requirements, with low area occupation and integrated in the same CMOS technology as the digital baseband processor.

In this PhD research, two viable alternatives to meet these stringent requirements have been studied.

Fractional-N type-I sampling PLL

First, an extremely low-jitter and low area occupation frequency synthesizer is presented, based on a sampling type-I analog PLL, which avoids the large integrating loop-filter capacitance of type-II analog PLLs. An accurate time-variant model of the PLL implementing the frequency synthesizer is presented, allowing to perform a fast design-space exploration, selecting the parameters for output noise minimization.
To suppress the phase detector (PD) noise contribution, a high-gain sampling PD is used, which is also functional to achieve a low reference spur, thus overcoming one of the conventional issues of type-I architectures. Recently published sampling type-I PLLs are bounded to integer-N operation, due to the phase detector narrow-range. In this work, fractional-N operation is enabled by canceling the quantization noise before phase comparison, by means of a digital-to-time converter (DTC). Thus, the narrow-range PD limitation is overcome, and the benefits of sampling type-I PLLs are extended also to fractional-N frequency synthesis. A digital phase-error correction (DPEC) technique is introduced to limit the sampling PD phase error with no extra quantization noise and extract, at the same time, a digitized version of the phase error. This digital signal is leveraged to implement a number of digital calibration algorithms to assist PLL operation and improve system performances, via the correction of analog impairments (e.g. the adaptive calibration of the DTC range). A simple and robust digital frequency-locked loop (DFLL) is proposed to achieve proper PLL locking over the whole oscillator tuning range, successfully coping with the usual lock-range limitation of type-I PLLs.
The proposed sampling type-I PLL has been fabricated in a 28nm bulk CMOS technology process (die micrograph in Fig. 1). The measured fractional-N RMS integrated jitter is 58.2fs around a 12.5GHz carrier (Fig. 2), with a reference spur as low as -73.5 dBc and fractional spurs lower than -63.2 dBc. With a power consumption of 18 mW, the achieved figure-of-merit (FoM) is -252 dB, outperforming other recently published fractional-N frequency synthesizer. Thanks to the type-I architecture and the proposed digital algorithms, the area occupation is only 0.16 mm², including decoupling capacitances and routing space, which features the smallest footprint among published sub-100fs jitter PLLs.

Fractional-N digital PLL with automatic bandwidth control

Second, an attractive alternative for the frequency synthesizer implementation is presented, which relies on a fractional-N digital PLL.
In the presented architecture, the phase detector is replaced with a 1-bit time-to-digital converter (TDC), and the analog loop-filter is substituted with its digital implementation, thus reducing area occupation and power consumption. Extensive use of digital adaptive algorithm is leveraged, and a technique to accurately control the PLL bandwidth, desensitizing the loop gain from any analog dependence, is proposed and analyzed. The digital bang-bang PLL has been fabricated in a 65nm CMOS process, with the emphasis on the adaptive bandwidth control performance. Even in the presence of an input noise level higher than 14 dB with respect to the nominal condition, the discussed algorithm is able to track and compensate the loop-gain variations, guaranteeing a constant PLL bandwidth with an accuracy better than 4%. The fractional-N frequency synthesizer occupies a core area of 0.22 mm², with a 400fs RMS jitter around a 3.6GHz carrier.

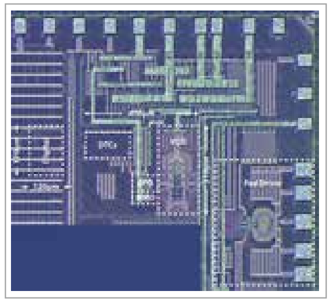

Fig. 1 - Die micrograph of the fractional-N sampling type-I PLL fabricated in 28nm CMOS.

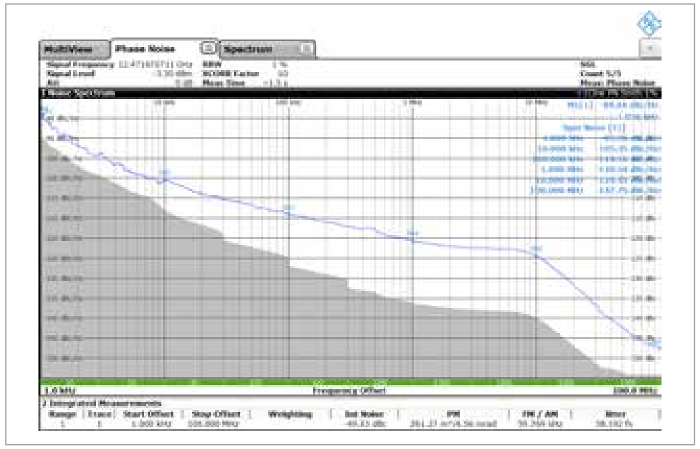

Fig. 2 - Measured phase noise spectrum and integrated jitter of the fractional-N type-I sampling PLL around a 12.5GHz carrier.

NETWORK SENSING AND CONTEXT AWARENESS

Marouan Mizmizi - Supervisor: Prof. Luca Reggiani

Context-awareness is the main driver of the ubiquitous computing paradigm. Well-designed context modeling and context retrieval approaches are essential in any context-aware system. Location is one of the primary aspects of all major context models - together with identity and time. From the technical point of view, sensing, fusing and distributing location and other context information is as important as providing context-awareness to applications and services in pervasive systems.

An intense research work by the scientific community and the industry is carried today in order to design and build localization systems that can operate in different environments and achieve a degree of precision, and reliability. The availability of such Positioning Systems (PS) will permit numerous advances in the disciplines of location-aware, pervasive computing, ambient intelligence, and it will facilitate the deployment of LBS.

In outdoor environment, positioning and navigation are based on the well-established GPS. Thanks to the development of the European Satellite Navigation System, Galileo, the modernization of GPS, and the recent advances in High Sensitivity GNSS technology has opened up new horizons, leading to a new location and context-based applications.

Nevertheless, GPS technology may not always deliver the necessary navigation and positioning information in a number of difficult environments, as well as when there is accidental or intentional interference with the transmitted signals. Indoor environment, underground car parks, and road tunnels are three examples of difficult environments, where the reception of satellite signals is affected.

Similarly, the malicious interference of satellite signals near airports will render the satellite-derived navigation and positioning information unreliable. The risk of such interference may be low, but difficult environments are always present in a number of LBSs, involving the continuous tracking of goods or individuals.

For indoor environments, in order to achieve a large-scale reality, not only a business opportunity must be present, but also the technology should be low-cost, low-power, with low maintenance expenses and it should require the minimum amount of new infrastructure w.r.t. those already present and installed for Wireless Personal or Local Area Networks (WPAN / WLAN).

Although there is no lack of such technologies, high-accuracy indoor positioning is an unconquered domain for the commercial solutions in the mass-market. The research in the area ranges from WLAN coverage area modeling to WLAN fingerprinting, from UWB (Ultra-Wide Band) ranging to RFID (Radio Frequency Identification) tags, from BLE (Bluetooth Low Energy) devices to visual-based solutions.

Systems based on WLAN technology are probably the most interesting in terms of costs. This is because those systems are based on pre-installed Access Points (APs) and also because the most of mobile terminals (new smartphones) are equipped with a WLAN radio. So, infrastructure and end-user devices are already available. However, the barrier is the availability of reference data: whereas the commercialized mass-market medium-accuracy WLAN positioning solutions retrieve the reference data from Global Navigation Satellite System (GNSS) (especially in crowd-sourcing), there is no such source of independent reference position data for database collection in indoor environments. Also, it should be noted that WLAN is an exception in terms of existing infrastructure. For instance, both UWB and RFID-based solutions require a new infrastructure and new radios into mobile devices.

The high-accuracy indoor positioning and navigation will be based on dedicated positioning-specific tags and this is due to compromises between costs and performances. In fact, the choice of WLAN technology, based on the reuse of existing APs, offer certainly the best solutions in terms of costs, but it has many limitations in terms of performances and reliability. However, the global large-scale uptake of such technologies requires that the positioning is based on some existing radio interface. This ensures that the mobile terminals require no new hardware components and that the radio components are already in mass-production keeping the costs of deployment and new applications really low. In addition to the medium range WLAN radio, also the short range BLE is an example of technology with several advantages in terms of costs and deployment.

The main contributions of this thesis are all related to positioning and navigation, spanning from large scale to Nano-scale systems. The principal innovations are in the following:

- Indoor Positioning System (IPS): the main aspects of these systems are reviewed with particular focus on the Fingerprinting technique. Indeed, it is widely used in IPS, especially when technologies such as Bluetooth or WLAN are adopted. Related to this, the innovative aspects are:
  1) analytical model: designing a fingerprinting-based IPS is a complex task because the multitude of parameters that the designer has to deal with. With the analytical tool presented, it is possible to predict the performance of the system before the time consuming and expensive practical installation. Moreover, we present a detailed description of the steps and decision necessary to implement an IPS;
  2) binary fingerprinting: a novel approach to fingerprinting, where the measured RSSI, is represented by a single bit, 1 or 0 depending on the user position. The Radio-Map (RM) can be interpreted as a codebook, where each fingerprint related to a known position is a codeword. The concepts from binary coding theory can be used to assess the performance of a localization algorithm, the Hamming distance becomes a key metric for estimating the position of user, and Gray mapping is a tool for beacon placement. Also, for this novel approach, we have developed an analytical tool to predict the performance of the system in design time. The key achievements are validated through numerical and experimental results.

- Space Labeling Design: inspired by the binary fingerprinting, we have extended the binary labeling beyond the localization problem. We introduce a novel access strategy based on the geographical position of the user devices. The benefits are a reduced signaling and control message from the network. To the best of our knowledge this the first work in this direction, and the presented results show the benefits of such approach;
- Vulnerable Road User (VRU) Protection: from an ongoing collaboration with Nokia Bell Labs, and within the 5GCAR, and EU funded project, we have studied a localization service in outdoor scenario. The goal is to build an application that protects pedestrian and cyclist from car crashes. A live demonstration is planned for June 27, 2019 in Paris. During this collaboration, many findings have been obtained, e.g. a novel approach to map-matching was implemented, this method is more flexible and reliable in critical scenarios, where the vehicle skid of the road;
- Nano-scale localization: context-awareness is a concept originated for electromechanical machines, but it can be even more important for biological machines. The position information in this context can be revolutionary. Nowadays, it is possible to modify genetically some bacteria and program them to sense and execute simple functions, e.g. release of drugs in specific position of the human body, and therefore to reduce undesired effects. We studied how to estimate distances in molecular communication environments, which is the first step for a localization algorithm.

DATA TRANSMISSION AND ANALYSIS OVER HYBRID WIRELESS/NARROWBAND-POWERLINE COMMUNICATION (WINPLC)

Sadaf Moaveninejad - Supervisor: Prof. Maurizio Magarini

Hybrid wireless/powerline communication has been proposed to improve reliability and performance in smart grids (SG). In this work we focus on a hybrid wireless/narrowband PLC (WiNPLC) system for communication links in SG such as that between smart meters (SM) and in-home devices (IHD). An effective approach to enhance the data rate in narrowband power line communication (NB-PLC) system is multicarrier modulation based on orthogonal frequency-division multiplexing (OFDM) and multiple-input multiple-output (MIMO) transmission over multiple power line phases.

Low voltage, in-home NB-PLC networks allow direct communication between SM and IHD. In order to minimize security issues, in many deployment scenarios transmission takes place only towards the IHD to display consumption data, with no backwards channel. As a result, channel estimation is difficult and a key challenge for achieving reliable communication over NB-PLC is to use robust transmission and detection techniques to mitigate the effect of the impulsive noise within the PLC channel and to make possible recovering transmitted data. To this aim, it is fundamental to give an appropriate characterization of such a noise. In fact, substantial components of the noise in NB-PLC systems exhibit a cyclostationary behavior with a period of half the alternating current (AC) cycle. Moreover, when MIMO transmission is adopted, an important issue that must be considered is the cross-correlation between different phases.

Frequency-shift (FRESH) filter is a recently proposed approach that is able to reproduce the effects of cyclostationary NB-PLC noise obtained from measurements. In this work, we propose to classify the noise generated by FRESH filter into three classes, based on the estimation of the respective probability density functions and on the evaluation of their second order statistics. First, we show that while one class exhibits a normal distribution, the other two exhibit an impulsive behavior for which we propose a generalized Student's t-distribution. Simulation results show that the bit error rate (BER) of MIMO-OFDM NB-PLC significantly changes between different classes of noise. Hence, we develop an algorithm for switching data delivery between MIMO-OFDM NB-PLC and MIMO-OFDM wireless transmission in unlicensed frequency band that takes into account knowledge of the periodicity of the three classes of noises. The result is a hybrid MIMO-OFDM wireless/NB-PLC system, which we refer to as, hybrid MIMO-OFDM WiNPLC. Figure 1 shows different realizations of such a hybrid system. We implemented WiNPLC system based on Fig. 1(a), because it is the most practical one from hardware point of view. While Fig. 1(b) and Fig. 1(c) are designed assuming hybrid transmitters and receiver. Our simulation results demonstrate BER improvement of the proposed hybrid system over individual MIMO-OFDM NB-PLC or MIMO-OFDM wireless systems. Further improvement in performance of hybrid system could be obtained by evaluating capacity of the MIMO NB-PLC system in presence of different classes of the noise. This thesis obtained capacity by spatio-temporal whitening of the cyclostationary correlated noise samples generated through FRESH filtering. This capacity is useful for adapting the modulation order and obtaining optimum performance based on the observed class of noise.

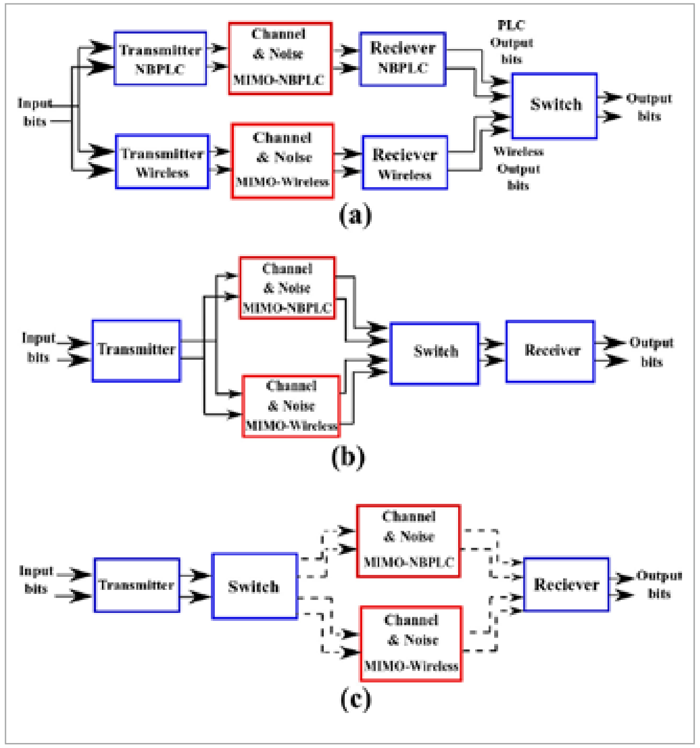

Fig. 1 - Block diagram of the WiNPLC system by switching between MIMO NB-PLC and MIMO Wireless links.

Due to the cyclostationarity of the noise, similar behavior is repeated in next periods and can take advantages from this pre-processing.

To support the future works of other researchers in the field of NB-PLC, we propose a simple and more adaptive method to generate noise samples with characteristics similar to those obtained using the FRESH filter. To this aim, the filterbank approach of FRESH filter is replaced by applying a spectral and temporal shaping to a white Gaussian (WGN) noise random process. In addition, by changing the slope of temporal shaping, distribution of each class could change from Gaussian to impulsive and vice versa. The proposed noise generation approach is compared with FRESH filter generator in terms of normalized mean square error (NMSE) in the cyclic auto-correlation between the measured and generated noise samples.

It is worth noting that the noise introduced by electrical appliances to the communication data could be used as device signatures, which is a useful information for energy monitoring. In this regard, data received to all SMs must be collected and analyzed for improving energy consumption management. Due to the notable rise in the number of installed SMs, Non-Intrusive Load Monitoring (NILM) has received growing attention in the recent years. NILM aims at replacing several SMs by a single SM and estimating the power demand of individual appliances from a household's aggregate electricity consumption, as presented in Figure 2. In the present work, after reviewing different categories of household appliances, the state-of-the-art load signatures, including both macroscopic and microscopic features, are introduced. Next, commonly used supervised and unsupervised disaggregation algorithms, which are employed to classify the appliances based on the extracted features, are discussed. Publicly accessible datasets and open-source tools, which have been released in the recent years to assist the NILM research and to facilitate the comparison of disaggregation algorithms, are reviewed. Finally, main applications of energy disaggregation, including providing itemized energy bills, enabling more accurate demand prediction, identifying malfunctioning appliances, and assisting occupancy monitoring, are presented.

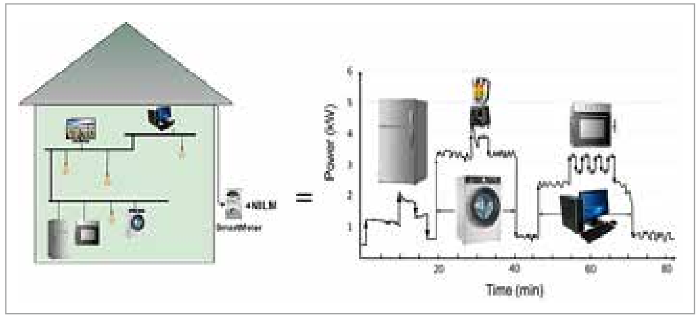

Fig. 2 - Using NILM for estimating the power demand of individual appliances from a household's aggregate electricity consumption obtained from a single SM.

DEVELOPMENT OF A 3" LABR3 SIPM-BASED DETECTION MODULE FOR HIGH RESOLUTION GAMMA RAY SPECTROSCOPY AND IMAGING

Giovanni Ludovico Montagnani - Supervisor: Prof. Carlo Ettore Fiorini

Gamma radiation detection finds many applications in different fields, including astrophysics, nuclear physics and medical diagnostics. Nowadays large Lanthanum Bromide crystals coupled to Photomultiplier Tubes (PMTs) represent the state of the art for gamma detection modules, in particular for spectroscopic measurements. Nevertheless, there is an interest in substituting photomultiplier tubes with solid state photodetectors like Silicon Photomultipliers (SiPMs), owing to the latter's significant advantages. These include insensitivity to magnetic fields, low bias voltage, compactness, fast response and mechanical robustness. The aim of this thesis work, which was carried out within the context of the GAMMA project supported by Istituto Nazionale di Fisica Nucleare (INFN), is the design, development and experimental characterization of a -ray spectrometer based on large Lanthanum Bromide scintillator crystals coupled with Silicon Photomultipliers. This detector specifications are compliant with nuclear physics experiments with energies ranging from 100 keV to 20MeV, characterized by state-of-the-art energy resolution and imaging capability, in a compact, modular and robust structure. In order to perform the readout of large scintillator crystals, a matrix of 144 Silicon Photomultipliers was designed using NUV-HD SiPMs from Fondazione Bruno Kessler (FBK). These were chosen due to their high Photon Detection Efficiency in correspondence with the peak emission wavelength of the crystal, the high cell density and low Dark Count Rate.

This thesis work entails the design of each element of the instrument. Starting from the SiPM tile and the optimization of the instrument mechanics, through to the development of the electronics boards and custom ASICs. The Gain Amplitude Modulated Multichannel ASIC (GAMMA) was developed in order to match the project requirements for a large charge dynamic range and full scale range. The ASIC is meant to cope with the 14 dB charge range provided by the SIPMs and to provide this information minimizing the statistical contribution to resolution degradation. The exploitation of a Gated Integrator filter with self triggering capabilities was meant to optimize the signal collection. The ASIC main schematic is represented by Figure 2: the core of the chip is the 16 channels stack which exploits a current input stage, the Gated Integrator with Track and Hold feature, an Active Gain Control and a Baseline Holder circuit.

Data acquisition and the biasing of the detectors were also designed in order to simplify the use of the instrument in operative conditions. Experimental measurements were performed in intermediate development steps, confirming the high performance of the developed instrument. The main motivation of this work was to develop the first instrument capable of detecting rays from hundreds of thousands of electronvolts to tens of millions of electronvolts with high efficiency and state of the art energy resolution. The measurements performed on a preliminary detection module have demonstrated the ability of the system to consistently achieve results in line with expectations - even better in some cases - within the restrictions given by the limited number of readout channels and by the simplified microcontroller-based DAQ. An energy resolution of 2.58% FWHM at 662 keV has been achieved coupling the designed detector with a co-doped $LaBr_3$ crystal by Saint-Gobain. This is, to

Fig. 1 - Developed spectrometer enclosure. The cylindric shape is designed to contain the crystal, the hexagonal flange allows the exploitation of the system in laboratory setups and the upper box hosts the electronics.

the author's knowledge, the best resolution obtained coupling SiPM with scintillator crystals. A spatial resolution better than 10 mm at the center of the matrix has been obtained dividing the matrix in just 8 macro-pixels. The measurement dataset, shown in Figure 3, was used to train an algorithm capable of reconstructing the horizontal position of interaction with sufficient precision. The next step of the project is to complete the 144-SiPM matrix readout system coupling each SiPM with a dedicated ASIC input: this is achieved by exploiting a total of 9 16-channels ASIC, of which a first release has been tested in parallel to this thesis and the results are reported in the dedicated chapter. The writing of this thesis occurred during the beginning of the testing phase. However, up to now, the only spectroscopic results available were obtained with single 8-channel or 16-channel ASICs in a reduced energy range. It will be possible to read individually the 120 SiPM of the matrix, increasing the energy full scale range to meet the project specification of 20 MeV. The combination of the preliminary system developed during this thesis work with the new 16 channels ASICs and the FPGA-based acquisition system will allow to obtain the ultimate performances targeted by the GAMMA project, satisfying all the requirements of the application: 120 non-merged SiPMs will allow to reach the target energy resolution while reaching the full required dynamic range, in conjunction with excellent spatial resolutions thanks to a large number of small pixels and to the promising neural-network-driven algorithm whose operation will be further refined. The new data acquisition system will permit to acquire data from a all the ASICs even at high event rates, together with an easy management of the SiPM bias voltage to compensate for gain variation due to temperature.

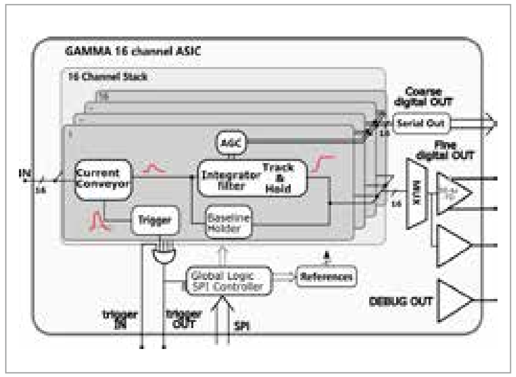

Fig. 2 - GAMMA ASIC main schematic. The analog channel exploits a current input stage, the Gated Integrator with Track and Hold feature, an Active Gain Control and a Baseline Holder circuit.

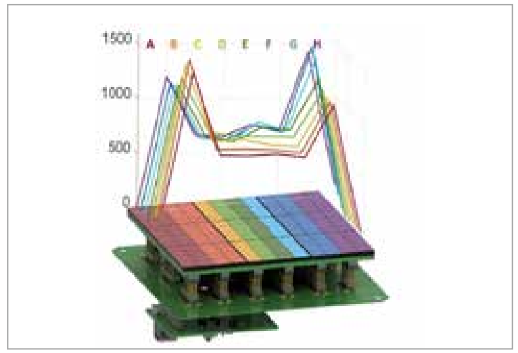

Fig. 3 - Training dataset used for position reconstruction algorithm. Average signal distribution (vertical) for the 8 different positions corresponding to the centers (horizontal) of the pixels and interaction point reconstruction results using a cross-correlation-based algorithm.

SUB-μA POLYSILICON MEMS REAL-TIME CLOCK WITH DETERMINISTIC JITTER COMPENSATION

Giorgio Mussi - Supervisor: Prof. Giacomo Langfelder

This Thesis aims to further develop the use of Microelectromechanical Systems (MEMS) in Real-Time Clocks (RTC). RTCs are electronic components that provide a time base – usually 32.768 kHz – to a more complex electronic system. Such clocks, usually based on quartz resonators, exhibit stabilities in the order of ±100 ppm that can be further reduced to tens of ppm when temperature compensation schemes are employed. The consumption of the clock is typically about 1 μA. Nonetheless, new applications have fostered an interest in MEMS-based RTCs, due to their CMOS compatibility, ease of packaging and overall reduced size. To overcome the intrinsic poorer frequency stability of MEMS resonators ±1850 ppm, -40°C to +85°C) with respect to their quartz counterparts (±100 ppm), attempts have been made to develop new fabrication processes that mitigate the native device instability down to few hundreds of ppm. Such techniques include the use of highly-doped Single-Crystal Silicon wafers or composite materials. This effort, coupled to complex mixed-signal integrated circuits (IC) able to further lower the frequency drift of the reference oscillator down to few ppm, led to the first successful commercial MEMS RTC in 2015.

The goal of this Thesis is to explore the use of standard (hence cheap) polysilicon resonators, characterized by a large linear Temperature Coefficient of Frequency (TCf) of -30 ppm/K, leading to ±1850 ppm drift in temperature. This idea is motivated by the intuition that custom fabrication processes show a less repeatable dependence of the resonant frequency on the temperature f(T), which leads to long and expensive multi-point calibration procedures. Contrarily, standard MEMS resonators, produced with lightly-doped polysilicon are not characterized by the same source of variability and could, in principle, be calibrated only at room temperature, thus reducing the overall part cost. This claim was verified by measuring the f(T) characteristic of 40 polysilicon MEMS resonators, whose structure is represented in Figure 1. The average 2^{nd}-order fitting of the f(T) characteristic was then subtracted from each f(T), showing a part-to-part spread of ±20 ppm, likely limited by the climatic chamber instability, and substantially validating the initial hypothesis of a better f(T) repeatability of polysilicon MEMS resonators.

Then, an IC was designed, including a low-power Pierce oscillator to couple to the MEMS resonator and a frequency compensating scheme, described in Figure 2. In this circuit, the reference oscillator enters as the CLK signal running at ≈540 kHz. Then, a multi-modulus divider divides the input frequency by either N or N+1 (N=16) according to its modulus control input to achieve on average a fractional division factor. The modulus control is driven by a Digital ΣΔ Modulator, implemented as a digital accumulator, whose single-bit output stream encodes in its mean value the Command Word CW. CW is given as an external input in this work but in a final application can be derived from the output of an on-chip temperature sensor. In this way, the RTC output with average frequency of 32.768 kHz stabilized in temperature is obtained at the node DIV. Due to the very low levels of power consumption to achieve, the fractional divider was partly designed using gates in dynamic logic. Although a valid RTC output, DIV is characterized by a large jitter, almost entirely due to the ΣΔ quantization noise. Since a jitter in the order of 30 ns is of interest in certain applications, a subsequent block was introduced to reduce this jitter. This is achieved with the use of a deterministic jitter cancellation, due to the a-priori knowledge of the ΣΔ quantization noise. A Digital-to-Time Converter (DTC) is built around a chain of delay elements whose Full-Scale Delay is matched to the period of the reference oscillator through a Delay-Locked Loop acting as a servo-loop. In this way 32 replicas of the DIV signal are created with increasing delay. A multiplexer then selects the correct phase to output according to the current quantization noise. The overall chip occupies an overall area of 290 μm x 240 μm.

The IC was then coupled to the MEMS resonator to prove its functionality. The oscillator frequency was measured on 10 samples across the temperature range in a climatic chamber. The part-to-part spread of the f(T) was ±6 ppm using the same f(T) fitting used in the initial experimental campaign, validating again the repeatability of polysilicon MEMS resonators. The total measured consumption was 642 nA, 430 nA due to the reference oscillator and 212 nA related to the compensation scheme of Figure 2. The DTC is able to correctly suppress the ΣΔ-induced jitter below 40 ns consuming only 180 nA and with a silicon area of 100 μm x 175 μm.

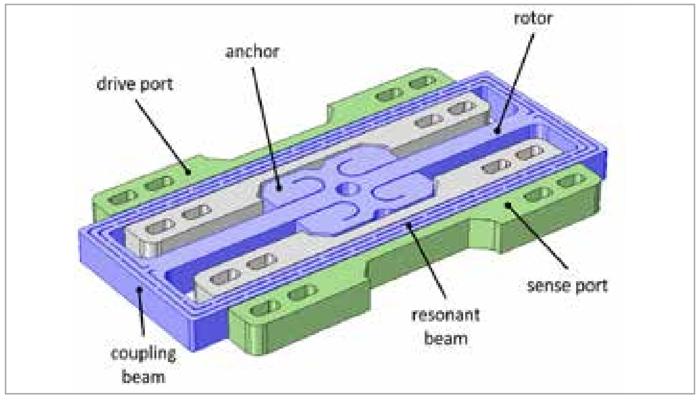

Fig. 1 - 3D view of the used MEMS resonator. Two resonant elements are suspended and anchored at the center of the structure through coupling beams. Slots within the beams are present to reduce the losses due to thermo-elastic damping thus increasing the quality factor.

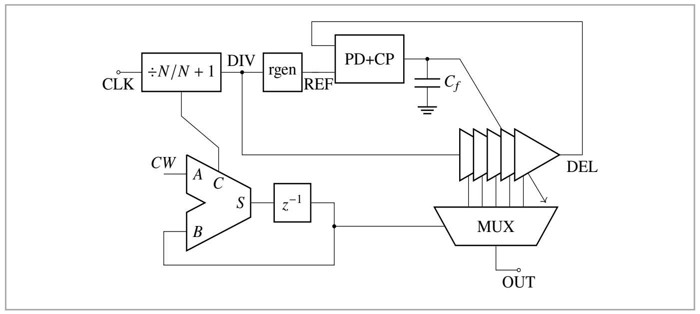

Fig. 2 - Block scheme of the frequency and phase compensating architecture.

SENSING AND DYNAMICS CONTROL FOR ASSISTED AND AUTOMATED DRIVE IN MOTORCYCLES

Dario Nava - Supervisor: Prof. Giulio Panzani

Autonomous driving is one of the most important challenges within the automotive field: indeed, many are the benefits that could drawn from its application in the long run. A prime example is provided by the increased safety on the roads when considering the use of autonomous connected vehicles, able to sense the environment to respond promptly to hazards and communicate with each other in order to avoid accidents. At the same time, such vehicles might provide users with added comfort: there might be no more drivers but only passengers, who should not keep their focus on the road but might engage in any other activity during the travel time.

Nowadays we are still far from making autonomous all the vehicles circulating on our roads. This is true first of all from a purely technological point of view, especially considering complex driving scenarios. Moreover, the ethical implications that the adoption of such technology entail - particularly how an unmanned vehicle is supposed to behave in situations that would imply a moral dilemma for a human driver - are yet to be properly investigated.

Despite these still unresolved problems, the benefits deriving from the steps made in the automated driving direction are already clearly significant. In such sense, it is enough to mention all the Advanced Driver Assistance Systems (ADAS) technologies already available since more than a decade on the market. Indeed, the development and consequent large-scale adoption of such systems are a direct consequence of the technological research aimed at the obtainment of unmanned vehicles.

The ADAS ecosystem comprises a multitude of different devices, apt at assisting the driver in dangerous situations and enhancing his comfort. Basic ADAS - for instance Blind Spot Detection and Lane Departure Warning systems - have the function to alert the driver of possible dangers without however actively intervening to avert them. More sophisticated systems, such as Adaptive Cruise Control (ACC), Lane Keeping Assist and Collision Avoidance systems, share the authority with the driver and can directly act on the vehicle dynamics, representing a first step in the direction of automated driving. Of course, the ever-increasing adoption of ADAS in new cars has a key role in mitigating the risk of accidents and, broadly speaking, improving the overall driving experience.

What stated so far is generally true for four-wheeled vehicles. The development of autonomous drive and, more in general, riders' assistance systems explicitly designed for motorized single-track vehicles (motorcycles, scooters, moped, etc.) is at its very early stages. In fact, the adoption of Advanced Rider Assistance Systems (ARAS), that are the counterpart of ADAS designed for two-wheeled vehicles, is limited and the interest on the topic has grown only in recent years.

This technological unbalance between the four and two-wheeled vehicle is due to several reasons. First of all, it is worth mentioning that a two-wheeled vehicle could hardly be conceived to transport passengers, reducing the field of application and thus the general interest around the technology. Moreover, the transferring of all the developed logics from a car application to a two-wheeled scenario is particularly challenging because of the significant differences in the vehicle dynamics between two and four wheeled vehicles (for instance, the roll angle). A clear example of what just stated is represented by the Anti-lock Braking System (ABS). Such system, introduced in 1978 by Bosch, helps to reduce the stopping distances with undeniable beneficial effects in terms of crash reduction, and is required to be mounted on marketed cars since 2004; however, its adoption on new motorcycles has become mandatory in the European Union only from 2016, with a delay of more than a decade.

By any means, the development of systems devoted to the concept of autonomous riding motorcycles can lead to considerable benefits: for instance, being a motorbike much more agile than a car and less prone to remaining stuck in traffic, it could be seen as a means to support fast delivery services or similar applications. Besides, as for four-wheeled vehicles, the improvements in this field might bring by beneficial effects to innovative systems apt at assisting the rider, thanks to the related research on vehicle dynamics, control and sensing systems. This could help to increase the safety and comfort to the vehicle category, that is intrinsically less favourably placed in case of accidents with respect to cars.

In view of what said above, my Dissertation deals with the development of algorithms oriented towards the autonomous and assisted drive for motorcycles. In this context, being the research field very broad and encompassing many sub-problems, two aspects have been investigated: the sensing problem, intended as the inspection and reconstruction of the surrounding environment with detection of possible dangers, and the control of the vehicle's dynamics.

With respect to the former topic, firstly the problem of detecting road lanes based on a single camera for two-wheeled vehicles has been examined; such task is at the foundations of the autonomous drive problem in urban/highways scenarios. Two algorithms are proposed: the first developed under the hypothesis of no knowledge about the roll angle value, the second explicitly considering the roll dynamics during the identification and tracking of the lane markings. The identified lane information is further exploited to develop a camera based two-wheeled vehicle oriented collision detection algorithm, capable of identifying the nearest preceding vehicle, estimating its distance and determining whether or not it is braking by classifying its braking lights status. The overall algorithms have been validated on real data collected using an instrumented motorcycle.

Furthermore, an autonomous parking strategy for a three-wheeled scooter - stable in vertical position - has been studied, based on a front low resolution solid state Lidar. At the beginning, the algorithm is developed under the limiting assumption of no vehicle odometry available, using only the instantaneous distance measurements to detect the parking area and drive the vehicle into the parking spot. Such hypothesis is then relaxed, and the parking algorithm is extended exploiting an Occupancy Grid mapping algorithm to keep track of the Lidar observations enhancing robustness to the overall strategy. Both the parking strategies performance have been assessed in simulation, while the odometry-free strategy has been validated also experimentally in a real parking environment.

Finally, the path tracking problem for an autonomous two-wheeled scooter has been addressed, considering moderate speed conditions. At first, a simplified model of the vehicle's dynamics is proposed and validated. A path tracking controller is developed, aiming at minimizing the error with respect to the target trajectory. In order to assess the performance of the controller, a simulation comparison with a strategy available from the scientific literature is proposed, exploiting a complex multibody simulator of the vehicle's dynamics.

In the Dissertation's Appendix, an algorithm for the driver's driving style characterization is proposed. Such characterization is meant to be used for the personalization of the driving behaviour of an Adaptive Cruise Control system. Despite the algorithm being tested against experimental data collected on a four-wheeled vehicle, it could be easily extended to the two-wheeled vehicles case. Moreover, it fits one of the main modern trends in the ADAS development: growing care to the user experience and personalization, which is of great importance especially in the motorcycle field, where the interaction between the driver and the vehicle is significant and critical for the user acceptance of such systems.

ADVANCED DRIVER ASSISTANCE SYSTEMS FOR OFF-HIGHWAY VEHICLES

Luca Onesto - Advisor: Prof. Matteo Corno

Co-advisor: Prof. Sergio Matteo Savaresi

Autonomous driving is one of the most important challenges within the automotive field. Autonomous vehicles (AV) have the potential to improve the current transportation system: preventing deadly crashes, providing critical mobility to elderly and disabled, increasing road capacity, increasing the energy efficiency. Autonomous driving is classified is 6 different levels, from level 0 (i.e. no automation) to level 5 (i.e. full autonomy in any driving scenario).
The current market scenario is characterized by different car manufacturers such as Tesla, Ford, GM and software corporation such as Google and Uber.
Most advanced cars are equipped with sophisticated Advanced Driver Assistance Systems (ADAS) able to automatically change lanes, to summon the car to/from a garage parking spot and to fully control the vehicle under specific traffic circumstances.
Even though the major car manufacturers set the goal of reaching the Full Autonomous level within 2030, several accidents involving AVs are reported. This implies that massive research efforts still have to be put into self-driving technologies.
Thanks to the growing market, different ADAS are commercially available also for most popular cars, providing semi-autonomous features. Among them the Automatic Cruise Control (ACC), Forward Collision Warning (FCW), Lane Departure Warning (LDW), Tire Pressure Monitoring System (TPMS), Automatic Emergency Braking (AEB), Lane Keeping Assist (LKA), Traffic Signal Recognition (TSR) and Lane Centering (LC).
ADAS can be classified according to the capability to intervene on the vehicle's dynamics. A system is classified as passive, if its action is limited to provide information and/or warnings to the driver and it has no capability to influence the vehicle movements. Active systems, instead, have the possibility to communicate control signals to an actuator to actively intervene on the vehicle's motion. The main objective of ADAS consists in preventing possible accidents, improving safety but also the driving experience.
What stated so far is generally true for four-wheeled-road vehicles: new technologies are first developed and tested on cars and then, after several years, they are redesigned for other vehicle classes. A clear example of what has just been stated is represented by the Anti-lock Braking System (ABS), introduced by Bosch in 1978. This system reduces the stopping distances with beneficial effects in terms of crash reduction and it is mandatory since 2004 on marketed cars. However, its adoption on new motorcycles has become mandatory in the European Union only from 2016.
This Dissertation focuses on the case of off-highway vehicles, where the open scientific and technical literature is rather scarce on ADAS for this class of vehicles.
This is due to several factors: a smaller market and the fact that, in most cases, solution developed for road vehicles are also applicable to this minor class. Despite, from January 1st 2019 to August 31st 2019, the Consumer Federation of America (CFA) identified 388 fatalities due off-highway vehicles in the USA, they are interested by a less-established legislation concerning licensing, safety and minimum age. These accidents interest both to recreational and utility task vehicles.
Moreover, the application of ADAS or, more in general, the use of automation for utility task vehicles can improve their productivity.
Precision farming, for example, makes use of information technologies in agriculture where both the hardware and the software, developed in first instance by the car manufacturers, are integrated into all procedures connected to farming. As for the transportation system, even in this field, the final objective is a fully automated farming.
In view of what said above, this Dissertation deals with the development of algorithms oriented towards the autonomous and assisted drive for off-highway vehicles. In this context, being the research field very broad and encompassing many sub-problems, two aspects have been investigated: the sensing problem, intended as the inspection and the reconstruction of the surrounding environment, and the control of the vehicle's dynamics.
With respect to the former topic, the estimation of road gradient for a heavy-duty vehicle, namely an agricultural tractor, has been examined. Road gradient is one of the most useful information on the vehicle state since it can be used to improve the tractor performance and safety. Moreover, in general, it can also be used for localization purposes. Firstly, exploiting an external camera, a pitch dynamics analysis based on a computer vision approach is proposed. The estimation of the road gradient has been widely assessed for what concern road vehicles, however the pitch dynamics of a tractor are different from the one of a car. To perform a complete vehicle pitch dynamics analysis, the cabin and the body pitch angles, the longitudinal and the vertical compression of the front and rear tires are analysed. Moreover, also the vehicle slip is considered. Since the mentioned variables are not measurable with the onboard sensors, an image processing approach, based on an external camera is proposed. Then, the results are used to evaluate the feasibility of a road gradient estimator based on the inertial measurement of a mono axial accelerometer installed in the cabin.
Furthermore, an obstacle detection system for a snow groomer is developed. A snow groomer is a tracked vehicle equipped in front with a blade and behind with a tiller (a heavy, hydraulically operated, metal rotating cylinder with metal teeth). The main task of this heavy-duty vehicle is to maintain ski slopes of ski resorts. Snow groomers usually operate during night time, after the close of the ski area, to avoid possible collisions with skiers. The design of the proposed obstacle detection system is divided in two parts: the first one is based on a 2D laser scanner and the second one is based on an infrared monocular camera. In particular, the LiDAR sensor is in charge of the detection of obstacles of all kinds, while the infrared camera is focused on the pedestrian detection. The LiDAR measurements are used to describe the surrounding environment through an Occupancy Grid framework, which is extended for this case study. While the infrared monocular camera is used to perform a pedestrian detection: the presence of skiers is detected through warm areas detection approach; moreover, the system is able to detect and exclude the portion of the image occupied by the front blade.
As regards the control of the vehicle's dynamics, the concept of Anti-lock braking systems to tracked vehicles, in particular to snowmobiles, is addressed. A track-deceleration based ABS algorithm is designed relying only on the track speed measurements. The system is tested it both in straight riding and cornering. The analysis shows that, on snowmobiles, Anti-lock braking systems have negligible advantages in term of stopping distance, but are beneficial in terms of steerability and stability, especially during cornering.
Finally, an automatic steering system for an agricultural tractor for vineyards is proposed. This problem involves both the sensing of the surrounding environment and the vehicle's dynamics control. The objective is to keep the tractor parallel and at the reference distance with respect to the crop rows. The pose estimation with respect to the crop rows is done through the measurements of ultrasonic sensors.
Three different perception algorithms are proposed, and their performance are evaluated through a 3D laser scanner. A grey-box cinematic model of the vehicle is proposed, and the controller consists in a state-feedback control logic, developed in simulation. The overall system is then tested and validated through experimental results.

IN-MEMORY COMPUTING WITH MEMRISTIVE DEVICES

Giacomo Pedretti - Supervisor: Prof. Daniele Ielmini

Powerful computing machines are the key enabler for advancements in scientific knowledge and wealth growth. Aggressive scaling in computers size and consequent increase of computing performance of the last decades, has exponentially incremented human ability to solve mathematical problems, understand and cure diseases and predict catastrophic events. The technology scaling has been driven by Moore's Law which predicted that the number of transistors in a dense integrated circuit doubles about every two years. Following this, semiconductor and computing systems company have been keep up with technology improvement that laid the foundation of nowadays, hyper-connected society. Computing systems are commonly organized in von-Neumann architecture where one or multiple types of memory store the data and the computing unit execute the computation. Also, modern Graphic Processing Units (GPU) and High Performance Computers (HPC) are based on evolved type of this architecture where memory is organized in deep hierarchy and computing split between different nodes. The problem of this architecture is that for nowadays computing problems, where lots of data have to be analyzed, most of the time and energy is spent for handling data travelling from memory to processor and back. This issue is known as von-Neumann bottleneck. Moreover, Moore's law is slowing down, due to physical limits of device scaling and increasing cost for manufacturing. To fulfill the hyper-scaling era needs, new materials, new device concepts, new technologies and new architecture are needed.

Merging novel in-memory, brain inspired and specialized architectures with nanoelectronics emerging memories is the goal of this Ph.D. dissertation. Different computing paradigms, namely spiking neural networks and analog computing, have been explored and networks based on different emerging memories have been realized, demonstrating in hardware new computing concepts.

The first full hardware demonstration of Spike Timing Dependent Plasticity (STDP) have been carried out, through a mixed analog-digital circuit with Resistive Random Access Memory (RRAM) learning synapses. Different applications in pattern learning have been demonstrated, such as online unsupervised learning, WTA networks, grey-scale unsupervised image learning and unsupervised pattern tracking. A Monte Carlo model was developed to finely mimic the network behavior and predict the learning outcome. To design and optimize the network more easily, a compact model based on two rate equations describing the average synaptic behavior through time has been developed. The two model have been extensively compared with statistical experimental data on a memristive network, to study the impact of different parameters. With the optimized values, a two-layers mixed unsupervised-supervised network has been implemented and tested to classify MNIST dataset.

A spatiotemporal spiking neural network (SNN) has been designed and experimentally demonstrated with a RRAM network. First, RRAM devices have been characterized for analog spatiotemporal behavior. Then, a semi-supervised sequence learning has been experimentally carried out in hardware, demonstrating WTA behavior. Finally, the spatiotemporal SNN has been applied to a sound recognition problems, to stress the similarity with biological brain. It has been experimentally demonstrated the ability of recognizing the direction of an incoming sound, in a brain-inspired fashion.

A Hopfield spiking recurrent eural network (RNN) has been designed with 1-transistor-1-resistor Phase Change Memory (1T1R PCM) arranged in array configuration on a printed circuit board. The network has been demonstrated able of performing cognitive computing tasks such as attractor learning and recall. A network Monte Carlo model has been developed to mimic the network behavior. The model has been validated with statistical measurement on the network for recall ability. Competition between attractors has been experimentally demonstrated as a primitive for complex decision making tasks.

The primitive has been extended to Constraint Satisfaction Problem (CSP), in particular Sudoku, which has been experimentally demonstrated on a toy problem, validating with extensive measurements on the time to solve the problem, the aforementioned Monte Carlo model. To accelerate the solution of larger CSP problem and increase the probability of success, a stochastic spiking neuron based on 1T1R PCM device has been developed, to perform computational annealing. Incoming deterministic pulses are converted in stochastic train of spikes embracing the high variability of intermediate-crystallization states of PCM devices. The neuron has been tested on large scale Sudoku-like problem up to Hexadoku, with 16 digit. The stochastic neuron introduces random noise all around the network, making it possible a fast solution. To further accelerate the solution, a two-layer network with a feed-forward input layer and recurring output units has been developed and validated on large scale problem. The network has been then benchmarked against the state of the art, showing high performance.

The concept of analog loop computing has been developed. A crosspoint circuit with analog feedback loops has been designed and experimentally demonstrated with RRAM devices arranged on a printed circuit board, and operational amplifiers giving the analog feedback. The circuit has been experimentally demonstrated able to solve linear system in one step, without iteration. Larger simulations of the network have been carried out, to show the scaling possibility. The impact of RRAM device variation and parasitic resistances have been carried out, to show the design issues in scaling the network. The complexity study of the time to solve a problem for a given error, has been demonstrated to be constant approaching the O(1) physical limits and depending only on the matrix eigenvalue. A similar circuit has been designed and demonstrated to solve eigenvector problems, such as Pagerank and the solution of differential equations. A third circuit comprising two RRAM crosspoint memories and feedback loops has been designed and demonstrated able to compute the Moore-penrose pseudoinverse, thus leading to linear and logistic regression in one step, without iteration. To further demonstrated the circuit features, large demonstration of the Boston house pricing regression has been simulated, with extensive study on the impact of RRAM variation and parasitic. Finally, the circuit has been demonstrated able of computing the weights of an Extreme Learning Machine for classifying 10 digits in just 10 step.

The new circuits designed and the experimental results achieved in this Ph.D. dissertation open the way for circuit co-integration of CMOS and memristive technologies for computing applications. Neuromorphic low- power optimizers based on SNN can be developed and integrated to demonstrated the power of in memory neuromorphic computing for solving complex real world problems. The loop computing concept can be extended to a large scale experimental result, to confirm the simulated constant complexity for solving linear systems. Finally, taking advantages of recurrent networks, analog feedback circuit and memristive technology, complex analog crosspoint networks can be designed to tackle harder problems reducing the time complexity for solution and approaching the potentialities of the most promising computing technology, i.e. quantum computing, with a low power, cost-effective solution.

TITLE: MONITORING DATA UTILIZATION IN FOG COMPUTING ENVIRONMENTS

Xuesong Peng - Supervisor: Prof. Barbara Pernici

Co-supervisor: Monica Vitali

This thesis proposes a framework to improve monitoring data utilization in fog computing environments, reducing the data volume while maintaining the data quality and structure performance.

Over the past few years, there has been a significant increase in the use of Monitoring Data for various domains such as IoT, cloud computing, smart spaces. In all the mentioned domains, applications use Monitoring Data as a critical pipeline to observe the target system (space) and to support the smart decision-making process. This work has been motivated by the increasing attention of monitoring data and the rising volume and dimensions of monitoring data, which brings the rapid grow of maintenance cost to the system owners in various aspects: energy, network bandwidth, storage, etc.

In this work, we try to understand the situation and the environment of monitoring data utilization and how the monitoring data are collected and transmitted and modeled during its lifecycle. We propose a correlation-based approach to reduce the volume of monitoring data where some monitoring variables can predict some other variables based on their quantitative relation, which has been derived by analyses. We propose execution profiles to capture the resources demands of software modules and services to enable better management of computational resources. We propose a deployment methodology in fog computing environments, with hierarchical environment model and resources capacity constraints. We also offer a fault-tolerant scheme to ensure the data quality of raw sensor outputs in the monitoring system, to improve the performance of data utilization in the whole lifecycle. Combining the aspects above, we propose an adaptive monitoring data framework for fog computing environments. The framework learns the relationships between monitoring variables and reduces the monitoring dataset using data reduction services. It can respond to changes in the surrounding environment, keeping these relations updated. It distributes individual reduction services to the fog environment and optimizes the deployment of data reduction services. It also exploits the fault-tolerant scheme to the edge side of the monitoring system for reliable data collection.

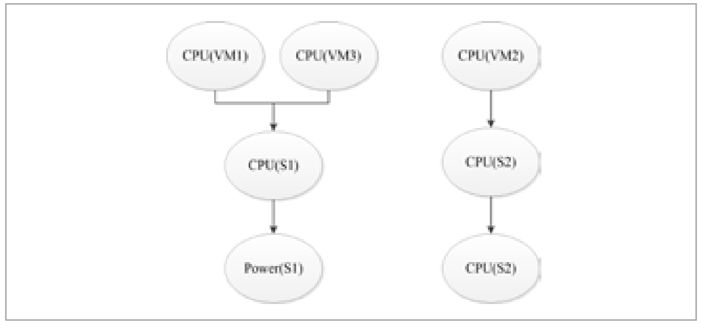

Fig. 1 - Correlations between monitoring variables.

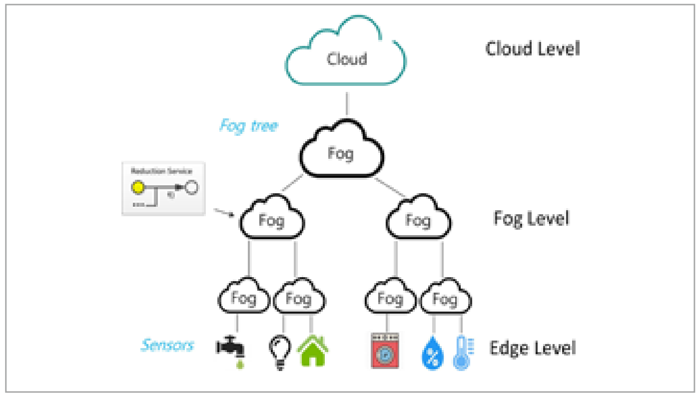

Fig. 2 - Adaptive Monitoring data Framework phases.

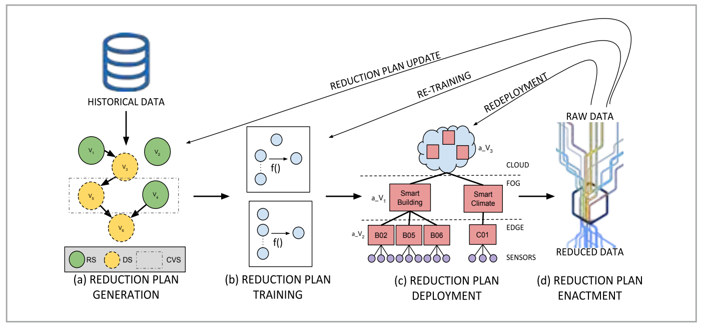

Fig. 3 - Hierarchical structure of the monitoring system.

SECURITY OF INDUSTRIAL SYSTEMS FOR MANUFACTURING

Marcello Pogliani - Supervisor: Prof. Stefano Zanero

Modern manufacturing systems are composed by a plethora of computer-controlled components interconnected by telecommunication networks. Although originally developed as isolated entities, these systems are nowadays highly integrated, in line with the Industry 4.0 vision: They routinely interact with information systems and remote networks for support and maintenance purposes, as well as cloud services for data analytics. This landscape opens a wide attack surface, which can become the avenue for sophisticated attacks. Indeed, several widely known incidents involving industrial control systems, such as Stuxnet and Triton, demonstrated how sophisticated threat actors are willing to carefully reverse engineer the specific control systems in use by their targets to carry out complex multi-stage attacks. Although not specifically directed towards manufacturing systems, such attacks serve as a worrying wake up call to the whole manufacturing sector.

Modern and highly automated manufacturing systems are cyber-physical systems, i.e., systems that act as a bridge between the physical world (the controlled process) and the digital world. Hence, their security requirements and threat model are different than those of conventional Information Technology (IT) systems: While a IT system is concerned about the confidentiality, integrity and availability properties, a manufacturing system is concerned mainly with safety (people's safety as well as environmental safety), and secondarily with process availability and physical system integrity. Additionally, modern "smart" components in the factory floor, such as "Industrial Internet of Things" devices, coexist with legacy devices and protocols, which cannot be easily replaced, updated, or patched—exacerbating the impact of any security vulnerability, and complicating its remediation.

In this dissertation, we aim to understand and analyze the security of a modern manufacturing system. We tackle this challenge from two angles: on the one hand, we aim to study what are the threats targeted at industrial control systems used in manufacturing; on the other hand, we aim to analyze the security of manufacturing systems, and the security implications of programming languages used in manufacturing.

In the first part, we analyze the ***existence of threats targeted at industrial control systems used in the manufacturing sector***: To this extent, we design and deploy a set of low-interaction honeypots on various network vantage points to study the "background noise" traffic (i.e., untargeted probing traffic) directed towards Internet-exposed devices that expose industrial control system protocols, and a high-interaction honeypot to study sophisticated attacks targeted at control systems. By analyzing the data collected by our low interaction honeypots through the analysis pipeline we designed, we find that untargeted ICS scanning is dominated by a few recurrent scanners, with only 4 actors comprising 83% of the traffic and 10 actors comprising 92% of the traffic. Despite most ICS probing traffic being apparently benign and research-oriented, we found instances of harmful behavior (i.e., an attempt to perform a Siemens S7 register write) as well as likely reconnaissance and fingerprinting activities.

In the second part of this dissertation, we analyze the ***security and resilience to attacks of manufacturing systems***. We start by analyzing the components of the direct and indirect attack surface of an entire manufacturing system, showing the existence of multiple attack vectors that can act as entry points of attacks, such as Internet-exposed industrial control gateways. Motivated by the importance of digitally controlled manufacturing systems such as CNC machines, robots, or additive manufacturing systems, we analyze their security. Specifically, starting from the threat scenarios that can apply in the manufacturing context (i.e., production outcome altering, production plant halting, physical damage, unauthorized access), we devise a set of industrial-specific attacks that are based on subverting the basic requirements of manufacturing machines concerning the cyber-physical interactions: accuracy, integrity, safety.

In order to demonstrate the feasibility of the proposed attacks, we apply them to the case of industrial robots, a class of complex, multi-purpose and extremely widespread manufacturing system. We perform a comprehensive technical security analysis concerning the network and physical attack surface of two representative models of robots (a conventional caged ABB robot and a collaborative model by Universal-Robots), finding a set of vulnerabilities that allow an attacker to compromise the robot from the network attack surface. Using these vulnerabilities and composing them into specific attack vectors, we implement the attacks we devised and validate their feasibility and impact. For example, we instantiate the "control loop alteration" attack by exploiting a vulnerability in the controller of our testbed ABB robot to gain remote code execution and modify the value of the proportional gain of the PID controller for each joint in an (encrypted) configuration file. This way, we "detuned" the PID controller, reducing the accuracy of the movement in a way not easily understandable by the operator. Additionally, we analyze the physical attack surface that is "digitally" exploitable through the interaction with the operator.

In the third part, we consider that ***manufacturing machines are controlled thanks to sophisticated programming languages***, such as robot-specific languages derived from ALGOL and BASIC for industrial robots, and G-code for CNC and additive manufacturing machines. Following up with the case study presented in the second part of this dissertation, we analyze the security impact of the complexity of the languages used to develop "task programs" for industrial robots. We study the most important primitives that such languages offer in the categories of file system access, dynamic program loading, and communication functionalities, and we find that such primitives allow a developer to introduce software vulnerabilities in a task program. Indeed, we have found vulnerabilities in real-world task programs: We found a task program implemented in the RAPID language containing a path traversal vulnerability. Motivated by this finding, we implement a prototype static source code analysis tool able to analyze task programs written in two of the main robotics programming languages: KRL and RAPID. Our tool detects by means of taint analysis whether there is a data flow from one or more sensitive sources (e.g., data coming from the network) to one or more sensitive sinks (e.g. robot's movement, file open, late binding calls). We validate the efficacy of the tool by quantifying the presence of unsafe patterns in a set of publicly available task programs drawn from online repositories, composed of 22 ABB and 49 KRL projects. We analyze publicly available code as a proxy, arguing that the patterns we found in public code are found also in production code. We found that 10 out of 22 (45.4%) ABB programs and 6 out of 49 (12.2%) KRL programs present vulnerable data dependencies, and that a common class of vulnerable code is represented by so-called adapters, such as the ones to integrate middleware (e.g., ROS-Industrial) with industrial robots. Unsafe patterns are indeed a recurring use case in robotics applications—especially for socket-controlled robot movement. This justifies the need for a more secure support from behalf of the programming platform, as well as a greater awareness of the security implications of such patterns.

ADVANCES IN PROPULSION SYSTEMS MODELING, OPTIMIZATION, AND CONTROL

Gabriele Pozzato – Supervisor: Prof. Sergio Matteo Savaresi

In the last years, challenges associated with vehicle powertrain modeling, optimization, and control have gained increasing prominence. As a matter of fact, the quest for vehicular emissions reduction, fuel economy improvement, and energy efficiency have led automotive companies to devote time, effort, and money for the development of alternate sources of propulsion for the next generation of vehicles. According to BloombergNEF's annual report on electric vehicles, by 2040, the 57% of passenger vehicle sales will be electric. However, the transition to electric mobility is still ongoing. As a matter of fact, the spread of battery Electric Vehicles (EVs) is limited by three main factors, namely: the low energy density of battery packs, the long charging times, and the higher cost if compared to standard Internal Combustion Engine Vehicles (ICEVs), where propulsion is provided by the engine only. Therefore, automotive companies are currently working towards the development and improvement of hybrid electric powertrain solutions. This technology is a fundamental step for the transition from the fossil fuel based mobility to an electric one. The main feature of hybrid electric propulsion systems is the presence of two energy sources, complementing each other: an internal combustion engine and a battery pack. Thus, two levels of optimization are generally needed in order to fully exploit the capabilities of Hybrid Electric Vehicles (HEVs). First, optimized *component-level* control strategies for the available movers must be introduced. As a matter of fact, the efficiency of internal combustion engines is function of the combustion process in the cylinders. As far as battery packs are concerned, *ad hoc* management strategies are fundamental in order to ensure the effort of all the battery cells to be comparable and to avoid a premature aging of the storage system. Secondly, *supervisory* Energy Management Strategies (EMSs) must be introduced to combine the available power sources efficiently. In this scenario, for the pursue of optimized powertrain solutions, it is crucial to have a deep understanding of both the optimization levels.

Against this background, the doctoral research focuses on the development of both component-level control strategies and supervisory policies. Thus, the hybrid electric powertrain is subdivided into its primal constituents, *i.e.*, the ICEV and the EV. In this scenario, control problems specific to the aforementioned powertrain technologies are taken into account.

1. Internal combustion engine vehicles
ICEVs are still a prevalent solution. Therefore, issues concerning emissions control or the improvement of the engine performance are crucial. Among all these problems, in spark-ignition engines the phenomenon of knocking is recognized as an undesirable event caused by unburnt air/fuel mixture self-ignition, which causes serious engine damages. Therefore, knock detection and control are of paramount importance to prevent damages while optimizing the combustion in the chambers.

On the one hand, the research is focused on the development of an innovative knock detection technique based on the application of principal component analysis to in-cylinder pressure measurements. Relying on this approach, a set of basis functions is used to classify knocking and not-knocking cycles via the logistic regression. On the other hand, a knock control architecture is developed. As opposed to previous approaches proposed in literature, in this strategy the controllers tuning is easily obtained by means of standard control theory techniques. Moreover, the proposed architecture is tested on an in-line three-cylinder engine. This newly developed approach allows for a remarkable reduction of the control action variance (*i.e.*, the spark timing) and shows satisfactory performances in terms of reference knock rate tracking capabilities.

2. Electric vehicles
EVs are considered an effective solution for everyday urban mobility because of the absence of local emissions, the generally low price of electrical energy, and the good performance. The diffusion of EVs on the large-scale is limited by their high cost. In this scenario, the battery pack and its possible replacement during the vehicle's lifetime have a great impact in the economy of electric powertrain solutions. Therefore, an interesting open challenge is matching the lifetime of the battery with the life of the vehicle. To solve this issue, models and operational strategies must be developed to understand, monitor, and control the phenomenon of battery aging.

Against this background, the research activity is focused on the development of an active battery aging control strategy; a problem that, to the author's knowledge, has never been addressed before. Two control actions are used to modify the battery aging behavior: the maximum current that is allowed to be drawn from the battery and the battery depth of discharge. The control objective is defined as minimizing the capacity degradation while guaranteeing an acceptable driving performance. First, the problem is solved offline and for known driving cycles to determine benchmark solutions. Then, an online optimization algorithm is developed to compute the optimal actions while learning the driver's behavior by means of a Markov chain stochastic model, which is updated in real-time. The performance of the online strategy is satisfactory and close to the globally optimal solution. Moreover, robustness with respect to model uncertainties and unexpected aging conditions is shown.

3. Hybrid electric vehicles
A well-known drawback of EV's technology is the low energy density of batteries if compared to fossil fuels, which leads to a limited driving range. To solve this issue, Range EXtenders (REXs) have been successfully introduced. A REX is composed of an internal combustion engine directly connected to an electric generator. This device is electrically coupled to the EV powertrain to extend the driving range, *e.g.*, when no charging stations are available along the trip. In this scenario, the EV+REX architecture is modeled as a series HEV in which two power sources are available: an electric Li-ion battery and the REX. Therefore, a power request can be split between the two sources.

The research work is focused on the development and implementation of energy management strategies for HEVs. The energy management policy is computed solving an optimal control problem formalized as a mixed-integer convex program accounting for the following cost terms: battery electricity consumption and aging, and REX fuel consumption, noise emissions, start-up. All the quantities to be minimized are expressed as monetary variables. Moreover, for the first time noise emissions and start-up costs are introduced in the objective function. First, an offline solution for the EMS problem is proposed. Then, two approaches for online energy management of HEVs are analyzed. The first one computes the optimal solution relying on the Economic Model Predictive Control (EMPC) framework. In this scenario, some dissipativity properties for steady-state and periodic operation of the system under investigation are proved. Therefore, results for close to optimum convergence of the EMPC are provided. The second one is regarded as an explicit policy. As a matter of fact, the EMS problem is formulated as a steady-state optimal control problem. Solving this program for each vehicle operating condition allows to retrieve explicit power split maps which can be used for online energy management of HEVs. This approach guarantees good energy management performances and a reduced computational effort. Eventually, the proposed management strategies are tested, in a simulation environment, considering an electric bus case-study.

HIGH-SPEED AND LOW-NOISE MULTICHANNEL ELECTRONICS FOR BROADBAND COHERENT RAMAN IMAGING

Andrea Ragni - Supervisor: Prof. Giorgio Ferrari

The aim of this research was to develop and test a novel instrumentation for fast Raman imaging combining broadband SRS spectroscopy with a cutting-edge low-noise multichannel electronics based on the lock-in technique. This work is part of the ERC project VIBRA Very fast Imaging by Broadband coherent RAman. The VIBRA project is headed by Prof. Dario Polli of the Dipartimento di Fisica of Politecnico di Milano (Italy) and its goal is to develop a coherent Raman microscope for near-video-rate broadband vibrational imaging. Final applications concern non-invasive cancerous cells detection and identification. A broadband laser consisting of sub-20fs pulses with 80MHz repetition rate is adopted in the experimental setup, enabling in principle, a fast acquisition of the Raman Spectrum. The femtosecond laser, containing the Raman information over a wide range of wavelengths, is spatially diffused on a photodiode array so that each wavelength of the spectrum can be separately acquired by a specific element of the array. A second photodiode array is used as reference in an in-line balanced (IBD) scheme in order to compensate the laser noise. Since the Raman signal is very weak compared to the laser average power (10÷1000ppm), it is modulated in the range 1÷10MHz and the lock-in technique is adopted as detection scheme. The acquisition in the MHz range allows to avoid the flicker noise of the electronics as well as the low frequency intensity noise of the laser. To perform fast measurements, the electronics must be able to acquire, amplify and demodulate in parallel the current photogenerated by every array element. The main requirement of VIBRA, consisting in acquiring with a compact and efficient platform 32 portions of the Raman spectrum in less than 100µs, constitutes the key element for an integration-oriented design of the front-end. Before the IC-design process, a discrete-component prototype with 4 differential channels has been designed and validated in a single-frequency Raman spectroscopy experiment. The spectrum of liquid-phase Methanol has been correctly acquired with modulation frequency of 1MHz and a time constant of 330µs. The pseudo-differential structure can compensate the common mode laser fluctuations, up to 100 times the shot noise level. Even with an intentionally low power (40µW on the photodiode) to simulate the broadband experiment, the prototype is able to reach the shot noise limit with a sensitivity lower than 10ppm (SNR=1). Following the same acquisition scheme, a new 4-ch system (VIBRA4) with a custom IC (0.35µm AMS CMOS process) has been specifically designed for broadband Raman imaging. The main features implemented on this system have been validated through different experimental results. In the modulation range 1÷10MHz, the input-referred current noise ranges approximately from 0.5pA/sqrt(Hz) to 6pA/sqrt(Hz) depending on the gain selection. It results that a shot-noise limited measurement is possible. The autobalancing can compensate any mismatch, in terms of power, between Signal and Reference up to ±40% and with a settling time of around 10µs, fast enough for imaging. This translates in a shot noise limited measurement also when using a noisy laser and in partially unbalanced situation. The acquisition of a Raman image (4-ch, 100x100 pixels, tdwell=10µs, Ppd=300µW), where PS and PMMA can be identified because of their different Raman spectrum, validates the whole acquisition system (optical, electronics and software). Thanks to the experience gained from the design and the experiments on

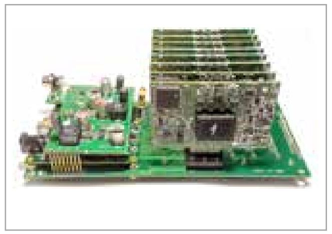

Fig. 1 - Photo of the VIBRA32 platform composed by 8 modules installed on the motherboard with PCIe connectors, main power supply board without shielding and the Opal Kelly module below.

the previous systems, a final 32-ch modular platform (VIBRA32, shown in Fig.1) for highspeed broadband Raman imaging has been developed and tested. This compact platform combines the low-noise front-end of the custom IC, replicated for 32 channels, to a fast-parallel acquisition and elaboration thanks to a Xilinx Artix-7 FPGA (mounted on the Opal Kelly XEM7310 module). The first Raman images of a PS-PMMA sample (Fig.2) acquired with VIBRA32 operating two modules (8-ch, 100x100 pixels, tdwell=10µs, Ppd=300µW), certify the proper operation of this modular platform. To validate the software elaboration and visualization under real conditions, the missing channels were however simulated by the FPGA and transmitted to the visual studio interface (32MB/s data stream). These experiments are conducted in the VIBRA LAB, situated in the Department of Physics of Politecnico di Milano. I acknowledge support by the European Research Council Consolidator Grant VIBRA (ERC-2014-CoG No. 648615) - Very fast Imaging by Broadband coherent RAman.

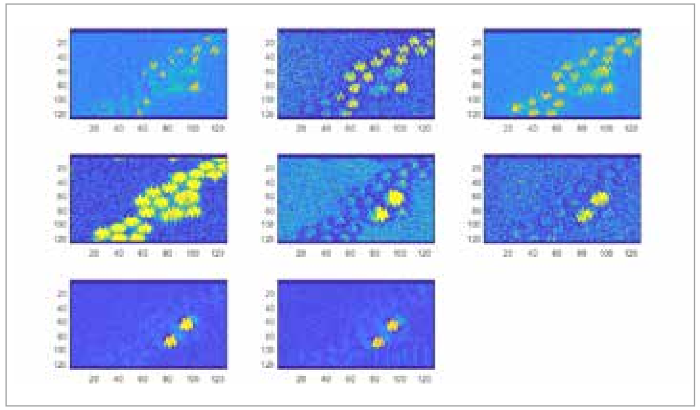

Fig. 2 - First Raman images (8ch, 100x100 pixels) acquired with the platform VIBRA32 in a first preliminary broadband SRS experiment. The plastic beads in PS (10µm diameter) and PMMA (5µm diameter) can be differentiated because of their different Raman spectrum. Parameters: tdwell = 10µs, Ppd = 300µW and fmod = 4MHz. No digital averaging performed nor post-processing for noise reduction, that is the real image acquired with tdwell = 10µs.

MODEL PREDICTIVE CONTROL APPROACHES FOR ENERGY EFFICIENCY IN BUILDINGS WITH HEAT PUMP AND STORAGE SYSTEMS

Soroush Rastegarpour - Supervisor: Prof. Luca Ferrarini

Recent studies highlight that floor heating, known as radiant-floor heating, can be an economic and efficient alternative to more common forms of heating. Unlike the radiator-based heating system, radiant floor heating works at a lower temperature level without any cold spots in the conditioned volumes, as it covers the entire floor. It also offers a more efficient solution from a carbon footprint aspect, especially when it is integrated with heat pump systems and hot water tank storages. On the other hand, radiant-floor buildings as Low-temperature heat emission systems, are well suited for combination with heat pump systems. The large heat exchange area allows for low supply water temperatures, while the high thermal capacity allows to reduce thermal power peaks. Both have a positive impact on the coefficient of performance (COP) of the heat pump.

The heat pump system is one of the most common and energy-efficient methods of space heating systems, since it can convey heat from a heat source (typically, ambient air) to a sink source (tank water) instead of generating heat directly, e.g. by using an electrical element. It offers economical alternatives of recovering heat from different sources for use in various industrial, commercial and residential applications hence it has become a key-component in an energy recovery system with great potential for energy savings. From a control point of view, any heat pump can be described via its coefficient of performance (COP), which relates the compressor electrical power to the thermal power delivered to the sink source.

COP of the heat pump is a time-varying function of several variables, including external inputs such as outside air temperature and humidity, and load-dependent ones such as inlet water temperature coming from the load. To this end, the lower the inlet water temperature is, the higher the heat pump performance is. Heat pump systems can also be beneficial for demand-response applications, where the main focus is to shift demands from on-peak to off-peak periods. In this case, heat pumps are usually connected to a buffer hot water tank (HWT) both to increase the COP and to decouple the (thermal) load and the (electrical) heat generation.

According to these motivations, remarkable efforts have been put in the research field of classic and advanced control techniques for the application of the building energy control and, in a higher level, microgrid optimal control. Among all the possible solutions, particularly interesting appear to be those based on Model Predictive Control (MPC). Several reasons can be listed to support this claim. First of all, MPC is an optimal multi-variable control technique and, among all advanced control approaches, it shows to be more adopted by the industrial world, as it can cope with all operational constraints. Secondly, it is able to consider different, possibly conflicting, goals into the control problem without degrading the stability criteria of the system as in the last years fundamental properties such as stability of the closed-loop systems and robustness with respect to several classes of external disturbances have been proved for many different MPC formulations. Lastly, it is able to consider the empirical models obtained from the process, e.g. through impulse response of the system, to design the controller.

Moreover, when controlling large-scale systems, it is possible to exploit the modular structure of the system considering a distribute MPC technique, where values of inputs, states and outputs are explicitly computed over all the prediction horizon at each sampling time by a local controller designed with MPC and they can be used as information to be transmitted to other local controllers to coordinate their actions. To this end, each MPC has different possibly conflicting objective function, and also information is transmitted and received once in each time interval. In this framework, the logic/algebraic and dynamical system model may be described in the form of a mixed-logic-dynamical model by which the logic constraints of system, such as charging/discharging of storages and power buying/selling from/to utility grid, can also be included in optimization problem.

In this dissertation, a small microgrid including the radiant-floor buildings as the load and several energy storages and generators, such as thermal and electrical energy storages, heat pump and PV- panels, are considered and modelled to be used into the MPC formulation for the prediction purposes. We aim then at evaluating the advantages and disadvantages of different MPC methods considering a complex nonlinear model of the system. To be more precise, four different MPC techniques are studied, which are labelled standard linear MPC (SMPC), linear time-varying MPC (LTV-MPC), nonlinear MPC (NMPC) and distributed linear MPC (DMPC). On the other hand, this thesis studies the impact of using different types of energy storage integrated with a heat pump for energy efficiency in a group of radiant-floor buildings. To this end, it presents an optimal control formulation based on an Economic Nonlinear MPC scheme, in order to find the best compromise among making the heat pump work in time windows when it is more efficient, storing electrical energy when it is cheap, storing thermal energy in the tank when the heat pump is more effective, modulating the inlet water temperature to satisfy the users' comfort constraints, exploiting the buildings thermal inertia. Moreover, the thesis explores the impact of the simplification level of the heat pump model on the overall quality of temperature control in a building, and on electrical energy consumption. As for the heat pump modelling, a detailed analysis has been conducted, where the heat pump is modelled by its coefficient of performance (COP) and validated experimentally with tests on the real heat pump at SYSLAB, Department for Electrical Engineering, Risoe Campus, Denmark Technical University (DTU), based on different weather and load conditions considering all the operational limitations.

TIME- RESOLVED MULTICHANNEL OPTOELECTRONIC INSTRUMENTATION BASED ON PULSED LASERS AND SINGLE- PHOTON DETECTORS

Marco Renna - Supervisor: Prof. Alberto Tosi

Advancements of optoelectronic instrumentation and measurement techniques based on single-photon detection fostered a wide diffusion of photonic components in a large variety of different applications, from the automotive and the industrial fields, to material sciences, quantum communication networks, biomedical and clinical applications. The technique to non-invasively probe diffusive media in depth by means of near-infrared light is known as Near InfraRed Spectroscopy (NIRS), and has been successfully adopted for brain and muscle oximetry, optical mammography, in-vivo molecular imaging and many other applications. Pulsed laser sources and single-photon detectors, thanks to the Time-Correlated Single-Photon Counting (TCSPC) technique, enable for time-resolved NIRS measurements, allowing to exploit the strong connection between depth penetration into the sample and the photon arrival times, and to retrieve absolute concentration of constituents and discriminate contributions coming from different layers of the sample.
The first part of my Ph.D. research activity focuses on the design of a complete system for time-resolved NIRS, featuring eight fiber-coupled custom pulsed diode lasers emitting in the 630 – 1050 nm range, two wide-area single-photon detectors based on Silicon Photomultipliers, and two time-measurement instruments based on a custom Time-to-Digital Converter. The instrument aims at improving state-of-the-art of time-resolved NIRS instrumentation, thanks to enhanced optical performance still guaranteeing noteworthy reductions of both costs and dimensions. The system has been designed within the framework of the E.U. funded H2020 LUCA project, which aims at developing a complete multimodal instrument based on Ultrasound and Diffuse Optics techniques for the early and non-invasive analysis of thyroid nodules. The achieved advancements on photonic components have been adopted also to improve optical performance of a portable dual-wavelength system for time-resolved NIRS, previously designed during my M.Sc. thesis, allowing for an in-vivo measurement campaign on freely-moving subjects. Finally, a stand-alone single-channel version of the developed laser diode pulser has been designed to drive different commercially-available laser diodes in gain switching regime with state-of-the-art optical performance.
The second part of my research activity focuses on the design and development on single-photon detectors based on the fast-gated operation of Single-Photon Avalanche Diodes (SPADs). SPAD detectors are nowadays gaining ever increasing interest thanks to low noise, picosecond timing resolution and ease-of-use. Additionally, SPADs can be operated in gated mode by driving the detector ON/OFF with sub-nanosecond transitions, allowing to perform a sharp time filtering of the incoming light signals and increase measurement dynamic range and signal-to-noise ratio.
With the goal of overcoming current limitations of time-gated single-photon detectors, two instruments have been designed: a single-pixel fast-gated SPAD module and a 16 × 1 fast-gated SPAD array module. The single-pixel module has been designed to operate either silicon or InGaAs/InP SPADs at high count rate, for applications such as Quantum Key Distribution (QKD), thanks to high gate repetition frequency and sub-nanosecond optical gates. The silicon version will be employed for the development of a complete QKD network within the framework of the E.U. funded H2020 QuantERA SQUARE project. The 16 × 1 fast-gated SPAD array module is based on an array of CMOS SPADs driven in fast-gated regime by an Active Quenching Circuit (AQC) array ASIC. The system has been specifically designed for Non-Line-Of-Sight (NLOS) imaging applications within the DARPA REVEAL project, where a complete imaging system has been developed for the reconstruction of scenarios hidden from the direct field-of-view.

HIGH-EFFICIENCY INDUCTORLESS FREQUENCY SYNTHESIS

Alessio Santiccioli - Supervisor: Prof. Salvatore Levantino

The roaring demand for wireless connectivity at a low price point has, in recent years, spurred the interest for highly-integrated transceiver solutions able to cut down on expensive silicon area requirements. In this context, one of the major limiting factors is generally represented by the frequency synthesizer used to generate the local oscillator signal for the transceiver. Conventionally implemented as phase-locked loops (PLLs) based around LC-oscillators, they require large amounts of area due to their use of integrated inductors, that additionally do not benefit from process scaling. However, they are generally preferred to ring-oscillator-based solutions, as they offer a superior jitter-vs-power tradeoff, which benefits the overall transceiver efficiency.

A promising alternative to LC-based PLLs, which overcomes the efficiency limitations of conventional inductorless implementations, is represented by the recently proposed multiplying delay-locked loop (MDLL) architecture. Leveraging injection-locking, these kinds of systems provide aggressive filtering for ring-oscillator phase noise, ensuring both low-jitter and low-power operation in the integer-*N* mode. Unfortunately, when extended to fractional-*N* frequency synthesis, MDLLs suffer from a large jitter penalty due to the noise introduced by the digital-to-time converter (DTC), which is required to perform edge-synchronization. This, in turn, leads to a substantial performance gap between the integer- and the fractional-*N* mode, which prevents their use in demanding applications.

This work aims at closing the performance gap, by proposing a number of design methods and techniques to improve the jitter-power tradeoff of fractional-*N* MDLLs. First, an accurate time-variant model for MDLL phase noise is developed and analyzed to derive close-form expressions for the output spectrum. Then, leveraging the results obtained from this model, the MDLL jitter-power tradeoff is examined and conditions for optimum performance are identified. Finally, the DTC design tradeoffs are evaluated and a technique to significantly reduce its jitter contribution is proposed.

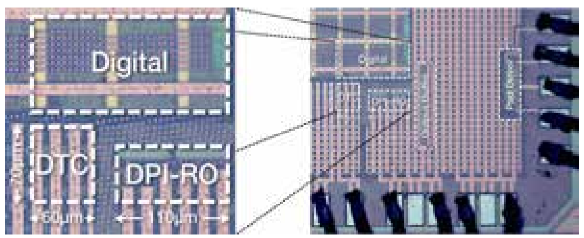

Fig. 1 - (Die micrograph of the fractional-N MDLL prototype, fabricated in 65nm CMOS.

To validate these results, a fractional-*N* MDLL has been fabricated in 65nm CMOS (Fig. 1). The prototype occupies a core area of 0.0275 mm^2 and draws 2.5 mW power from a 1.2 V supply. The maximum RMS jitter is 334 and 397 fs for the integer-*N* and the fractional-*N* case, respectively. Achieving a jitter-power figure-of-merit (FoM) of -244 dB in the fractional-*N* mode, the proposed system effectively bridges the gap to integer-*N* implementations.

TRAFFIC MANAGEMENT IN NETWORKS WITH PROGRAMMABLE DATA PLANES

Davide Sanvito - Supervisor: Prof. Antonio Capone

Traditional computer networks include packet forwarding and specialized devices (middle-boxes) running vertically integrated data and control planes, each one with their own low-level vendor-proprietary configuration interface, making networks complex and difficult to manage. Software-Defined Networking (SDN) is a computer networking paradigm based on the decoupling of the control plane from the data plane. The control plane is logically centralized in an external device (the network controller) while network devices just run the data plane, effectively forwarding the traffic, and expose a common programming interface to the controller to configure the network forwarding policies. Rather than designing complex distributed protocols to achieve the desired global network behavior, network managers can now directly operate on the global network view offered by the controller to configure the network and to deploy new services and applications more easily. With SDN, innovation happens at software speed: a network can be reconfigured by running a different application on top of the controller, bringing to the networking domains software engineering best practices such as code modularity and reusability. An SDN device can be re-purposed without waiting for the slow development cycle from vendors and at the same time the introduction of new networking programming abstractions and high-level languages simplifies the network management and facilitates the evolution of the network. In this thesis work we analyzed the network programmability opportunities for traffic management offered by SDN at three different layers: control plane programmability, data plane programmability and in-network computing. The thesis can be accordingly divided in three macro sections.

We started from the control plane programmability and exploited its global view to design a Traffic Engineering framework to enable online traffic optimization based on periodic traffic measurements and predictions. Even if the SDN paradigm provides the needed flexibility for frequent network updates, the achievable network reconfiguration rate is limited by several aspects including the uncertain nature of the traffic, the limited speed of flow programming in hardware and the traffic monitoring overhead. In order to avoid too frequent network reconfigurations, we defined an optimization model to cluster the observed Traffic Matrices in the time, traffic and routing domains to build a limited set of corresponding routing configurations to be proactively applied with a guaranteed minimum holding time. By tuning the model's parameters we can explore the trade-off between Static TE and Dynamic TE approaches and reduce the number of reconfigurations with a small deviation of performance with respect to the continuous update of the network configuration. We also allow adjacent clusters to overlap to make smoother the transition between routing configurations and help consistent update mechanisms. Moreover, we show how to implement our solution in a production-ready SDN platform for network operators (ONOS) through a new service which have been included in the project's codebase as official open-source contribution.

Despite the robust nature of each routing configuration, the controller intervention is still required to handle exceptional events such as network failures and congestion. These kinds of scenarios challenge the strict decoupling of the control plane from the data plane of the SDN paradigm, because the intrinsic control channel delay and control traffic processing time at the controller prevent a prompt and scalable reaction. Switches are indeed unable to modify their configuration without relying on the external controller. To cope with such events, in the second part of the thesis we designed two applications based on the capabilities provided by advanced stateful programmable data planes which enable the offloading of control functions to the device to support the adaptation of the forwarding configuration in the fast-path according to the locally observable state of the network. In the first application we offload to the data plane the detection and the recovery of network failures in order to reduce both the recovery latency and packet loss while avoiding to rely on the external controller, even for the case of distant failures. In the second application we instead focus on data-center scenarios and show that we can delegate to the networking device the tasks of monitoring and re-routing large flows colliding on a same path in order to quickly balance the load and improve the flow completion times without the intervention of the remote controller.

Latest advances in the data plane programmability pushed further the potential of SDN. The possibility of programming the packet parser and defining custom per-packet actions enables a new class of applications beyond pure packet forwarding under the in-network computing paradigm: the network is not just a pipe which merely forwards packets among endpoints, rather it can modify the application data while it flows into the network, effectively making the network part of the pool of computing resources of the infrastructure. As a third contribute, we investigated the option of offloading the execution of neural network inferences on a commodity SmartNICs. By adopting neural network quantization techniques, we simplified the set of operations required to execute a neural network inference to fit the computational and memory limits of the networking hardware. The ability to run a traffic analysis task directly where the forwarding decision is accordingly taken allows to achieve lower latency and higher throughput compared to an equivalent system implemented on a general purpose CPU while freeing an entire CPU core.

PRINTED CARBON NANOTUBES BASED TRANSISTORS: FROM CHARGE TRANSPORT STUDIES TO BIOSENSING APPLICATIONS

Francesca Scuratti - Supervisor: Prof. Marco Sampietro

In recent years, solution-processed carbon-based electronics has emerged as a groundbreaking technology, capable of delivering new features and functionalities that could pave the way to a plethora of unprecedented applications in the fields of wearables, logistics, healthcare and bio-sensors.
Indeed, functional carbon-based materials like semiconducting polymers and single walled carbon nanotubes (s-SWCNTs) possess unique mechanical properties essential for the implementation of most of the aforementioned applications, such as flexibility, stretchability and softness - impossible to replicate with silicon or other inorganic materials. In addition, the possibility of depositing these materials from solution allows the use of printing and direct-writing techniques, enabling low-cost and high-throughput processes compatible with a wide range of conformable substrates such as plastic, paper or fabric.
The combination of these techniques and organic compounds proved to be an excellent choice for the realization of electronic devices with optimal processability and ever-increasing performance, first of all for the fabrication of field-effect transistors (FETs), the basic building block of complex circuits and applications. In particular, s-SWCNTs are currently object of intensive research thanks to their impressive mechanical and electronic transport properties which, in conjunction with their capability to be processed from solution at low temperature, make them ideal candidates for high-performance printed and flexible electronics.
From a practical point of view, since the growth process of carbon nanotubes always produces mixtures of metallic and semiconducting chiralities, it is necessary to separate them before further processing: by noncovalent functionalization of the side-walls through polymeric chains (*polymer wrapping*), nowadays dispersions of semiconducting SWCNTs with purities > 99.9 % can be obtained, leading to the deposition of networks that display field-effect mobilities higher than state-of-the-art polymeric materials.
However, despite their superior performances when compared to other carbon-based materials, the full potential of carbon nanotubes in printed electronics applications is still hindered by a general lack of understanding of charge transport mechanisms in mixed networks. Indeed, a thorough description of the factors hampering charge transport in these systems is highly desirable to rationally tailor and guide their processing, in view of drastically improving the performances of carbon nanotube based solution-processed devices. The interest in unraveling the dynamics of transport to enhance the devices' figures of merit is transversal to multiple fields, as carbon nanotubes networks can be employed not only for standard solid-state electronics and circuitry, but also for soft and bio electronics: indeed, their extraordinary stability in harsh environments and low biofouling in biological samples make them suitable for enabling novel biosensing and biomonitoring applications not achievable with polymeric materials.
In this context, increasing efforts are being dedicated to the development of cellular models as *in vitro* alternatives to animal experimentation. Cell-based biosensors constitute a fundamental tool in biotechnology, and their relevance has greatly increased as a result of a surging demand for reduced animal testing and for high-throughput and cost-effective in vitro screening platforms dedicated to environmental and biomedical diagnostics, drug development and toxicology.
At the moment, electrochemical cell-based biosensors are being extensively studied and employed for such applications. However, this class of devices presents several limitations, including limited compatibility with complementary optical cell-probing techniques, and need for frequency-dependent measurements. The excellent stability and optical transparency of carbon nanotubes, intergrated in a simple but powerful platform such as electrolyte gated transistors (EGFETs), could help overcome these limitations.
The experimental work presented in this document aims at providing fundamental understandings of charge transport mechanisms in SWCNT networks, with the idea of developing useful guidelines to further improve their performances in practical devices, as well as at illustrating the potential of carbon nanotube based transistors in novel bioelectronic applications, by leveraging on their stability and biocompatibility properties.
Firstly, the charge transport dynamics in aerosol-jet printed, mixed semiconducting networks of carbon nanotubes were studied through charge modulation spectroscopy; to the best of our knowledge, it is the first time that this technique has been applied to s-SWCNTs. We demonstrated the large influence of the diameter distribution in such semiconductors, finding that mobile charges are predominantly located on and move through those chiralities with the smallest bandgaps, even if they represent only a small proportion of the network.
This notion was confirmed also in the charge transport investigations of inkjet printed s-SWCNTs with variable network densities and compositions. By tailoring the printing process was possible to control the electrical properties of the semiconducting layers, shifting from ambipolar to p-type behavior. A thorough analysis of the transport properties evidenced the existence of a delicate balance between network morphology and charge transport efficiency, induced by the reconfiguration of charge paths upon variation of the network density. Temperature-dependent and charge modulation spectroscopy characterizations indeed evidenced that the preferential transport through small bandgap chiralities is disturbed by the creation of energetic barriers for charges with increasing printed layers, lowering the network transport efficiency and hence the overall device performances.
These investigations provide useful indications for the design of carbon nanotube networks in field-effect devices, suggesting that narrow diameter distributions of tubes wrapped by a low-bandgap polymer are better suited for achieving higher currents and mobilities, as well as demonstrating that by carefully tailoring the deposition process is possible to achieve unhampered and band-like charge transport.
Finally, the outstanding stability of SWCNTs in aqueous environments was exploited to develop a novel biosensing application. By profiting from the electrostatic interactions occurring at the cell/ semiconductor interface, electrolyte-gated field-effect transistors based on solution processed carbon nanotubes networks allowed to electrically monitor adhesion and detachment processes of *in vitro* cell cultures, an application that has never been explored for this category of devices. The remarkable stability of SWCNTs in cell culture medium granted the efficient operation of the device over a time span of few days upon continuous immersion in the electrolyte.
The approach proposed in this work has the advantage of employing a very simple acquisition set-up with respect to other emerging technologies for *in vitro* monitoring. In addition, the compatibility of our devices with well-established optical probes and with large-scale deposition techniques could hopefully open the path to a novel set of tools for bioavailability assays. Carbon nanotubes based high-performance electronics and cell-proliferation monitoring through field-effect transistors are not yet mature fields, and appropriate efforts, turned to both technological and scientific criticalities, must still be dedicated to fully unravel their potential.
To this extent, the progresses reported in this dissertation aim at providing new fundamental insights and tools, thus possibly contributing to the further advancement of large-area electronics and bioelectronics.

MICROELECTRONICS FOR SINGLE-PHOTON TIME-OF-FLIGHT MEASUREMENTS

Vincenzo Sesta - Supervisor: Prof. Franco Zappa

In the last years, the accurate measurement of time intervals is required in many single-photon detection applications, based on Time-of-Flight (TOF) and Time-Correlated Single-Photon Counting (TCSPC) techniques, such as fluorescence lifetime microscopy, time-resolved spectroscopy, or 3D ranging based on LiDAR (light detection and ranging). A number of technologies can reach single-photon sensitivity; however, among those, only SPADs (Single-Photon Avalanche Diodes) are the only solution capable of achieving single-photon sensitivity and time-resolved detection in a rugged and compact solution that can be realistically employed in outdoor environments. This PhD research aims at developing state-of-the-art microelectronic circuits for high-resolution time interval measurements, namely Time-to-Digital Converters (TDC) and related electronics for integration either as a stand-alone device or inside SPAD array detectors.

The first step was the design of a high performance TDC core for time interval measurements, using a standard cost-effective 0.35 µm CMOS technology, compatible to state-of-art SPADs. The converter was developed to achieve about 6 ps LSB resolution and about 80 ns full-scale range, with Differential Non-Linearity (DNL), lower than 2% LSB and conversion rate up to $2 \cdot 10^7$, as required by TCSPC and TOF based applications. To achieve such performance, the TDC was based on a 4-bit counter followed by two interpolators, which implement the cyclic sliding-scale technique to significantly improve converter's linearity. Each interpolator exploits two interpolation stages, thus allowing picoseconds resolution while keeping conversion time lower than 55 ns.

The stand-alone TDC chip was fabricated and then characterized to prove the architecture and performance. Some significant issues were discovered in the implementation and a second chip was developed, in a 0.16 µm BCD technology. The changing of technology was due to the requirement of the "REVEAL" project ("Revolutionary Enhancement of Visibility by Exploiting Active Light-fields"), funded by DARPA, which required the development of a 16 × 16 SPAD array with picosecond resolution TDC. This detector was targeted to non-line-of-sight (NLOS) imaging, aiming at reconstructing the shape of hidden objects from multiple scattered light beams. To this purpose, a 16×16 SPAD array with 16 on-chip TDCs and also a stand-alone TDC was developed, which are now under fabrication.

As a part of this research, a fast timing electronics chip was also developed to integrate in large area time-gated digital Silicon Photomultiplier (SiPM), to be exploited in innovative 'smart optodes' for the H2020 'SOLUS' ("Smart Optical and Ultrasound Diagnostics of Breast Cancer") project. The 'SOLUS' project aims at developing an innovative multi-modal imaging system for in-vivo diagnosis of breast cancer, by combining multi-point optical investigation with traditional ultrasound imaging. To do this, 8 optodes need to be placed around the ultrasound transducer in the probe, each one combining pulsed laser sources and a detector. The detector with integrated timing electronics must achieve state-of-art performance, facing very strict packaging, thermal and integration challenges in order to result in a very compact optode. The developed timing electronics consisted of a TDC and a histogram builder circuitry for creating an on-chip histogram of photons' arrival times, able to convert at least 10^7 events per seconds. The TDC was designed in a 0.35 µm CMOS technology, achieving 72 ps LSB resolution and 9 ns full-scale range.

A histogram builder circuit processed the TDC output code and created a histogram with 128 channels and 12-bit depth. Moreover, a deterministic dithering was introduced to improve the overall conversion linearity.

The detection chip with on-chip timing electronics was fabricated and then tested and characterized, using a custom board and Xilinx Spartan-6 FPGA board to elaborate the raw measurement data. The tested timing electronics shows optimal performance suitable to TD-NIRS applications. The Full-Scale Range is about 10 ns with maximum channel width of about 106 ps and TDC conversion time is shorter than 100 ns. The converter precision was measured showing a Full-Width Half Maximum (FWHM) better than 150 ps; the root-mean-square value of the DNL is about 3.9% LSB with dithering enabled, while it increases to 14% LSB with dithering disabled. The experimental characterization of the detector and the timing electronics shows that its performance are adequate for TD-NIRS applications at short source-distance separation, and the small size and reduced power consumption makes it possible to integrate the chip in an extremely compact optode, suitable for the integration in the final probe for the SOLUS system.

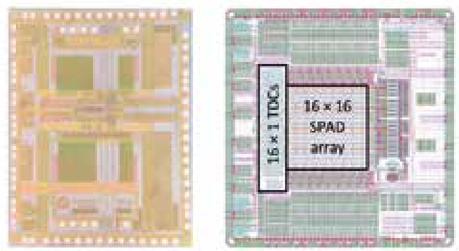

Fig. 1 - Micrograph of first high performance TDC (left) and layout of 16 x 16 SPAD imager with redesigned TDCs (right).

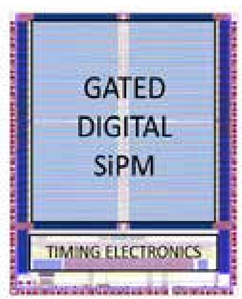

Fig. 2 - Layout of time gated digital SiPM with designed timing electronics.

Fig. 3 - Photograph of circuit board with bonded digital SiPM IC and prism mounted on top for the SOLUS optode.

A MORPHOACOUSTIC APPROACH TOWARDS THE HEAD-RELATED TRANSFER FUNCTION PERSONALIZATION

Muhammad Shahnawaz – Supervisor: Prof. Augusto Sarti

The head-related transfer function (HRTF) describes the transfer characteristics of a sound waves as they travel from a sound source placed at a certain position in the space to the ear canal in free space. These transfer functions greatly depend on individual's head, torso and ear morphology and are highly idiosyncratic in nature.

The knowledge of these individualized acoustic transfer functions is crucial to present personalized 3D audio through binaural rendering. This thesis study aims to build on the currently available knowledge and propose some simple HRTF personalization methods which let one estimate the individualized HRTFs on the bases of morphology knowledge without running expensive and laborious acoustic measurements or BEM simulations. This thesis aims to do so by widening the knowledge base for HRTF personalization process and create a better understanding on relationship between the morphology of the listener and the corresponding HRTFs.

This work is a composite study of many fields and concepts and involving simple signal processing techniques such as spectral analysis and Principal Component Analysis (PCA), the physics of Numerical Simulations like Fast-Multipole Boundary Element Methods (FM-BEM), and functional space analysis of shapes like Large Deformation Diffeomorphic Metric Mapping (LDDMM) in this thesis use the simple digital signal processing techniques such as Notch Extraction, Sparse Representation based modeling, Large Deformation Diffeomorphic Metric Mapping (LDDMM) and Kernel Principal Component Analysis (KPCA). The works done in this thesis study can be divided into two main groups. The first set of works provides some preliminary studies which can be used to personalize the HRTFs on the basis of anthropometric data. These studies are mainly performed on CIPIC database, while the second set of the works presents the approaches which use the 3D morphology of the data and use a morphoacoustic approach in order to better understand the relationship between the outer ear shapes and the corresponding acoustics by studying the variations in both spaces separately and then finding a mapping between both spaces using some kind of mapping techniques. All these studies are performed on the SYMARE database.

There are two studies in the first group. The first study provides a statistical analysis of the center frequencies of primary notches in the HRTFs and analyze how these notches evolve as a function of elevation angle. For this purpose, the notches of the median plane HRTFs of both databases are extracted using simple digital signal processing techniques. The notch frequencies are then clustered into three clusters using k-means considering each of the notch corresponds to one of the contours of the outer ear shape. The means of the clusters show how the notch frequencies evolve over time for the whole database. Finally, through a comparison process these cues are compared amongst both databases as well as for left and right ears for the same databases. The results suggest that there are some binaural cues also present in the median plane and can aid a listener to localize a sound source.

In the second study in first group a preliminary HRTF personalization study based on sparse representation is presented. This study is based on a very strong assumption that a very same sparse representation can be used for both anthropometric features as well as the corresponding HRTFs. However, the study presented in this work is different from the traditional studies of this sort for two reasons. The first reason is that it provides a simple statistical analysis to understand which of the anthropometric parameters are more relevant for creating the sparse combination and the second that it includes a method for incorporating the relative importance for every anthropometric parameter used in the study. That is why we named this study as weighted sparse representation. Furthermore, this compares the results of the method with some famous closest-match based personalization schemes and show that it outperforms the previous techniques.

In the second group of studies the first work analyzes the effects of affine transformations of the ear shapes on the corresponding HRTFs. As a counter product this study creates a synthetic database from SYMARE (one of its own kind), which we call affine models for SYMARE population. For the affine models the ear shapes are affine matched with the template ear shape to have same scale, orientation and position and then attached to the template head and torso shapes. These affine matched ears were firstly created to study the morphable model of the ear shapes on the bases of KPCA.

The affine model has many benefits, 1) it creates a simplistic paradigm to study the morphoacoustics of the ear shape by limiting the variations to only shape variations of the ear shape removing all the variations due to different head and torso, ear size, ear rotation and position. Second, it simplifies the process of modeling the ear shape as one has to model the shape variations only using LDDMM and KPCA not the scale and rotation. Third it supposedly simplifies the modeling process of the acoustics as all the ear shapes are at the same scale, position and rotation and are placed on the same head and torso shape. This study provides an analysis on how simple corrections such as rotation of HRTF directivity patterns and scaling of frequencies can compensate for these affine transformations providing a close estimate of the HRTFs of actual ear shapes on actual head and torsos. Finally, the study calculates the optimal frequency scaling factor from purely acoustic point of view which matches the the affine modeled HRTFs to the original HRTFs in the best way. These optimal scaling factors are then related to the physical scaling factors by using linear regression. The results show these scaling factors can be inferred simply by knowing the ear shape scaling factors coming from the affine matching process.

The second study in this group provides a simple Spatial Principal Component Analysis (SPCA) based modeling method to analyze the variations in the acoustic directivity patterns of the HRTFs as a function of frequency. The directivity patterns of different frequencies are modeled separately and the number of principal components required to model the directivity patterns for a given frequency are quantified for a frequency range from 0.2-17 kHz. This study reasserts the importance of the affine models by showing that the directivity patterns of the affine models can be modeled by using only eight principal components of even 17 kHz with an average standard spectral difference (SDD) of less than 3 dBs. Furthermore, this study uses these morphable model of the ear shapes and analyses how much variation in the ear shapes can be captured using first eight shape principal components. Finally, these shape components are related to the acoustic principal components through linear regression to provide a simple personalization method for HRTFs.

Finally, in the last study a novel idea of morphological weighting to create a weighted morphable model for ear shapes on the basis of Weighted Kernel Principal Component Analysis (WKPCA) and the LDDMM is presented. This study reports a simple yet a very interesting tool which not only improves the personalization of the HRTFs but also present a wonderful way of identifying the contributions of different parts of ear shape to different spatial cues of HRTFs and can be a variant of morphoacoustic perturbation analysis.

LEARNING-BASED MODEL PREDICTIVE CONTROL: THEORY AND APPLICATIONS

Enrico Terzi - Supervisor: Prof. Riccardo Scattolini

The thesis deals with learning-based Model Predictive Control (MPC) design, discussing both identification and control synthesis, and it is composed of three main parts: the first and the second one are more focused on theoretical aspects and algorithms, while the third one deals with a real case study from process industry.

PART I: Identification and control algorithms

In part I we discuss the derivation of a novel “identification & control” algorithm for Single Input Single Output (SISO) linear systems where the interlink between the two phases is tight, and the discussion flows from the dataset to the working controller, that is inspired by tube-based philosophy. At first, specifically in chapter 1, the Set Membership (SM)-based design of “multistep” prediction models is presented. The idea is to consider independent models tailored to predict the p step-ahead output variable without iterating a unique one-step ahead model. This novel technique allows not to increase the computational burden, that only requires the solution to Linear Programs (LP), and at the same time to consider a longer horizon in the identification. Also, thanks to the use of Set Membership tools, an error bound for each predictor will be defined, provably smaller than the one associated to any iterated 1-step model. The estimate of this bound is obtained from data, and it is proved to converge to the theoretical value as the number of data and information content increases. Then, in view of robust control design, we rely on the Feasible Parameter Sets (FPS) previously introduced to derive a model uncertainty associated to the 1-step ahead model. Also, multistep predictors are included in the regulator, tailored for tracking piece-wise constant references, given their lower worst-case error bound. In Chapter 2, the extension of the algorithm to the Multi-Input Multi-Output (MIMO) case is discussed and tested on a realistic benchmark, i.e. the quadruple tank process, together with a preliminary extension to the nonlinear case. Finally, in Chapter 3, thanks to the multi-step models, a nonstandard state space form of the system is defined, that evolves over a multiple of the sampling time, i.e. taking a “long” step composed of several “short” steps at each state iteration. With these ingredients a tailored multi-rate control scheme is investigated for the regulation problem, taking advantage of such state space form and of the independent p-steps ahead model as output equation, in order not to lose accuracy.

PART II: Model Predictive Control of Neural Networks

The aim of Part II is to investigate quite recent and powerful tools such as Neural Networks for identification and control, thus addressing also nonlinear systems. The use of Neural Networks for time-series prediction in different scientific fields has been discussed in the literature, as well as for pattern recognition problems; also, NNs have been applied for image, speech, text and music classification, and some control applications are present. A common trait of the vast majority of the literature about Neural Networks is the prevalence of empirical and experimental evaluation against theoretical analysis, that is hindered to simple cases. In the thesis we focus on Recurrent neural networks, and specifically on a couple of architectures proved to be able to overcome the well-known “vanishing gradient problem”, namely the Echo State Networks (ESN), and the Long Short-Term Memory networks (LSTM). Though these networks found applications in several fields, quite surprisingly, their use for control purposes, and specifically as predictive models in MPC schemes, has been rather limited, essentially in a couple of main contributions. Also, their theoretical analysis and properties as nonlinear systems have not been systematically characterized, especially with control perspective, and this hinders their applicability for control even more. For these reasons, in Part II of the thesis, specifically in Chapters 4 and 5, we address both ESN and LSTM networks from a system theoretical standpoint, establishing sufficient conditions to guarantee the incremental input-to-state stability property (δISS). These conditions are enforced during the training. Moreover, in the case of ESN, we provide a theoretically sound algorithm for dimensionality reduction based on Kalman decomposition, and study an alternative training algorithm with probabilistic guarantees based on the Scenario theory. Then, we use the network as a prediction model, and design a tailored observer with guaranteed convergence of the state estimates. Finally, a number of predictive control schemes both for regulation and tracking problems is devised, some of which are endowed with stability guarantees. The general block scheme of the considered control architecture is represented in Figure 1. The capabilities of these networks are tested on a simulator of a pH neutralization process, that is a standard nonlinear benchmark in the literature. Also, we use the LSTM prediction model and control scheme, which proved to be superior, within a case study, that constitutes the third part of the thesis, and is described next.

Part III: applicative case study

The third part is about a cooling station serving a large business center located in Milan, northern Italy. The system is composed of four chillers with on/off commands and an absorption chiller, that exploits waste heat coming from external utilities, working to provide cooled water to a set of users, see Figure 2, where only one chiller is reported for space limitations. The chillers are currently controlled with a relay-based controller, while the absorber follows an internal and non accessible logic. The aim of the case study is the optimization of the performances of the control system in terms of energy saving without compromising the demand satisfaction. Unfortunately, a lot of information is missing about the plant, thus the modelling comprises a mix of grey and black box techniques, inspired by the physics of the systems. As anticipated, LSTM networks are then employed as a learned model of the system, and embedded in a full predictive control scheme, see Figure 1, with a dedicated algorithm that is eventually tested in simulation. Notably, this learning-based approach cuts dramatically the modelling time, at the price of hiding the “physics” of the system inside the LSTM’s architecture, and thus losing perception on “what is going on” in the model, and provides significant performance in simulation.

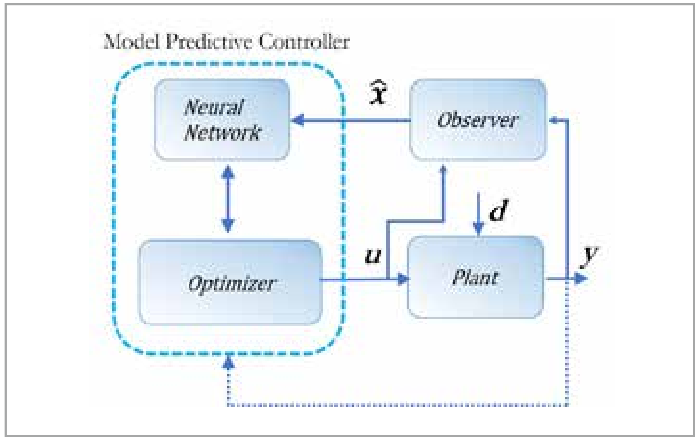

Fig. 1 - Block scheme of the Model Predictive Control scheme embedding recurrent Neural Networks as prediction models. u is the input, d the disturbance, y the output and the estimated state.

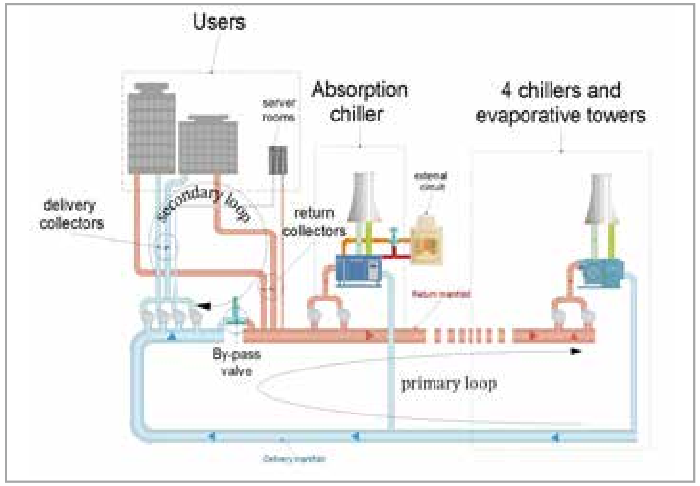

Fig. 2 - Scheme of the cooling station

VELOCITY ON THE WEB

Riccardo Tommasini - Advisor: Prof. Emanuele Della Valle

Starting from Smart Cities, to news analysis, many World Wide Web (Web) applications
shows the need for processing data reactively and continuously. This need appeared first in the Big Data context with the name of Data Velocity, i.e., the challenge of processing data as soon as they are available, and before it is too late.

The most significant hindrance to address Velocity on the Web is that the challenge does not appear isolated. Indeed, the Web is a decentralized environment, and Data Variety, i.e., the challenge of processing heterogeneous data coming from many different sources, is spread.

Data Variety aggravates the complexity of the analysis, which becomes a significant obstacle when insights must be reported in a timely fashion. Indeed, Data Variety poses a data integration challenge. Taming Data Variety and Velocity simultaneously requires to take into account background knowledge, identify relevant domain-entities, and qualify/quantify the relationships between these entities, with strict temporal constraints.

Problem Statement, Research Question, and Research Problems

Resources are the central abstraction of the Web. A global naming system called Uniform Resource Identifier (URI) is used to identify resources. Several files and data formats are used to represent resources. The Hyper-Text Transfer Protocol (HTTP) is the standard protocol that regulates the interactions with resources.

A sign that Data Velocity is permeating the Web comes from the re
cent evolution of the Web's underlying data infrastructure. Extensions of HTTP that are designed to deal with Data Velocity are becoming popular. HTTP Streaming and Web Sockets are designed to provision data continuously; Server-Sent Events and WebHooks are designed to supply data reactively. This PhD thesis henceforth refers to all these protocols together as HTTP*.

The Web does not put any restrictions to what a resource can be, and it prescribes that technologies can be invented and evolve independently. This approach allowed the Web to reach its current scale. However, Web resources result decentralized, heterogeneous, and incomplete.

The Web of Data (WoD) is an extension of the Web that addresses Data Variety. WoD encourages data sharing, and it provides a technological stack that enables interoperability. Foundational WoD technologies include but are not limited to (i) the Resource Description Framework, which is a semi-structured data model; (ii) the RDF Schema and Web Ontology Language, which are knowledge representation languages to build and share vocabularies; (iii) SPARQL , which is a protocol and a query language to interact with RDF resources. Like the Web, WoD is a dynamic environment, but its key technologies, i.e., RDF, OWL, and SPARQL, are not designed to tame Data Velocity. Thus, Web applications that present both variety and velocity-related requirements must choose what to address first.

Problem Statement. Web applications need to tame Data Velocity without neglecting Data Variety.

To solve this problem, we must leverage on the previous attempts to address Data Velocity despite Data Variety. In the literature, Data Streams, or simply streams, and Events are the key abstractions that allow applications to perform continuous and reactive analyses. The research area is known as Stream Processing (SP), and like WoD, it relates to the entire data infrastructure.

The technologies that constitute the Web architecture are designed according to
three architectural principles, i.e., *Simplicity, Extensibility, and Orthogonality,* and organized into three areas, i.e., **identification, representation, and interaction***.* on the Web, Data velocity affects all of them. In particular, it requires to introduce new kinds of resources, i.e., streams and events, that are typical of the stream processing domain. The following research questions summarize the research goal.

Research Question. Can we identify, represent, and interact with heterogeneous streams and events coming from a variety of Web sources?

Nevertheless, this research question identifies a macro problem that is too broad to be solved. This problem includes a number of challenges related to the Web technological areas.

The challenges related to **identification** refer directly to the Web resources. Indeed, a resource is everything that can be identified using a URI. URIs are the means to a global naming system that sustains a dense global network. Web agents leverage on such network to identify and interact with resources' representations.

The challenges in the scope of **representation** refer to the Linked Data principles. On the Web, the role of a representation is to allow the agents that access and exchange it to understand each other. Representing resources implies providing a model and a description: a model establishes a contract on the resource interpretation; a description mediates the communication between a resource owner and an applicant. If the description is machine-readable, owner and applicant can be autonomous agents.

The challenges in the scope of **interaction** refer to protocols, APIs, and languages, to access and manipulate streams and events. The literature on interaction on the Web is vast. However, for what concern this PhD thesis, interaction is limited to access and process.

On the one hand, *accessing* indicates protocols that allow obtaining the data of interest. On the other hand, the term *processing* is still too generic. Thus, it is necessary to clarify what it means in the context of this PhD thesis, i.e.,

- <u>Analyzing</u>, i.e., filtering, joining, and aggregating input flows to produce one or more output flows.
- <u>Detecting</u>, i.e., matching trigger conditions on the input flow and take an action that may combine input items to produce new information.
- <u>Reasoning</u>, i.e., deducing implicit information in the input flows by leveraging on the data representation and using inference algorithms.

From the attempt to restrict the problem scope and framing a more precise research goals, it emerges that work in the area of identification is stable. Moreover, new Web protocols have addressed the challenges related to stream/event data access. Thus, this PhD thesis takes streaming/event data access as a solved problem.

On the other hand, research on representation still presents many open challenges. In particular, challenges related to conceptual modeling of streams and events as resources. This PhD thesis refers to these problems as the representation problem.

The **Representation** calls for improving the Web of Data by enabling mod- eling and description of Web streams and Web events, satisfying the requirements of users interested in representing them.

Research on interaction still presents many open challenges. A lot is left to do in terms of querying and expressive reasoning. In particular, the problem of designing more expressive data-manipulation languages for the Web streams. This PhD thesis henceforth refers to all these problems as the processing problem.

The **Processing Problem** calls for improving the Web of Data by enabling expressive yet efficient processing of Web streams and events.

Finally, recent work highlights the lack of a systematic approach to validation research for Web Stream Processing. Two issues call for yet another research referred as the validation problem: (i) existing prototypes are not designed for comparative research, and (ii) existing benchmarks lack a shared experimental methodology.

The **Validation Problem** calls for improving validation research for the processing problem by enabling a systematic comparative exploration of the solution space, which satisfies re-searchers' requirements.

Conclusion

This PhD thesis investigates the aforementioned problems using the Design Science (DS) research methodology. DS studies how artifacts, i.e., software components, algorithms, or techniques, interact with a problem context that needs improvement.

MACHINE-LEARNING DEFINED NETWORKING: APPLICATIONS FOR THE 5G METRO-CORE

Sebastian Troìa - Supervisor: Prof. Guido Maier

With the advent of 5G technology and an ever-increasing traffic demand, today Communication Service Providers (CSPs) experience a progressive congestion of their networks. The operational complexity, the use of manual configuration, the static nature of current technologies together with fast-changing traffic profiles lead to: inefficient network utilization, over-provisioning of resources and very high Capital Expenditures (CapEx) and Operational Expenses (OpEx). This situation is forcing the CSPs to change their underlying network technologies, and have started to look at new technological solutions that increase the level of programmability, control, and flexibility of configuration, while reducing the overall costs related to network operations. Software Define Networking (SDN) and Network Function Virtualization (NFV) are accepted as effective solutions to reduce CapEx and OpEx and to boost network innovation. Although the implementation of SDN and NFV in networking gained big momentum in the last years, it also brings a whole new level of complexity. Virtualization breaks traditional networking into dynamic components and layers that have to work in unison and that can change at any given time. The high complexity introduced by these new technologies has led to the research for increasingly smart algorithms to optimize the network resource allocation. This thesis investigates new Machine Learning (ML) based algorithms in order to efficiently optimize resources in 5G metro-core SDN/NFV networks. The main goal is to provide the modern CSP with intelligent and dynamic network optimization tools in order to address the requirements of increasing traffic demand and 5G technology. The present study can be divided in two main activities: 1) propose novel ML algorithms in order to optimize the resource allocation of 5G metro-core networks; 2) implement the proposed algorithms in different emulated and real network scenarios. The first activity starts by investigating the three branches of ML, that are: supervised, unsupervised and reinforcement learning. Exploiting different kind of network data, we develop supervised learning algorithms in order to predict network traffic in mobile-metro-core networks in order to optimize resource allocation in advance and to proactively route connections. Then, we focus on unsupervised learning algorithms in order to identify typical traffic patterns in SDN/NFV mobile-metro networks. Taking advantage of this patterns, the goal is to predict how and when to optimize the resource allocation in the underlying physical network. After investigating supervised and unsupervised techniques, we focus on reinforcement learning techniques with the aim to build an intelligent system capable of optimizing the network resource allocation given the actual and historical observations of the state of the network. Finally, we implement ML-based algorithms on different emulated and real SDN/NFV testbed scenarios. First, we develop the network planner module under the framework of an European project called Metro-Haul, in which ML-based algorithms interact with the control plane of the proposed SDN/NFV orchestration infrastructure. Then, we develop an experimental SDN/NFV service orchestrator called SENATUS, able to host and quickly test various ML algorithms. We present two testbed scenarios to show the capability of SENATUS to integrate both Information Technology (IT) and networking resources. As last, we develop different ML-based performance monitoring tools for Enterprise Networking (EN) by means of Software Defined Wide Area Networking (SD-WAN) technology. It represents a revolutionary way to simplify Enterprise networks and achieve optimal application performance using centrally managed WAN virtualization. In particular, we present two testbeds based on open source components with the aim to test classical and ML-based performance monitoring algorithms.

A STUDY OF HIGH-PERFORMANCE FREQUENCY SYNTHESIZER BASED ON DIGITAL BANG-BANG PHASE-LOCKED LOOP FOR WIRELESS APPLICATIONS

Tuan Minh Vo - Supervisor: Prof. Salvatore Levantino

High-performance frequency synthesizer is a fundamental part of almost any modern wireless communication device, for example, used for coherent demodulation/modulation in wireless transceivers. The frequency synthesizer based on phase-locked loop (PLL) architecture, serving as a local oscillator in a transceiver, is indeed a negative feedback control system generating an output signal whose frequency is multiple of the reference signal frequency. The multiple can be an integer or a fractional number. Though fractional-*N* PLLs entail the key advantage of a finer frequency resolution, the noise-power figure-of-merit (FoM) of state-of-the-art integer-*N* PLLs is still better than in the fractional-*N* case. In addition, digital PLL synthesizers are taking over conventional analog ones, because of their benefits in terms of power consumption and area occupation in ultra-scaled complementary metal-oxide-semiconductor (CMOS) technologies. The digital solution simplifies the design and, as this is portable to the next technology nodes, may potentially reduce the time-to-market. In this study, a type-II fractional-*N* digital PLL having a phase detector (PD) with only two digital output levels of 1 and −1 is of interest. Though this topology of fractional-*N* digital PLL has been demonstrated in practice being able to obtain a FoM close to the best of the integer-*N* ones, there is a lack of theoretical literature that explains in detail adopted techniques as well as algorithms in the system.
It is known that the one-bit (also known as bang-bang, BB) phase detector, when employed in frequency synthesizers, only acts like a linear element when phase (time) error at the PD input is dominated by random noise. Since the deterministic quantization noise of the digital Delta Sigma modulator (DSM) dithering the modulus control of the frequency divider is not white or/and much larger than other thermal random noise in the system, this noise may cause a limit cycle in the BB-PLLs. To address this issue, a digital/time converter (DTC) is placed in the feedback path between the frequency divider and the BBPD. The control gain of the DTC is automatically adjusted in background by a calibration loop operating based on the principle of the least-mean-square (LMS) adaptive filter. The calibration loop helps the PLL to adapt to changes in the digitally-controlled oscillator (DCO) period as well as the DTC characteristics in practice. By far, in the presented publications, to guarantee a short convergence time, at least second-order DSM is required when a fine frequency resolution is desired. This follows a large DTC time range of twice of the DCO period in the system and this compromises jitter performance. In addition to the issue related to the quantization noise, the BB-PLLs also face an extremely long transient process when a large jump of the output frequency is required. Indeed, the frequency locking time of the BB-PLLs is shown to be inversely proportional to the values of the loop filter gains while these gains are very small to guarantee the loop stability. To solve the problem, a frequency-aid technique has been proposed. This technique is essentially based on exploiting digital ternary phase detectors (TPDs) to create multi filters in the feed-forward path of the BB-PLL. The outputs of the filters are used to tune a multi-bank DCO in corresponding order. The frequency-aid circuit is only triggered when the time error at the BBPD input is larger than a fixed value, i.e., the dead-zone of the TPD. In the published report, the lower bound of the dead-zone is large which may not be optimum for the frequency locking time and affect to the phase noise performance of the system. Furthermore, even when this technique is adopted, the frequency locking transient in the conventional BB-PLL still requires a relatively long time in cases the DCO tuning words are at their worst conditions.
The objective of this study is to give an insight, for the first time, into the operation of the fractional-*N* digital BB-PLL during the transient via the behavior of the time error at the BBPD input. The study is carried on in two separated cases, i.e., with the frequency aid technique in the first case and with the LMS calibration loop in the second one. In order to reach this goal, analysis is carried on in the time domain for the frequency aid technique. Then, based on the analyzed result, we propose a novel frequency aid technique to further improve the frequency locking speed. In the worst case of the frequency locking, the proposed technique reduces the transient time by a factor of 3.5. The LMS calibration loop is evaluated versus the value of the fractional part of the frequency control word and the order of the above mentioned DSM. The analysis, that is carried on in both the time-domain and the *z*-domain, not only explains in detail the operation of the LMS loop but also gives quantity results of the DTC time range and the noise induced by the LMS loop. Moreover, two novel calibration schemes are proposed in order to use a smaller delay range DTC while keeping a short convergence time. At the same convergence time, the required DTC time range in the first proposed scheme is 0.57 times, and, the one in the second proposed scheme is 0.55 times as of the ones in the conventional schemes, respectively. All of analysis is verified by simulation based on accurate behavioral models. The models are built with real design parameters, and, designed for the BB-PLLs synthesizing an output frequency from 3.2 GHz to 4.0 GHz from a reference frequency of 52 MHz.