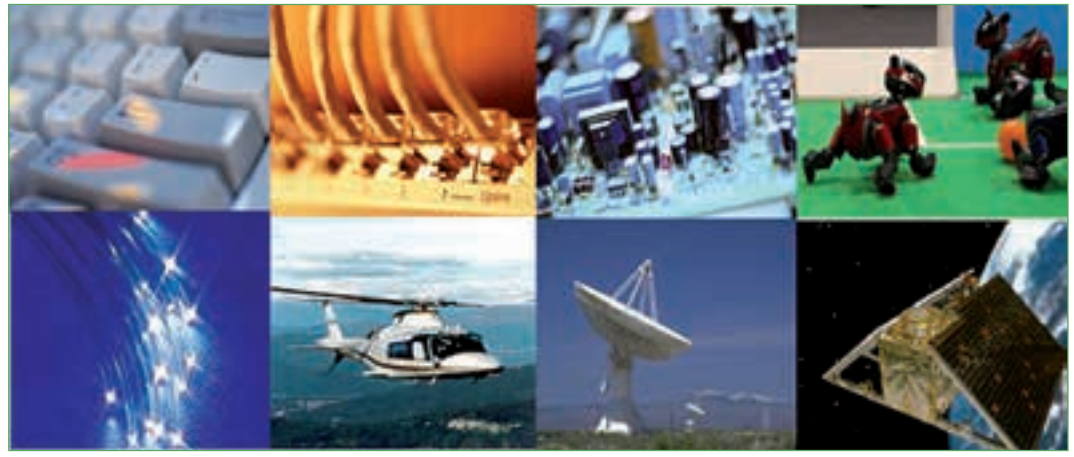
Chair:
Prof. Carlo Fiorini

DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY



Introduction

The PhD program in Information Technology (IT) goes back to the year 2001, when the two traditional programs in Automation-Computer Engineering and Electronics-Telecommunications were merged. As such, the unified course covers the research interests in four scientific areas, namely Computer Science and Engineering, Electronics, Systems and Control, and Telecommunications. This broad variety of research activities is completely focused in the ICT area, and perfectly corresponds to the core mission of the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB). However, pursuant the history of the Department, and following the new trends of the modern society, some cross-related research fields are also encouraged, such as ecology, environmental modelling, operations research, and transportation systems. The PhD program in IT is the largest at the Politecnico in terms of number of students. There are more than 60 first year students and about 210 in total. The students are subject to an examination every year to evaluate the progress achieved in their research and course work.

Topics

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

Industrial collaborations

Due to its intrinsic technological nature, the PhD curriculum is corroborated by many industrial collaborations. About 25% of the total number of scholarships are funded by industry or by international research projects involving industrial partners. In the school vision, the collaboration between university and industry is ideally based on the challenge of turning invention into technological innovation. This amounts to shaping new technology frontiers and to building a fertile atmosphere for a mixture of world-class research at universities and in private companies. This also amounts to creating a common terrain of friendly culture, to size the risk and to believe in strong basic research. The external referee board is composed by members of public and private companies, working in industry and in applied research. The board is in charge of monitoring the activities of the PhD program and giving suggestions for its development. The board meets once a year to point out the new emerging research areas worth to be investigated and to monitor the visibility of the course in the industrial world.

Educational aspects

The teaching organization and subject of the courses reflect the scientific interests of DEIB faculties. The curricula include a wide choice of courses (about 30 per year), of different nature. The challenge is to promote interdisciplinary research while offering technical advanced courses that spur innovative and cutting edge research. Therefore, particular attention is devoted to help each student to make the best choice according to an internal regulation scheme.

Internationalization

Every year about 5 courses are delivered by foreign professors. Moreover, the PhD program encourages joint curricula through agreements with foreign institutions. At present we count joint agreements for a Double PhD Program with New Jersey Institute of Technology (NJIT) in the electrical and computer engineering disciplines; academic cooperation for a Double Doctoral Degree with the Graduate School of Engineering and Graduate School of Life and Medical Science, Doshisha University, Japan; double PhD Program with Queensland University of Technology, Australia; between DEIB and the Drexel University, Philadelphia, Pennsylvania, USA; Double PhD Program with the University of Western Ontario, Canada; Joint PhD Degree Program between DEIB and Georgia Institute of Technology – Atlanta USA; Agreement for co-supervison of a doctoral Thesis with Ecole Polythechnique de Montreal – Canada.
The PhD program in Information Technology participates in ICO-NEH (International Curriculum Option of Doctoral, in Erasmus mundus Programs: STRONG-TIES Strengthening

Training and Research Through Networking and Globalisation of Teaching in Engineering Studies and GREEN IT for the benefit of civil Society (North Africa – for incoming and outgoing mobility) and a Cooperation agreement with "Escuela superior Politecnica del Litoral" (Ecuador) for admission of Doctoral candidates. Some activities are organized within the Network of Excellence EURO-NF (Network of the Future) and of ACCM, Austrian Centre of Competence in Mechatronics.

Every year at least 4 fellowships are assigned to foreign students, selected on the basis of their CV and research interest.

Conclusions

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

Prizes and awards

In 2014 the following awards have been obtained by PhD students:

Chorafas Foundation Award: Silvio Del Giudice, Valerio Panzica La Manna
Best Ph.D. in Information Technology Award: Federico Pepe
IBM Ph.D. Fellowship Award: Daniele Dell'Aglio
ICWE 2014 Best Paper Award: Andrea Mauri
IEEE WETICE 2014 Best Paper Award: Mahsa Teimourikia

VULNERABILITY DETECTION AND COUNTERMEASURES IN BUILDING AUTOMATION NETWORKS AND CYBER-PHYSICAL SYSTEMS

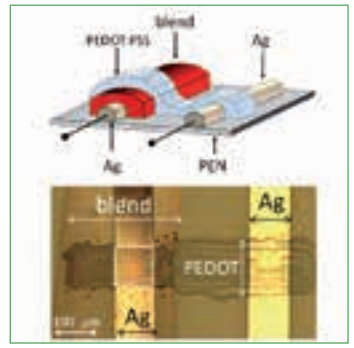
Alessio Antonini - Supervisor: **Luca Breveglieri**

Cyber-Physical Systems (CPSs) are systems where software and hardware entities monitor and manage physical devices using communication channels. They have become ubiquitous in many application domains, including health monitoring, smart vehicles and energy efficiency, as in the power supply provisioning and management (e.g., smart buildings and smart grid infrastructures), thanks to the rapid growth of the embedded system technologies.
On a more general perspective, these systems bridge the cyber-world of computing and communication with the physical world of the more complex systems, which include mechanical parts, sensors and actuators of different kinds. Examples of CPSs include the smart electricity-grid systems, the Supervisory Control And Data Acquisition systems (SCADA) for managing the functioning of industrial plants, and the distributed control systems of Building Automation Networks (BANs) – designed to monitor and control the security and safety of the mechanical systems in a building, as well as their services.
All the aforementioned systems were commonly designed as stand-alone networks in physically protected locations, using proprietary technologies. The extension of their connectivity and the improvement of their management into more efficient and effective CPSs, paved the way also to new security and safety threats. In particular, the security of CPSs involves the need to secure a set of systems that were explicitly designed assuming the existence of a physical barrier to confine the attackers, which has fallen when the need to connect them with the external world has surged. This is particularly problematic as CPSs have a multitude of possible attack vectors, by their own nature, since a number of malicious actions can be undertaken at every level of their programming and working stack, from the low-level firmware residing on the devices, to the infrastructure coordinating their functionality, which typically resides on common general-purpose computers. To this end, it is crucial to approach the securization of CPSs at multiple levels. In particular, it is fundamental both to tackle the communication level security of such systems, by providing authentication to the involved endpoints and strong integrity guarantees on both the transmitted data and control signals, and to tackle the system security, lest an attacker can alter their functioning for malicious purposes. The research and industrial communities have developed several solutions to tackle with potential threats, but these solutions lie far from being unaffected by defects, as it is shown in this thesis.
In this research work, the attention has been focused on the CPSs that operate in a smart grid, considering that the introduction of these so-called smart devices has particularly attracted the attention of the cyber-security researchers, mainly due to the large scale impact that would occur in the case of a cyber-attack. There are many different CPSs that are worth being investigated; nevertheless, this research focuses on two of them, which appear to have a major role in the smart grid field: the BANs, which can operate at the peripheral level of the grid, and the PLCs, which are widely used to manage the generation and distribution of energy.
This thesis pursues several objectives, at different levels of complexity.
First, different technologies have been investigated for both the BAS and SCADA presented in the literature, in order to understand what kind of communication protocols, operating systems, and hardware are used. Each technology may have specific vulnerabilities that cannot be generalized. About the BAS architecture, the thesis focuses on the analysis of network protocols for BANs such as KNX, LonWorks, BACnet and Modbus; about the SCADA systems, it focuses on the Programmable Logic Controllers (PLCs) architecture.
Second, the previous technologies have been summarized and a qualitative comparison among them has been carried out.
Third, relying on the previous analysis, two case studies have been defined in order to better investigate the vulnerabilities of such systems. A first testbed has been realized using a KNX system as most representative because of the more stringent constraints of computational capacity, and a second one for the SCADA systems, based on Siemens PLCs as most representative because of their spread in the industrial market.
Fourth, after an extensive activity of reverse engineering, a set of viruses has been developed to exploit the identified vulnerabilities, trying to generalize the attack scheme. A first malware has been designed to attack a KNX system and to show the possibility of fully reprogramming the network devices, in order to obtain a malfunction or a complete Denial-of-Service. A second malware has been designed to attack the application program of Siemens PLCs, in order to show how it is possible to change the specific industrial process running on the PLCs. In the worst case scenario the malware can shut down the entire process. Both malwares have been tested on realistic testbeds, proving the feasibility of the attacks.
Fifth, some countermeasure have been proposed for each scenario. These countermeasures must exploit the attack schemes previously studied and must be resistant to several classes of attack schemes. In the case of the BAN, a novel protocol, with strong security guarantees, has been proposed. The protocol is based on a logical time slotting and an ephemeral shared secret, which is refreshed periodically by means of a multiparty key agreement scheme. The computational complexity of the novel multiparty key agreement scheme has been shown to be equivalent to a computational Diffie-Hellman problem. This solution provides support for sending secure commands in 1.3 seconds, and performs a key agreement, providing forward secrecy, in roughly one minute of computation and transmission time on the most constrained BAN network infrastructure: KNX. Furthermore, these performance figures improve by more than an order of magnitude on faster BANs, thus providing a widely applicable solution, proved to be secure against the strongest possible network bound attacker model, with the same mathematical strength of well characterized crypto-systems.
In the case of SCADA, a novel defensive approach has been proposed, which is based on the application of model checking techniques at firmware level, to verify the integrity of the industrial process execution. It has been shown how to apply those techniques to a real case based on the application program of a water turbine.
A Safety Checker (SC) for the system has been written, which is able to intercept at firmware level the wrong commands transmitted by an hypothetical malware acting at application program level. The SC performs a real-time check between the values that the firmware is going to send to the physical devices and the actual status of the industrial process, in order to determine the admissibility of the future state. Using a simulation, it has been shown that the proposed solution can run in realtime with the firmware without affecting the general performances of the system.
In conclusion, this research shows how it is possible for a hacker to perform cyber-attacks to CPSs with a discrete knowledge of the physical devices and industrial processes. Thanks to the numerous weaknesses of the CPSs, it is possible to obtain the full control of industrial processes and to perform different actions in order to accomplish different goals, from a Denial-of-Service to a more complex manipulation of the physical systems. This thesis also shows that it is possible to use different techniques to maintain the safety and security of CPSs. The results of this research have advanced the state-of-art in the security of CPSs and BANs, and have been published in the refereed proceedings of research venues.

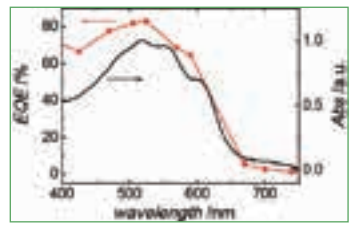
INKJET PRINTING OF ADVANCED FUNCTIONAL MATERIALS: TOWARD PATTERNED DEVICES FOR LARGE AREA APPLICATIONS

Giovanni Azzellino - Supervisor: **Prof. Marco Sampietro**

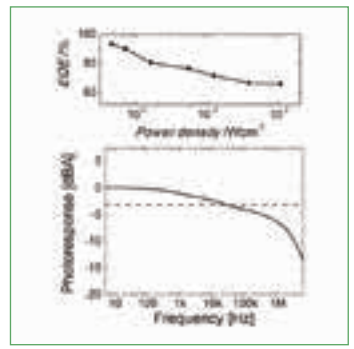
Solution processible semiconductor technology has attracted considerable research interest in the past few decades; in particular, devices based on conjugated polymers and semiconducting nanocrystals, both processible from solution, can pave the way for the development of low-cost, large area and flexible electronics.
The key advantages of these materials over crystalline semiconductor technology are mainly ascribable to their low-temperature manufacturing and compatibility with flexible substrates, and their unique and tunable optical capabilities.
The high absorption coeffcient of conjugated polymers makes them suitable for effcient light harvesting devices. Further, the quantum confinement effect in semiconducting nanocrystals is particularly useful for making both devices with narrow and pure emission and devices harvesting light at a certain wavelength.
In this work we focused, primarily on the realization of photosensors based on conjugated semiconductors, secondly on novel LEDs with pure and narrow emission, based on quantum dots. In particular, we addressed the need of patterning functional materials, coming from solution, to manufacture many devices on the same substrate and envision the realization of more complex electronic devices (even bendable) for imaging/lightening purpose with a pixilated structure.
Inkjet printing technology, which allows ambient condition and room temperature manufacturing, meets all requirements for patterning solution-based semiconductors and manufacturing our devices. Furthermore, the additive nature of this deposition technique is virtually compatible with whatever substrate, even flexible.
In particular, we worked with the piezoelectric inkjet printing: functional solutions were loaded into tiny nozzles that are able to propel small amounts (tens of pL) of the same solutions in response to an electric voltage pulse. Control over the jetted solution is mandatory in order to obtain well-defined structures and high resolution features (~ 10 μm): that is mainly achieved by properly tuning the voltage waveform which drives the piezoelectric nozzle.
Nonetheless, the major hurdle in depositing films by inkjet printing relies in the uniformity and the continuity of the films themselves. In fact, casted films suffer from the coffee stain effect which gives rise to non-uniform thickness and hollowed films. Noticeably, we addressed that issue and manufactured photosensitive devices: the adoption of tailored solutions, resulting from the mixture of solvents with different physical properties, contributed to avoid the coffee stain, pile up different layers, all of them by inkjet printing, and make optoelectronics devices with remarkable performances.
Finally, we reported pinholes-free, fully inkjet printed photosensors, based on a vertical 'inverted' architecture with the photoactive layer sandwiched between two electrodes, deposited on flexible substrate, with EQE (at 1V reverse bias) in excess of 60% over most of the visible spectral region, and a peak of 83% at 525nm. Noticeably, EQE and specific detectivity of our devices are comparable with the literature reported devices made by the more standard coating techniques.
In other words, the novelty of our approach, does not just prevent the manufacturing of high performance devices, but demonstrates the inkjet printing capability of patterning and making efficient electronic devices as well.
In-depth analysis of printed photosensors has enlightened the relationship between the manufacturing recipe and the device electrical properties; after an optimization step we were able to report devices with EQE independent on the incident power density (over a power density range larger than two decades), which is high desirable for photodetectors.
The same inkjet printing technology was adopted to pattern quantum dots for making LEDs. In that case, the piezoelectric inkjet is preferable over other printing techniques, like micro-contact or thermal, for two main reasons: the non-contact additive approach avoids material waste and mechanical contact between the dots and the stamp, and finally the piezoelectric driving prevents the heating of the dots-based ink that would be detrimental, in terms of quantum yield, for dots. We inkjet printed a solution based on core-shell dots PbS-CdS (emission peak at 1450 nm) in ambient condition at room temperature and we addressed the issues of coffee stain and thickness uniformity by tailoring the solution itself and modifying the surface energy of the underlying layer. We manufactured arrays of dot spots with 25μm resolution and finally demonstrated the electroluminescent behavior of devices incorporating arrays of printed dots as emissive layer.



1. Cartoon (top) and optical micrograph (bottom) of the inkjet printed photodetector with active area ≈ 100 x 100 μm². The dummy strip of silver is needed to minimize the resistive losses due to the transparent electrode based on PEDOT:PSS.



2. EQE spectrum (red) measured at 1V reverse bias under incident power density of 10 mW·cm⁻² and normalized active layer absorbance (black).



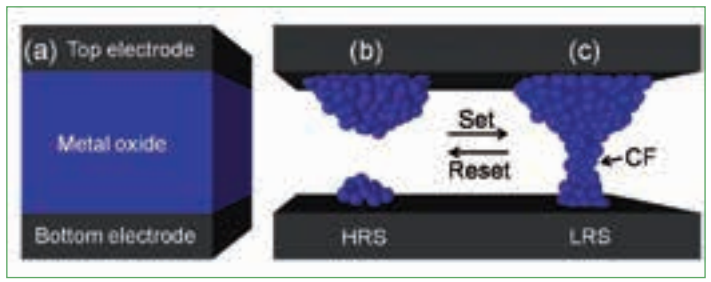
3. Top: EQE as function of the incident power density (measured at 570nm, at 1V reverse bias). Bottom: Device bandwidth measured at 570 nm under incident power density of 0.1 mW·cm⁻² at 1V reverse bias.

VARIABILITY, ENDURANCE AND NOVEL APPLICATIONS OF RESISTIVE SWITCHING DEVICES

Simone Balatti - Supervisor: **Ielmini Daniele**

In recent years there has been a huge increase in the storage capacity of integrated memories, still the technological solutions remained almost the same. The continuous demand of new portable, low-cost, low-power devices has forced a huge effort in R&D pushing the limits of the current technology. Flash memory, representing the main stream memory technology and experienced an impressive development, that has led this technology to a 16 nm node and to the implementation of 3D architectures. As we approach the scaling limit imposed by physics to Flash technology, the research is making more and more efforts towards the development of new devices that are able to combine the characteristics of Flash technology with the performances of faster memories, overcoming the scaling limits that will soon afflict the floating gate devices. At the same time, it is emerging the need of a new family of memory, called storage class memory, that combines the benefits of a solid-state memory, like high speed and robustness and the archival capabilities and low cost of conventional HDD. Among many competitors, oxide based resistive switching memories represent a strong candidate for next generation solid state non-volatile memory technology. In particular, the extreme ease of fabrication, the low power and energy consumption and the high write/erase speed suggest that this technology may became a valid alternative to Flash memories and an optimal choice for storage class memory.
The RRAM memory cell has an extremely simple structure, basically it is a parallel plates capacitor, where a metal oxide is sandwiched between two metal electrodes that act as top and bottom contacts, as shown in Fig. 1a. The application of a positive voltage to the top electrodes induces a ion migration from the top electrode toward the bottom electrode that creates a conductive filament that connects the two electrodes, as shown in Fig. 1c. This state is characterized by a low resistance (it is called low resistance state LRS) and it is identified as the logic value of 0. On the opposite, the application of a negative voltage at the top electrode leads to ion migration toward the top electrode and the formation of a new gap, Fig. 1b. This final state shows a high resistance (it is called high resistance state HRS) and it is identified as the logic state 1.

This thesis is focus on the electrical characterization and the modeling of several key issues of the RRAM, namely the switching mechanism, the program variability, the endurance failure and the random telegraph noise. Starting from the concept of complementary switching, a concept recently proposed to solve the sneak-path in crossbar-array, an innovative multilevel scheme for oxide RRAM was developed. The new scheme relies on the storage of two different states for any resistance level, where the two states differ by the orientation of the CF. The defect concentration and orientation can be independently controlled by the current compliance and voltage polarity, respectively. Storage and discrimination of 8 states (i.e. 3 bits) is demonstrated by experiments and numerical simulations. These results support the high functionality of nanoionics in storing and elaborating information in metal oxides.



Switching variability of the set state was studied and two different models where provided: (i) an analytical model, for the Poisson fluctuation of the defect number in the conductive filament and (ii) a Monte Carlo model, which describes the discrete injection of defects. Both models can capture the dependence of variability on the compliance current. It is also evidenced a new set-failure phenomenon induced by complementary switching. Set failure might be suppressed by accurate choice of the programming voltage, time and pulse shape, as well as by a careful RRAM stack engineering. Pulsed operation of oxide RRAM was studied, showing that the reset pulse-amplitude V_{stop} controls resistance window and switching variability, the high resistance state distribution improves for increasing V_{stop}, while low resistance state is degraded due to capacitive overshoot and incomplete set. Endurance failure at high V_{stop} is due to negative set, i.e. a defects injection from the bottom electrode in to the depleted gap that induces a non switching state with low high resistance state and prevents a set transition. Changing the set/reset pulse-width causes a negligible change of maximum endurance, which is explained by an Arrhenius model of failure.
Starting from a detailed electrical characterization random telegraph noise (RTN) in RRAM devices was studied. The random fluctuation between two levels is explained by the change of charge state in a bistable defect close to the CF. The model provides a physical quantitative description of both the electron transport in presence of a fluctuating defect and the temperature-dependent switching kinetics. The model accounts for the size dependence of RTN amplitude, which is due to the partial or full depletion of carriers depending on the CF diameter, and for bias dependence of RTN switching. The work on RTN, presented at the IEEE conference IRPS, was awarded with the Best Poster Award.
Nonvolatile logic operation in RRAM through conditional switching in serially-connected devices was developed in this work. The RRAM state variable can be 0 or 1. The state variable is used both as input or output of the logic operations. AND, IMP, NOT and transfer can be achieved in a single clock pulse, while OR, XOR and all other operations are achieved in multiple computation steps. The new nonvolatile logic approach allows to suppress the static leakage power dissipation, while reducing the area consumption thanks to the scalable 2-terminal structure of the RRAM switch. The RRAM logic was studied also from the circuit viewpoint, discussing the implementation architecture, the select/unselect scheme to prevent disturb and the implementation of a 1-bit adder. Two architectures are considered, namely a 1T1R architecture, where RRAM in both the top and the bottom crossbar arrays are selected by a transistor, and a hybrid 1T1R-1R1R architecture, where only RRAM in the bottom crossbar requires a select transistor. Finally, a 1-bit adder is designed and demonstrated by experiments and simulations. RRAM logic appears as a promising alternative to CMOS technology for area and energy scaling thanks to the reduced area and nonvolatile behavior of RRAM devices.

TECHNIQUES AND TOOLS FOR EFFICIENT, QOS-DRIVEN WAREHOUSE-SCALE COMPUTING

Davide Basilio Bartolini - Supervisor: **Prof. Marco D. Santambrogio**

Warehouse-scale computing, which is supported by datacenters, emerged in the last decade as a fundamental enabling technology for pervasive phenomena such as the Web 2.0, big data, and cloud computing.
Despite being assembled from commodity components (servers, interconnects, ...), these datacenters opened the way to a new paradigm for mainstream computing; as researchers work on understanding this new paradigm, two important themes emerge in a new way compared to traditional systems.
A major concern for datacenter operators is their efficiency and cost-effectiveness, which are crucial to supporting the growth in the services and value coming from big data and cloud computing.
Additionally, public cloud computing presents further challenges for both datacenter operators and users. A major issue for users that want to bring their workloads to the cloud to take advantage of utility computing is that performance on virtualized resources is hard to understand and often unpredictable. For this reason, using public clouds for applications that need to provide a required quality of service (QoS) level is not straightforward and often leads to increased inefficiency due to conservative resource allocations.

There is a tension between these two issues (efficiency and QoS), as techniques to improve efficiency (e.g., virtualization, power management, colocation, ...) impact performance, often unpredictably.
This dissertation attacks both sides of this tension and proposes novel techniques and tools to help solve it, towards future efficient QoS-driven warehouse-scale computing.

First, we analyze a well-known model for the total cost of ownership (TCO) of a datacenter and find that, as things stand today, opportunities to further reduce TCO, and allow datacenters to scale further, mostly lie in improvements in the efficiency of IT equipment, particularly the efficiency of servers.
On this basis, there are three main opportunities to improve efficiency: increasing server utilization, reducing static power consumption, reduce dynamic power consumption.
The challenge is being able to target these opportunities without hurting QoS. We show that traditional mechanisms and policies to pursue these goals are not well-suited for datacenters: colocating applications causes inefficiency and performance degradation due to contention on shared resources; deep sleep states impose high transition latencies and flush shared state, impairing performance; traditional controllers for dynamic voltage and frequency scaling (DVFS) reduce active power, but can heavily impact performance, because they are oblivious to the peculiarities of datacenter applications.

Then, we analyze metrics to quantify the performance of datacenter applications and define their QoS.
Throughput is a general metric to quantify rate of progress or load, but it is not enough to capture the performance of latency-critical applications, such as user-facing services, which need to provide performance guarantees on the end-to-end latency of each request.
Latency-critical applications are particularly interesting, because they define an operating context that is peculiar to datacenters; we analyze the behavior of five latency-critical applications, studying how latency is affected by different operating conditions. One important consequence of defining QoS with application-level metrics is that traditional systems that optimize for aggregated, low-level metrics cannot provide this type of QoS guarantees.

The main contribution of this dissertation is proposing novel approaches to the problem of achieving QoS enforcement in an efficient way in two complementary scenarios.
We analyze these two scenarios and propose two methodologies and practical systems (AutoPro and Rubik) that solve this problem:

- AutoPro tackles on the problem of providing predictable performance with automated resource allocation in public infrastructure-as-a-service (IaaS) cloud computing. AutoPro provides a practical solution based on a control-theoretical background for systems running compute-bound, throughput oriented applications. With AutoPro, we focus on current hardware and propose a solution that is directly deployable on modern datacenters with no hardware changes.

Rubik analyzes datacenters running latency-critical applications, along with other batch work and tackles the problem of reducing the TCO while maintaining QoS guarantees on the tail latency, thus improving efficiency. Rubik provides a solution based on a runtime system and few key hardware changes, mainly to provide partitioning of the memory hierarchy; this solution could be implemented with negligible overhead on next-generation servers.

Both AutoPro and Rubik demonstrate the importance of three principles that we suggest as guidelines for the development of next-generation computer architecture and operating systems for datacenters:

- Availability through the hardware/software stack of application-level information is key for effective control.
- Control systems used to tune system-level knobs need to be founded on solid theoretical bases (e.g., AutoPro uses control theory, Rubik uses statistics and control theory); ad-hoc empirical controllers do not generalize well and often fail due to unpredictable pathological cases.
- In order to support the dynamic execution context of datacenters, as opposed to the static runtime of traditional clusters, control systems need to operate at a high frequency; coarse-grained adaptation cannot adapt to quick changes and imposes overly conservative guardbands, leaving much on the table.

Completely solving the problems of providing QoS and operating datacenters efficiently remains an open research problem, and different techniques and approaches are needed depending on the specific context (application types, public versus public clouds, criticality of the QoS requirements, ...).
This dissertation analyzes these problems and provides two practical solutions for two somewhat complementary scenarios.

ANALYSIS AND CONTROL OF MECHATRONIC ACTUATION SYSTEMS IN INDUSTRIAL APPLICATIONS

Damiano Belloli - Advisor: **Prof. Sergio M. Savaresi**

The widespread of the actuation systems in the industrial applications leads to more and more demanding requirements in term of accuracy, dynamical response and energetic efficiency. Recently, the technological performances have been improved by the introduction of the electronic components, which allows the implementation of innovative control and estimation strategies: this new generation of devices, called mechatronic actuators, is the topic of this Thesis. In particular, three aspects that influence the mechatronic actuation systems are faced in the present work:

- *Modeling problems*: The knowledge of the dynamical behavior of the mechatronic actuators is the primary base for the improvement of performance, because it permits to deeply understand the system functioning and provide fundamental information for the development of the estimation and control strategies.
- *Estimations problems*: In industrial environment is not always possible to obtain the information or the measurements required for an ideal control, because of difficult operating conditions or absence of proper transducers: the estimation algorithms allow to overcome this limitation, providing methodologies which, under suitable conditions, guarantee robustness and accuracy
- *Control problems*: The proper design and tuning of the regulation systems are directly responsible of the performances. The knowledge and development of recent and innovate control strategies can provide solutions to overcome the actual limitations.

Since the mechatronic actuators studied in this Thesis have a practical impact on the industrial applications, the proposed methodologies and algorithms are validated by experimental tests. All the laboratory and industrial tests have been performed in collaboration with industrial partners: Consorzio Intellimech Kilometro Rosso, Scaglia Indeva and most of all Tenaris Dalmine R&D.

The Thesis is structured in two parts: in the first section, the development of suitable models, estimation algorithms and control strategies for different technologies of mechatronic actuators are illustrated; in the other section, the methodologies previously analyzed are applied to specific industrial environments, such as self-balancing manipulators and rolling mills for pipe production

The increasing number of industrial applications which involve the actuation systems and the more and more demanding requirements needed to face these technological challenges, lead to the necessity of innovative devices, such as the mechatronic actuators. The large diffusion of the electronic components bring to an enhancement of the performances, especially in the estimation and control fields, but there still is improvement margin, particularly in the overall system dynamic. For this reason, the mechatronic actuation systems are the main topic of this Thesis, with particular focus on the dynamical behavior analysis and control strategy design useful for industrial applications. As overview of the linear mechatronic actuators, three different actuation technologies are presented: the Electro Hydraulic Actuator (EHA), the Electro Mechanical Actuator (EMA) and the most recent and innovative Electro HydroStatic Actuator (EHSA). For each actuator typology, the main components and the function principle are detailed in addition to the description of the control strategy. Then, the main advantages and drawbacks of each technology are listed and related to the most common applications.

Then, the mathematical models developed to study the dynamic behavior of the mechatronic actuators (EHA, EMA and EHSA) are detailed. The gray-box approach is adopted to model all the different actuators: the dynamic behavior of the single components is described with physical equations and then joined to the overall models. The parameters are obtained in part from the datasheet characteristics and in part from suitable experimental measurements by a proper parameter identification procedure. Then, the overall models are validated comparing the simulation results with experimental data. The developed models are also adopted to develop suitable strategy to estimate variables that cannot be directly measured or comes from low quality sensor. In particular, the actuator position estimation is explored using the Kalman filtering approach and compared to the measurements provided by transducer of different technologies. Finally, the speed estimation has been also investigated with different algorithms.

Two different the control scenarios for linear actuators are illustrated. First, the Virtual Reference Feedback Tuning methodology is customized for the cascade control architecture to face the actuator positioning problem in the EHSA. Then, the pole placement and root locus criterion are used to design the regulator of the actuator force. Finally, on the base of the models previously implemented and tested, a global comparison of the different actuators is provided to highlight the performance differences between the EHAs, EMAs and EHSAs.

After a detailed technological analysis of the mechatronic actuators, two different example of industrial applications are presented. The first application case refers to a new regulation strategy for self-balancing manual manipulators. First, a brief introduction of the machinery and application environment is provided. Then, the experimental setup and physical model of the apparel used in this work are presented, with particular focus on the control architecture. The control strategy of the Virtual Feedback Reference Tuning, adopted also in the EHSA control, is applied to the selfbalancing manual manipulators, with the experimental validation of the obtained results. Then, the load mass estimation problem is faced with two different approaches and the performance are validated with proper experimental tests.

The second application case considers the regulation of the speed profile trigger for electrical motors in rolling mills, based on the load estimation. After an introduction of the hot rolling process for the seamless pipe production, illustrating the role and the importance of the mechatronic actuators in this industrial environment, the problem of the crop ends is discussed. The control strategy usually adopted to reduce the impact of the crop end phenomenon on the process productivity is presented, highlighting the main limitation. Then, an innovative control strategy is proposed to overcome the actual limitation: on the base of two different approaches for the motor load estimation, the proper trigger time is estimated and controlled in closed-loop. Finally, the effectiveness of the proposed regulation system is illustrated by the application to the real industrial plant.

Main Contributions

The main contributions of this thesis can be summarized as follows:

- *fine modeling and control analysis*

This Thesis proposes a detailed description of the mechatronic actuators of the main diffused technologies, with particular focus on the analysis of the dynamical behavior and on the design and tuning of specific methodologies for the estimation and the control of such devices. All the proposed algorithm are experimentally validated.

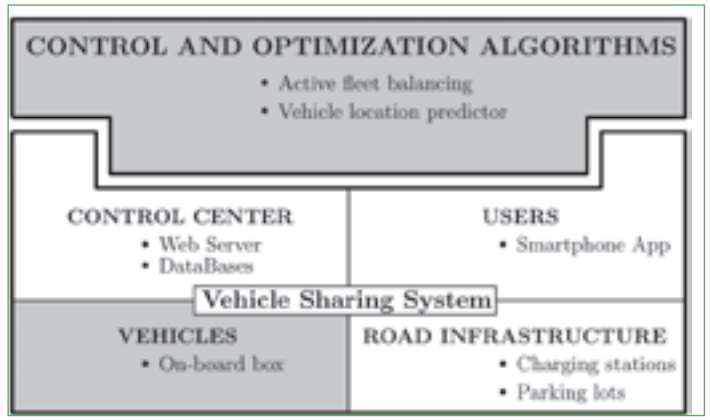
- *innovative control and estimation solutions applied to industrial cases*

A novel control strategy is provided for the velocity regulation of a self-balancing industrial manipulator, including load mass estimation: the results it the possibility of a transducer removal, obtaining the similar or better performances. An innovative methodology for the control of the speed profile trigger for electrical motors in rolling mills is also proposed, implemented and tested on a real productive plant, with notable benefit on the industrial productivity.

ANALYSIS AND DESIGN OF ADVANCED VEHICLE SHARING SYSTEMS: ON-BOARD TECHNOLOGIES, CONTROL AND OPTIMIZATION

Andrea Giovanni Bianchessi - Supervisor: **Prof. Sergio Savaresi**

Mobility has a huge impact in terms of economic costs and it accounts for one quarter of the world's greenhouse gas emissions. Traffic congestion, pollution and greenhouse gas emissions are some of the most prominent issues that modern metropolitan areas currently have to face. The goals set at the political level in this regard cannot be achieved without a significant re-think of urban mobility.
New models of sustainable mobility call for the integration of vehicles that differ in type (e.g. cars, scooters and bicycles), technology (e.g. electric, hybrid or with classical combustion engines) and ownership (they can be publicly or privately owned, fully or partially shared) within the same system, and they must offer end-users common functions, services and interfaces.
This Thesis deals with Vehicle Sharing Systems. Vehicle Sharing gained global popularity during the first decade of the 21st century as one of the possible solutions to environmental and increasing city traffic congestion problems. A Vehicle Sharing System is a 24/7 vehicle rental service (the fleet vehicle is heterogeneous) most dedicated to drivers who make occasional use of the car, typically for short periods of time and within city areas. Rates include both insurance and fuel, and to access the service users must be registered.
Specifically, this Thesis presents the Green Move project which is an innovative, high-technological and advanced Vehicle Sharing System based on an heterogeneous and electric vehicle fleet. The main components of such a wide and complex system are depicted in Figure 1. This Thesis focuses on the fleet vehicles related components, both at a very low (on-board electronic box) and high (control and optimization algorithms) levels. To visually inspect some of the Green Move solutions refer to: https://www.youtube.com/watch?v=kI3yiB50RWw.



1. Components of a Vehicle Sharing System on which control and optimization algorithms are hooked. This Thesis, among all the components, deals with the design and development of both a vehicle on-board electronic box prototype and control and optimization algorithms for fleet balancing and forecasting of vehicles locations.

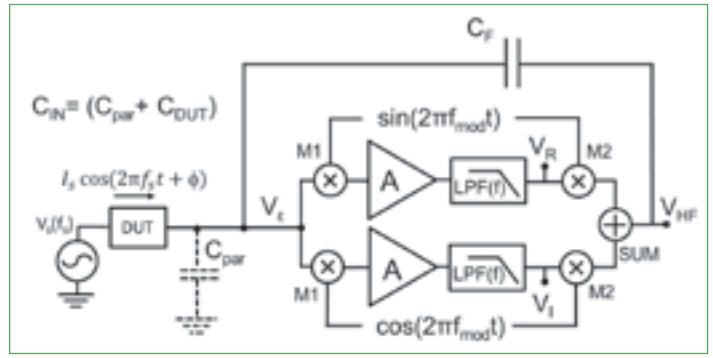
The services currently in place often appear as a simple rental service just for cars. However, the GM service is green, flexible, smartphone-based, free from intermediaries and provided with the most innovative fleet management algorithms.
A detailed description of the fleet vehicles is provided (e.g. size, number of seats and autonomy) along with technical aspects (i.e. vehicle data and buses, charging modes and commands).
All the vehicles have to be endowed with a Green-eBox (GEB) to be inserted into the GM fleet. The GEB is an Android based electronic on-board control unit; it implements several abstraction mechanisms that allow the seamless use of technologically different vehicles, and it provides a unique and standardized mode of access (the Vehicle Interface) for all the system actors. The GEB has been tested both on the fleet vehicles and, for four months, in a condominium- based VSS demonstrating great reliability.
Then, two innovative algorithms for managing and controlling the vehicle fleet were introduced. The Feedback Dynamic Pricing (FDP) technique models the VSS as a dynamical system. This opens up new control options for the fleet balancing problem. In this approach, a control strategy can be derived using the theory of feedback-regulated systems. Thus, assuming that people are sensitive to changes in the price of the service this can be actively and real-time con- trolled by acting on the trip fee. A full-fledged simulator has been developed and experimental data are encouraging and demonstrate the feasibility of the proposed approach.
The Nearest Available Vehicle (NAV) algorithm for predicting the user distance from the nearest vehicle overcomes the drawbacks of free-floating VSSs due to the absence of booking mechanisms. The idea is to use the data from vehicles past locations to make a prediction of future vehicle locations by providing users with the distance from the nearest vehicle. This technique has been successfully tested by using real car-sharing data from the Milano Car2go service.
The main contributions of this Thesis are summarized as follows:

- Development of an innovative system for vehicle sharing which eliminates all the intermediaries between the users and the system, Chapters 3 and 5. Green Move is the first Vehicle Sharing service which is entirely smartphone-based. It doesn't require neither a member-card to lock/unlock vehicle doors nor mechanisms to retrieve and give back the vehicle key. The key is simply sent by the control center to the user's smartphone as an electronic ticket.
- The possibility of adding to a Vehicle Sharing System vehicles that are heterogeneous in their types and owners, Chapter 5. The lacking of heterogeneity of the vehicle fleet is one of the most important limits of today Vehicle Sharing Systems. The Green Move project overcomes the most of the technological problems leading to an easy integration of different vehicle models.
- Development of an highly dynamic platform for managing the load- ing/unloading of application on the vehicle on-board box, Chapter 5. None of the currently available car-sharing kits offers the possibility to add/remove functions depending on the user, location or period. On the contrary, the Green e-Box is based on a novel middleware, which allows applications to be dynamically loaded/unloaded based on the current context.
- The modeling of a Vehicle Sharing System as a dynamical system, provided in Chapter 7. At the best of our knowledge, a Vehicle Sharing System has never been modeled as a dynamical system. This opens up new control options for the fleet balancing problem. In this approach a control strategy can be derived using the theory of feedback regulated systems.
- A novel approach called Feedback Dynamic Pricing (FDP) is presented in Chapter 7. It is based on the modeling of a VSS as a dynamical system and it aims to control the fleet balancing by varying the price of the service in real-time.
- An innovative technique (named Nearest Available Vehicle, NAV) for solving the problem of the lack of booking mechanisms in free-floating car-sharing services is introduced in Chapter 8. It provides users with the prediction of their distance from the nearest vehicle.

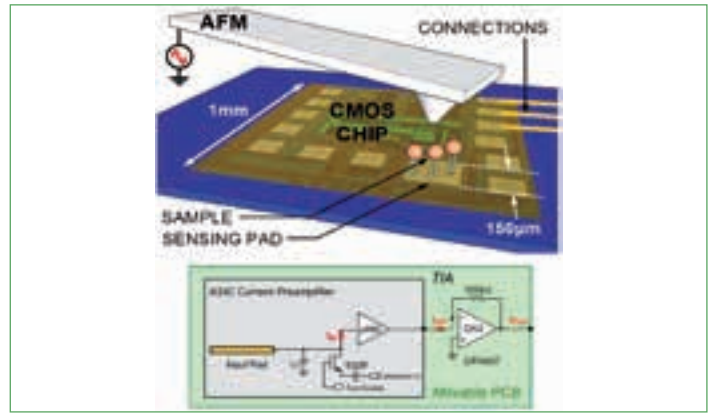
INTEGRATED INSTRUMENTATION FOR HIGH SENSITIVITY IMPEDANCE MEASUREMENTS ON NANOSAMPLES

Davide Bianchi - Advisor: **Prof. Marco Sampietro**

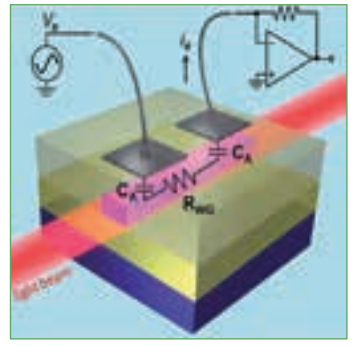
The dramatic progress in miniaturization, made available by the integrated circuit technology, allows nowadays to implement sophisticated measurement systems on chip that can perform in a millimeter sized package most of the functions available with standard discrete components instrumentation. Circuit integration not only allows to develop massively parallel systems, otherwise too complex and bulky to be managed, but it is often also the key for dramatically improving the sensitivity of the readout systems which are designed for investigating the micro- and nano- world. Sensitivity is indeed strongly related to the capacitance (electrodes and instrumental) at the input of the measuring system, C_{IN} which actually defines the noise level of the instrument: an IC bonded to a nanoelectrode with an input capacitance of few hundreds of femtoFarad would allow to get about a x100 improvement in the sensitivity, if compared to the setup built with a bulky instrument, with cable connections and about 10pF input capacitance. The miniaturization of the preamplification chain is not sufficient by itself, but must be coupled by both well-conceived setup and device to be measured. The setup must allow to actually shorten the connection to the device under test (DUT), physically approaching it as much as possible to the IC; the integration of the readout measurement system together with the DUT implies the development of embedded custom systems, which must also cope with special requirements (for example biological, or optical, mechanical, chemical) of the DUT itself. The device itself must be finally properly devised to match as much as possible the readout requirements: maximize signal and signal-to-noise ratio when connected to the readout electronics, minimize its input capacitance. The work presented in this thesis is focused on the development of instrumentation for high sensitivity measurements, and therefore focuses on the development of the IC preamplification chain, but also on the setup and on the device development. The first work presented is mainly focused on the development of an integrated transimpedance amplifier, made of a very peculiar two channel modulation/amplification/demodulation structure embedded into a feedback loop. In this topology, whose principle structure is shown in Fig.1, the incoming signal is firstly modulated to DC frequency by multiplier M1, then amplified by amplifier A with a gain G_A independent of f_s, low-pass filtered by LPF(f) to prevent $2f_s$ component and other spurious harmonics to propagate into the loop, and finally demodulated back to the input frequency by M2. For standard operating conditions, when $f_{mod} = f_s$, the circuit acquires the DUT input current and provides an output V_{HF} at the measurement frequency, whose phase and amplitude are proportional to the input DUT admittance. Since the signal is translated into DC, two paths in parallel are required to process both the in-phase (I) and quadrature (Q) input signal components; the unity gain SUM block adds them together as they are translated back into the measurement frequency. The translation into DC allows to obtain into the loop two DC outputs which behave as an external lock-in system connected to V_{HF} . Despite the increased loop complexity, the DC translation has therefore the benefit of avoiding an additional lock-in system, as well as advantages in terms of sensitivity and compactness. Besides the IC, a dedicated setup suitable for embedded applications has been also implemented, allowing to realize a complete instrument for impedance spectroscopy. The new IC topology, coupled with the dedicated PCB, allows for the first time to reach a resolution as good as 0.4aF in the 100*kHz* – 150*MHz* range. The second work presented focuses instead on the implementation of a setup conceived for electrical measurements with an Atomic Force Microscope (AFM): as sketched in Fig.2, in this measurement system the AFM probe is used as input electrode, while the metal input pad of an appositely optimized CMOS amplifier is directly used as a conductive substrate electrode. This solution is determinant for minimizing the input stray capacitance and the capacitive coupling with the tip holder, and therefore the equivalent input current noise. Although more delicate than standard passive electrodes, this active substrate allows to explore the ultimate detection limits of this technique, without requiring any modification to the commercial AFM instrument, and leading to a system sensitivity up to 14.4zF. The third and last work presented focuses instead on the design and production of optimized devices which are conceived to sense the photoconductive effect caused by light, when it is coupled into conductive waveguides. This technology is very promising for the development of non-invasive light detection into microphotonic structures. The device concept is schematically shown in Fig.3, and consists of two electrodes surrounding the optical waveguide and spaced from its core by an electrically insulating layer: the core conductivity, which is increasing proportionally to the coupled light, can be measured for frequencies higher than $f_A = 1/(2C_A R_{WG})$. After theoretical study of the physical mechanisms involved and a first analysis of the material platforms suitable for this purpose, amorphous Silicon devices have been designed, produced and finally tested, proving the validity of their working principle on this technological platform.



1. Basic schematic of the impedance analyzer based on modulation/demodulation architecture IC.



2. Principle scheme of the new measurement system topology.



3. Illustration of the device structure, consisting of two metal electrodes deposited onto the electrically-insulating upper cladding and capacitively coupled (*CA*) to the electrically-resistive (*RWG*) silicon waveguide.

RADIO PLANNING AND MANAGEMENT OF ENERGY-EFFICIENT WIRELESS ACCESS NETWORKS

Silvia Boiardi - Supervisor: **Prof. Antonio Capone -** Co-supervisor: **Prof. Brunilde Sansò**

Legacy wireless access networks (WANs) are typically designed regardless of energy efficiency considerations.
The planning objective is to minimize the deployment costs while insuring complete area coverage and connection quality in any load condition, especially during peak traffic periods. In recent years, the growing demand for ubiquitous mobile communication access increased not only the amount of customers to be served, but also the awareness of the environmental impact. If on one hand the capacity of the current networks has to be upgraded to manage the extra traffic, on the other hand new energy-aware techniques should to be introduced in the regular network functioning.
Within this context, *green networking* has emerged as a new way of building and operating communication networks to improve their energy efficiency. Besides its obvious environmental advantages, power consumption reduction is beneficial to mobile operators for economic reasons. Two types of cost are incurred: capital expenditures (CapEx), related to the purchase and installation of radio equipment, and operation and management expenses (OpEx), consisting in energy, site rentals, marketing and maintenance costs.

While energy-aware network design models or operation management solutions have been largely treated in the research world, the challenge is to convey both CapEx and OpEx costs as well as power issues into a single modeling framework.
The work carried on during this doctorate program arose from the belief that a fundamental issue had been so far overlooked by the research on green mobile networking: an effective energy-aware operation through a cell sleeping technique is closely dependent on the planning decisions taken during the network design phase. While the design and operation problems had always been tackled separately, we developed an optimization model that *jointly* considers the *planning and management* of a wireless access network. The philosophy at the basis of the approach is summarized in Figure 1, where a three-cells toy topology is represented. Bold circles stand for the area coverage of turned-on access stations, thin circles stand for switched-off devices and black dots stand for traffic aggregators (also called test points). Figures 1(a) and 1(b) report examples of a topology deployed according to the minimum installation cost criterion, operated during high and low traffic load respectively. During off-peak periods, the operation savings correspond to the energy spared by turning off the largest cell, while the other two remain active to serve the traffic demand. Would it be possible to decrease further the energy consumption of the topology? Figure 1(c) answers this question, illustrating the main principle of the joint optimization. Given a traffic variation profile, the joint framework is able to build a *flexible topology*, specifically designed to take advantage of the demand variability. In the example, by adding an additional cell at the cost of a modest CapEx increase, it would be possible to switch off as much as three access stations instead of just one in low load conditions.

1. Effect of joint optimization on network operation management.

The *joint planning and energy management* (JPEM) optimization model, as we named it, is an original integer linear programming problem. Its objective is to minimize the sum of the capital investment and the energy costs due to devices' deployment and operation and, at the same time, to provide an example of the possible benefits derivable from a cell sleeping mechanism applied to the chosen topology. Tuning the objective function by means of a trade-off parameter, different network topologies can be produced to accommodate the designer's requirements. In particular, the major selling point of the JPEM is that, whether a service area is blank or already covered by legacy equipment, it allows the network planner to build the most energy-efficient network from the energy consumption point of view, maintaining the CapEx under control.
As main line of research, we designed the JPEM framework to optimize cellular access networks. An instance generator was developed using C++ programming language to produce several test scenarios, considering LTE technology and three types of access devices - macro, micro and pico cells. Once again first in the literature to tackle planning and management at once, we also developed a variation of the joint formulation to suit wireless mesh access technology. We designed an all-new instance generator, this time producing Wi-Fi mesh test instances. In both cases, we implemented the models using AMPL, a programming language specific for high-complexity mathematical computation, while CPLEX branch and bound solver was used to optimize them. Through remarkable results, we demonstrated that network planning and energy efficiency are closely related. To further confirm the validity of the joint approach, we compared the cellular and wireless mesh joint model results with those of a more traditional procedure, where network deployment and operation are addressed separately.
Due to the high number of optimization variables, the JPEM framework is extremely computationally expensive. Therefore, the last part of the doctorate program was devoted to develop an ad-hoc heuristic procedure in order to obtain results in a shorter amount of time and allow the investigation of larger test examples. Through the observation of the cells' activation pattern during the daily peak and off-peak time periods, a partial topology is provided as input to the JPEM and integrated to form a feasible solution. The heuristic approach, fully implemented using AMPL, proved successful and solved a large set of real size instances.

MULTI-OBJECTIVE METHODOLOGIES FOR VEHICLES RIDE QUALITY ENHANCING

Stefano Bottelli - Supervisor: **Prof. Sergio M. Savaresi**

This work was developed under the supervision of Prof. Sergio Matteo Savaresi, from Politecnico di Milano. It is this work goal to investigate the potential of semi-active suspensions and active aerodynamic surfaces for the purposes of developing centralized control strategies, which can improve several aspects of ride experience towards an ideal multi-objective vehicle control. It is in fact shown how these actuators can be used with multiple purposes, from comfort enhancing, through road holding and handling increasing, to steering and braking capabilities improvement. A new control strategy (namely the Multi-objective Chassis Control, also referred to as MCC) is also developed both for semi-active suspension and Active Aerodynamic Surfaces (AAS), which makes use of logical control rules and it is independent from vehicle models, similarly to the skyhook algorithm. The MCC is therefore characterised by an extreme implementation simplicity, and it can easily be employed on any 2-wheels or 4-wheels vehicle without re-tuning, being not modeldependent. In addition to these qualities, it can provide better performance with respect to decentralized control strategies (such as the skyhook), due to the synergistic usage of the actuators. It is also shown how MCC allows to decouple comfort objectives, namely heave, pitch and roll oscillations reduction, making it an extremely versatile and powerful algorithm.

In the beginning, a complete study on the potential effectiveness of active aerodynamic surfaces is carried out. In particular, two reasonable configurations are presented for AAS installation on each vehicle corner, and decentralized control techniques are developed with the goal of enhancing comfort and road holding respectively. Two model-based controllers are designed on the classical quarter-car model. The proposed design and the several trade-offs involved in the tuning are discussed considering several parameters. In particular, the effect of different tunings, the aerodynamic surface size, the road disturbance amplitude, the vehicle mass and the vehicle velocity are thoroughly discussed and numerically analysed, from the performance and power requirements point of view. Furthermore the AAS control is tested with semi-active suspension systems, showing that the two approaches are additive from the performance point of view. The performance improvement comes at the cost of power requirements and a complex mechatronic design. If compared against semi-active suspensions the cost-benefit ratio does not seem compelling, but in terms of pure performance improvement, the proposed AAS system outperforms semi-active suspensions. It yields performance comparable to that of fully active suspension at a fraction of the power requirements.

The most popular comfort-oriented algorithms developed for semi-active suspensions are typically designed on the quarter-car model. When it comes to actual implementation on real vehicles, four (or two in the half-car model) controllers usually take care of the comfort enhancing of each singular corner, neglecting the interactions among them. Thus, several simulations are carried out in this work, where a half-car and a full-car model are endowed with semi-active suspensions, each one controlled with the same decentralized strategy. Both frequency and time domain evaluations are performed in order to assess comfort performance, showing how classic control strategies perform when employed in a decentralized fashion.

After, a new centralised methodology that can improve comfort by means of the synergistic use of all the semi-active suspensions in a vehicle is presented. The advantages of the MCC technique lie in its simplicity: similarly to the classic logic rule based strategies for semi-active suspensions control, the MCC algorithm is extremely simple and its implementation on a real vehicle is trivial. The sensors that are needed in order to operate it are an inertial unit on the chassis, which is able to provide the controller with heave speed, pitch speed and roll speed, in addition to potentiometers measuring the suspensions deflection (so that the suspension deflection speed can be computed by derivation). Moreover, since the MCC methodology is not model-based, it can be readily employed on different vehicles without any re-tuning. Another great potential of the strategy is that the control objectives can be decoupled in order to define which variable the control system needs to focus on more, among chassis heave, pitch or roll reduction. The parameters that need to be modified in order to modulate the algorithm action on each objective are extremely concise and trivial to tune. Several time domain and frequency domain tests are carried out, showing that the classic semi-active control strategies performance can be outdone by the MCC methodology. In addition, due to the extreme simplicity of the algorithm, it is showed how MCC can be rearranged in order to control active aerodynamic surfaces. Several simulations are conducted in order to assess the new control system performance, highlighting the higher effectiveness of the methodology in enhancing vehicle comfort with respect to decentralised control strategies. Later, a preliminary concept of multi-objective vehicle control by means of active aerodynamic surfaces is presented. It is in fact shown how these actuators can be used with multiple purposes, switching among different control modes, accordingly to the driving scenario. A ride mode map is presented, together with the switching conditions among the ride modes, namely the driving mode, the braking mode and the steering mode. Per each ride mode a control strategy is developed, and the outcomes are analysed by means of time domain and frequency domain performance assessment tests. It is thus noted how AAS can be successfully employed for enhancing several aspects of the ride quality, such as comfort, braking efficacy and steering capability.

Finally, since it is shown that the knowledge of the actual vehicle mass is vital for the development of a model-based AAS controller, a load estimation algorithm is developed. The spectrum analysis based method is shown to be able to provide a reliable estimation of a vehicle load and of its distribution without the need of measuring neither road profile nor the wheel vertical motion, but simply by processing the vertical acceleration of the chassis and the pitch speed, the latter being required to provide additional information on the load distribution. The method is shown to be effective enough to discern among a copious but finite set of loads and distributions. Naturally, the potential of the developed estimation methodology goes beyond AAS controller effective tuning, since mass knowledge is considered a vital information to prevent the criticalities that arise in several other applications, when load can vary during the vehicle usage. This can happen for example with public transportation, garbage collectors, or private vehicles, especially motorbikes, in which the load variation with respect to the vehicle mass is high.

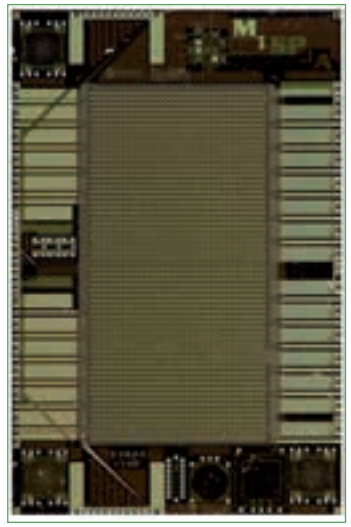
FRONT-SIDE AND BACK-SIDE ILLUMINATED SPAD ARRAYS FOR 2D IMAGING AND 3D RANGING

Danilo Bronzi – Supervisor: **Prof. Franco Zappa**

Over the last few years there has been a growing request for all-solid state smart imagers, fostered by the needs of a number of applications, which demand complex and advanced data collection with high sensitivity (possibly at the single-photon level) and very high acquisition speed (possibly thousands of frames per second). Among theses applications, we can mention: molecular imaging, fluorescence lifetime imaging, micro-array based biological analysis, confocal microscopy, and adaptive optics; as well as, Safety (e.g. automotive, environmental surveillance, traffic and workplace safety monitoring, product safety analysis, food and agriculture quality and safety assessment) and Security (access control, biometrics, surveillance systems, dangerous agents monitoring, homeland security, fire hazards) scenario, which simultaneously require:

- single-photon sensitivity in the 300 nm – 900 nm wavelength range;
- integration times as short as microseconds (for precise time-tagging or videos of fast optical transients);
- instrumental response function capable of measuring few tens of picoseconds for the estimation of the arrival time of a single-photon (e.g. for time-of-flight computation);
- sufficient horizontal and vertical pixel resolution to image a broad field of view;
- adequate image repetition rate to allow real-time image acquisition and image processing;
- low power consumption, compactness, and low-weight for reducing costs and for ease of installation and maintenance. Nowadays, the imager market offers a broad portfolio of either commercial- or scientific-grade cameras, ranging from consumer CMOS active pixel sensor (APS) cameras up to high-end CCD imagers. However, none of them simultaneously offer high speed and ultra-high sensitivity: CCDs reach sensitivity at close to single-photon level but necessarily require cooling and long integration times (i.e. very low frame rates); APS imagers provide video-rates but with relatively limited detection efficiency, thus requiring bright illumination scenes. Special kind of CCDs, like the intensified or electron-multiplying CCDs, can reach sensitivity at close to single-photon level but are costly and bulky and necessarily require a cooling system; photo-multiplier tubes and micro-channel plates are well-known single-photon detected but they cannot be integrated with complex CMOS electronics for advanced imaging. The aim of this doctoral work was to design monolithic a Single-Photon Avalanche Diode (SPAD) array and read-out electronics, able to provide simultaneously both high frame rates and single-photon sensitivity. The array was fabricated in an automotive-grade CMOS technology to ensure scalability, reliability, and low cost. Moreover, with the aim of enabling high performance smart system-on-a-chip implementations the single-photon detectors are paired with sophisticated in-pixel intelligence able not only to deliver two-dimensional (2D) intensity information – through photon counting in either free-running (down to 10 µs integration time) or time-gated mode – but also to perform light demodulation with in-pixel background suppression. The provided features allows different operating modes, thus enabling both time domain applications – such as flash detection, fluorescence lifetime imaging or fluorescence correlation spectroscopy – as well as frequency-domain lock-in depth measurement for three-dimensional (3D) ranging in automotive vision and lidar. The development of the first ever back-side illuminated CMOS SPAD is reported. In the future, this approach will allow one to achieve a higher pixel density (up to 100 %) and an enhanced spectral sensitivity in the near-infrared range. The designed SPAD imager (Fig. 1) was used to assemble a prototype 3D camera (Fig. 2) for advanced driver assistance and collision avoidance. The system electronics includes a FPGA module that processes data coming from the chip; a high–speed USB 2.0 interface for communication and a 32 MB SDRAM for data storage. A second back-mounted board accommodates the power converters, digital logics, and analog circuitry for implementing a Direct Digital Synthesizer, thus providing arbitrary waveform modulation for the light source. The camera is housed in an aluminium case supporting a 12 mm F/1.4 C-mount imaging lens, whose field-of-view is approximately 40°×20°. The whole system is very rugged and compact and a perfect solution for vehicle's cockpit, with dimensions of 80 mm × 45 mm × 70 mm, and less that 1 W dissipation. To provide the required optical power and allow fast modulation of the optical signal, the illumination source has a modular design based on a power supply board and five 808 nm laser driver cards. The system was operated at night and during daytime, in both internal and real traffic scenario, yielding high dynamic-range (118 dB) high-speed (over 200 fps) depth measurement with high precision (less than 50 cm at 45 m). The prototype was tested at Centro Ricerche Fiat and proved to be capable of acquiring information about the surrounding objects on the road in a both standing or moving car and of predicting the relative speed and movements between the obstacle and the car, thus warning the driver and avoiding the collision by properly acting on the car. Taking advantage of the high sensitivity and fast frame rate capability, the SPAD camera was used in collaboration with the Karolinska Institue and the Technical University of Berlin to evaluate a novel DNA sequencing method. The technique relies on the contemporary measurement of Fluorescence Correlation Spectroscopy (FCS) and Fluorescence Lifetime Imaging (FLIM) of 4 different dyes labelling the 4 different nucleotide. With respect to other method, this one is so sensitive that do not need DNA multiplication through Polymerase Chain Reaction (PCR) and thus would allow one to detect the smallest mutation in a DNA sequence. This research project is still going, but promising results were already obtained. A second collaboration with the Institute of Photonic Sciences (ICFO) was also launched. The measurements carried on at ICFO were aimed at exploiting and validating the high-sensitivity of the SPAD camera to overcome the typical low frame rate of CCD in speckle contrast spectroscopy (SCOS), and let scientists image at high speed the blood flow in living tissues in few seconds. Overall, the opportunities given by the designed CMOS SPAD opened the path towards important developments that can greatly influence several fields such as biology, medical science, imaging and security.

1. Designed Single-Photon Camera.



2. Photograph of the bonded SPAD imager.

ENERGY EFFICIENT LOCALIZATION TECHNIQUES FOR WSN COMMUNICATIONS

Salar Bybordi - Supervisor: **Prof. Luca Reggiani**

In this research, localization is treated in a more general way w.r.t. the existing definitions in the literature since here the concepts of static localization, fingerprinting in a single shot estimation of position information and tracking of position information over different time instants are integrated for achieving a common goal, i.e. an enhanced energy efficiency. Localization scenarios, considered in this research, cover active localization, in which the target participates in the estimation process of inter-node ranging information, and also passive scenarios, in which the target does not participate in the estimation process. The general trend in design of localization algorithms is towards achieving more accurate estimation of position. Additionally, considering the limited battery life in wireless sensor networks (WSN) nodes as a determining factor in the performance of WSNs, the algorithms should respect energy efficiency constraints. The contribution of this research is mainly focused on the investigation and proposal of localization algorithms which should achieve these two crucial design objectives:

- accuracy improvement
- energy efficiency

Considering the two above-mentioned goals, this research considers algorithms for both active and passive localization scenarios. The first section of the research deals with the investigation of energy-efficient localization algorithms for passive localization. The first part developed for passive scenarios is dedicated to the theoretical analysis of a hybrid tracking algorithm composed of active and passive steps. The analysis of this scheme, based on the Posterior Cramer Rao Bound (PCRB), confirms that mixing active and passive cases can be an effective tool towards energy efficiency. The second part of the investigation of passive scenarios deals with the energy efficiency under two perspectives. A hybrid tracking algorithm composed of Extended Kalman Filter (EKF) based tracking and Fingerprinting (FP) is proposed in order to tackle conventional problems related to the implementation of either tracking or fingerprinting separately. One of the common drawbacks of FP belongs to large data size and consequent large search space as a result of either vastness of surveillance area or finer grid resolution in FP grid map which limits the application of FP to small environments or scenarios with largely spaced grid points leading to poor localization performance. This hybrid algorithm enables FP to be applied in larger environments or environments with finer space grids. The second aspect of the scheme deals with the critical concern when tracking more than one passive target, i.e. to make the correct discrimination among measures as a result of occurred ambiguity in paths clusters scattered by different targets especially when targets are moving close to each other. The final phase of this analysis is dedicated to consideration of applied TOA-based ranging technique (so-called soft ranging) and its potential characteristics for providing more accurate ranging measures by means of feeding a kind of a-priori information. This issue provides EKF update steps with more precise ranging measures leading to a better localization performance. Simulation results validate a zero-energy tracking algorithm in which the mobile target does not consume energy. The second section of the research deals with energy efficient algorithms for active localization scenarios. Accordingly, power allocation among transmission power of beacons is an effective tool toward this objective. We present a pervasive literature review on existing power allocation (PA) schemes and, in particular, of the optimal ones, based on the minimization of a localization error bound (SPEB), function of transmission power from each beacon. Then, two new sub-optimal algorithms are investigated. The former is based on the definition of a parameter called uncertainty area which is a convex function of transmission power in the pair-wise selection of beacons. Numerical results confirms a notable performance advantage of localization with Power Allocation (PA) schemes w.r.t to the case of uniform power allocation among beacons specially in target locations in the vicinity of beacons. The latter proposed algorithm is based on the fact that the optimal SPEB based PA approach does not show any advantage when performance (i.e. Mean Square Error (MSE)) of the ranging estimator achieves a floor corresponding to a certain threshold in the received Signal to Noise Ratio (SNR). This corresponds to the behavior of a practical, realistic ranging estimator where the achievement of higher ranging accuracies becomes not possible by increasing transmission power over a certain threshold because of phenomena like, for example, the maximum sampling rate and computational load available in the sensors. Consequently, this sub-optimal, simplified PA algorithm is based on the distribution of transmit power of beacons with the SNR above the threshold to beacons with the SNR below the threshold, realizing a type of simple adaptive power allocation (Adaptive Power Allocation (APA)) directly based on the measured SNRs.
Simulation results confirm that such a simple strategy can be effective in medium low SNR regions, even w.r.t. more sophisticated optimization procedures.

MAIN RESULTS OBTAINED

I. *Achievement of energy efficiency via combination of active and passive steps in localization*
II. *Design of a power allocation algorithm for active localization scenario in order to improve localization performance*
III. *Investigation of impact of real ranging estimator on existing optimal power allocation strategies for WSN localization*
IV. *Design of a hybrid fingerprinting and EKF based tracking for indoor passive localization scenario*

OPTIMAL/ROBUST CONTROL FOR ANTI–SLIP/SKID DEVICES IN RAILWAY APPLICATIONS

Danilo Caporale - Advisor: **Prof. Patrizio Colaneri**

Control systems for wheel axle control of railway vehicles are the main focus of this work. We analyze the problem from two points of view: the single axle speed control and the whole train coordinated control. In both of them we propose innovative control algorithms to enhance the performance of current wheel–slip protection devices exploiting advantages offered from modern control theory. Results are theoretically validated and simulation experiments are given.

Chapters 1 and 2. Introduction, motivations and background

The operation of railway vehicles requires the ability to control the rotational speed of the wheels in every environmental condition. To this purpose, a wheel slip protection device (WSP) is necessary to prevent the wheels from slipping on contaminated rails. This is done by comparing the axles peripheral velocity with the train velocity and by controlling the speed to prevent the two from being too different. Despite the progresses control theory has experienced in the last decades we still face, especially in adverse meteorological conditions, inadequate railway traffic response due to difficulty of controlling the speed of the wheels on slippery rails. This problem is particularly evident in snowy regions or in urban areas where the autumnal mixture of foliage and water challenge even the most recent WSP devices. It is remarkable how technological improvements have not been able to solve this problem: an emblematic example is what happens in braking maneuvers, where initially poor performances where ascribable to the difficulty of controlling mechanical brakes with air pressure circuits and the use of electrical motors as brakes, with its faster response times, has not been a complete answer to the problem. In the spirit of overcoming such difficulties, in this work we devoted our attention to the use of modern control techniques to push the performances over the current industrial and scientific limits, paying attention to what, in our opinion, was not yet investigated and proposing innovative ideas to overcome current limitations.

Chapter 3 - Modeling approach

The dynamics of the mechanical system involved in wheel speed control is studied. Different models are used for different design strategies, and each of these is thoroughly presented. Particular attention is devoted to the understanding of the wheel–rail adherence characteristics and propose a model capable of describing all possible operating conditions with a relatively simple nonlinear parametrization. Then an innovative adaptive controller based on such parametrization is presented and its performances analyzed via simulation. Also a discrete controller based on optimal filtering and control theory is presented and its performances assessed via simulation on a detailed Simulink diagram kindly provided by Alstom Transport–Italy. It has also been possible to test part of this algorithm on a real time Hardware In the Loop simulator, always provided by Alstom Transport–Italy. Other than the single axle control, we will devote our attention to the whole train braking performance with an innovative distributed preview control approach and we will show how a simple communication between neighboring coaches is capable of improving stopping distance performances. The illustrated models range from a single wheel model without transmission dynamics to a simplified full train model. Each of these has different purposes. The single wheel model is the most useful to study traction and braking algorithm as it is the simplest model capable of describing all relevant phenomena. Considering transmission dynamics allows for performance evaluations in presence of resonances which often appear in practical applications. The full train model is used to propose a high–level strategy of braking effort coordination whenever communication allows for a neighbor–to–neighbor communication. The electrical model clarifies the context of operation of the present work.

Chapter 4 - Adherence characteristic

The main source of disturbance when designing a control system for the wheel speed is the friction force between the wheel and the rail, often called *adherence force*. This force is also the mean by which a torque applied to the wheel axle is transmitted to the rail, hence providing traction or braking force to the vehicle. In the following sections different existing models of the wheel–rail con- tact force are presented, which are based on the theory *elastic contact between solids*. Peculiarities of each model is summarized, and from these model we propose two simple control–oriented models which are capable of describing all operating conditions.

Chapter 5 - Discrete time adaptive controller

An algorithm for optimal wheel slip control to be readily used in the implementation of a discrete time controller is presented. Theoretical justification is given as well as simulation results obtained on a detailed metropolitan train model provided by Alstom Transport.

Chapter 6 - Nonlinear adaptive control

In this chapter we describe a nonlinear adaptive control technique for the control of a single axle. The main contribution of this chapter is a new technique for anti-slip control in railway vehicles. This technique is based on a new parametrization of the unknown adherence force. The control philosophy aims at enhancing vehicle performances by estimating the maximum available adhesion force with a nonlinear filtering, which is then exploited via a nonlinear controller. The effects of the implemented control scheme are the use of the maximum available deceleration in case of poor adhesion conditions and, consequently, a reduction in the wear rate of the rolling stock. The results are illustrated via simulation.

Chapter 7 - Distributed Control with Preview

In this chapter, we propose a method to take advantage of information exchanges between preceding and following coaches in order to improve the braking performance and to reduce the stopping distance of the train. Different braking control approaches are considered depending on the level of communication between the control units. In particular, a novel distributed braking control algorithm is proposed that, thanks to the preview of the future condition of the rail offered by the coaches at the front of the train, is capable of better tracking the maximum adherence condition. In so doing, it is possible to enhance the performance offered by the single control units and to reduce the stopping distance as compared to standard techniques based on limited (or lacking in) communication. Extensive simulations have been performed to test and compare our algorithm with standard techniques. The numerical experiments confirmed the soundness of the preview control based strategy and showed that it outperforms the other techniques. Special attention is devoted to understanding the part played by the various parameters in the problem, such as the distance between the actuated coaches and the convergence rate of the control units. Focus has been placed on the role of communication and preview for performance improvements. As such, the dependence of the adherence curve on the train state is expressed only by its dependence on the position and the traveling speed. The communication protocols used in railway applications are introduced, to show how the proposed algorithms can be implemented and spot the possible problems that might arise.

Chapter 8 - Simulation and tests

We provide some simulation results conducted in different scenario to assess the validity of the algorithm proposed in Chapter 5, where only a result has been shown which has been obtained on a real time simulator available in the Sesto San Giovanni Alstom's plant. Here more extensive simulations have been conducted on a Simulink model of the train.

TOWARDS THE DEFINITION OF A METHODOLOGY FOR THE DESIGN OF TUNABLE DEPENDABLE SYSTEMS

Matteo Carminati - Supervisor: **Prof. Cristiana Bolchini**

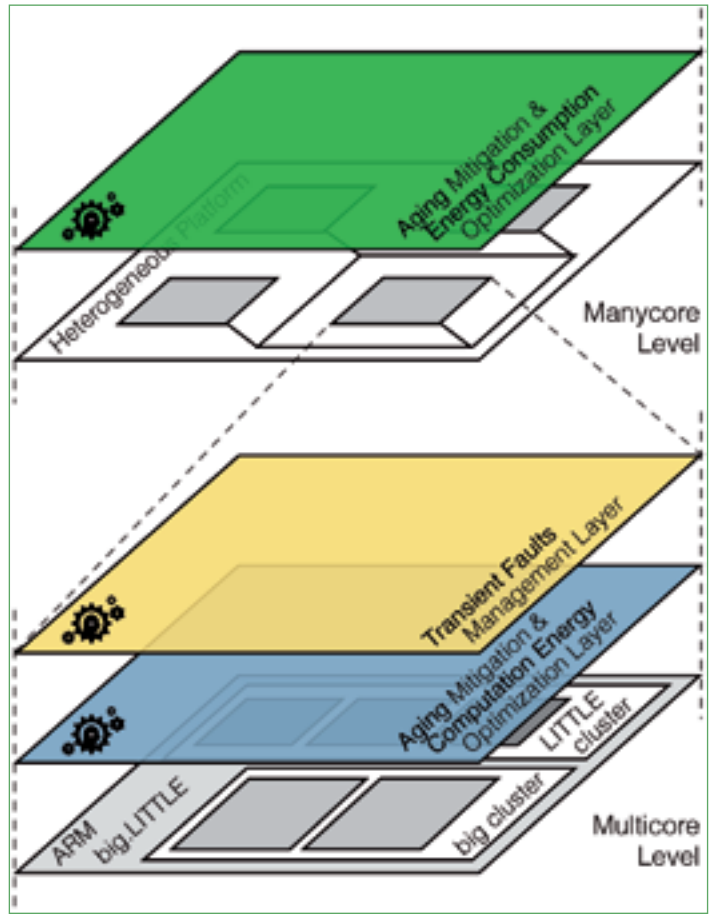
The problem of guaranteeing the correct behavior in digital systems even when faults occur has been investigated for several years. However, the researchers' efforts have been mainly devoted to safety- and mission-critical systems, where the occurrence of faults (both transient and permanent) can be extremely hazardous. Nowadays, the need to provide reliability also for non-critical application environments is gaining a lot of momentum, due to the pervasiveness of embedded systems and their increasing susceptibility due to technology scaling. While in critical applications the budget devoted to reliability is almost unlimited and it is not to be compromised, in non-critical scenarios the limited available budget used to guarantee the best performance and energy consumption is to be shared for providing reliability as well.

In the past, great effort has been devoted to provide strict reliability management. This led to the shared belief that reliability is to be considered from the early stages of the embedded systems design process. In fact, as this process is becoming more and more complex, approaches that do not consider reliability throughout all the design steps may lead to expensive or not-optimized solutions. Moreover, considering reliability in a holistic way allows to drive the several decisions by exploiting the synergy of both the most classical aspects and reliability-oriented ones. Postponing the reliability assessment to the later phases of the design flow on a system prototype is not appealing, because failure in achieving the desired level of reliability would be detected too late. However, the complexity of managing reliability, performance and power/energy consumption all at the same time grows exponentially, especially when several decision variables are available in the considered system. For this reason it is not possible to envision a system able to properly react to any possible scenario on the basis of decisions precomputed at design time; a new paradigm based on self-adaptability is to be designed. Self-adaptive systems are becoming quite common when dealing with such complex systems: relevant examples are available in literature if performance management is considered.

Given these motivations, we argue that the self-adaptive paradigm is to be implemented when designing embedded systems with the aim of considering reliability as a driving dimension. In particular, in this thesis we propose a comprehensive management framework for dealing with reliability in multi/manycore embedded systems. Reliability represents the main optimization dimension and is considered both for permanent/transient faults management and components aging mitigation. Energy consumption minimization has been introduced because it is directly and considerably affected by the knobs the framework exploits. Perfomance is taken into account as a constraint to be satisfied according to the soft real-time paradigm. The result is a cross-layer self-adaptive system for the combined optimization of reliability and energy under performance constraints.

The proposed framework autonomously takes care of the resource management problem, hiding its complexity. The overall work is organized in several layers as shown in Figure 1.

This framework implements a well-known control loop where the status of the system and the environment are sensed (observe), adaptation is defined through decisions made at runtime to meet the specified goals and constraints (decide), and the values of the system parameters are modified accordingly (act). The designed framework is integrated in a two-layer heterogeneous multi/manycore architecture, which is considered as the reference hardware platform. At the multi-node level, the designed framework employs a hybrid approach to minimize aging while optimizing communication and computation energy. A runtime orchestrator has been designed to smartly map tasks on the available nodes starting from pre-computed optimal mappings. Tasks are then re-mapped, at runtime by means of heuristics, to cope with the evolving conditions. Transient faults management is considered at the multicore level only, since creating, scheduling, and gathering results of redundant threads and voters/checkers benefit from a shared-memory bus-based architecture such as the intra-node one. A rule-based system has been designed to guide the orchestrator in selecting, at each instant of time, the best redundancy-based reliable technique to satisfy the user's reliability requirements and minimize the perfomance overhead. This layer is located, in each node, on top of another adaptation layer that takes care of aging mitigation and computation energy optimization. This is achieved by acting on different knobs (tasks mapping and scheduling, resource switch-on/off, DVFS) through the synergic orchestration on ad-hoc designed controllers. Each adaptive layer has been validated in a simulation environment by executing application traces collected from execution on real architectures.



1. A graphical overview of the proposed system composition.

The envisioned framework and the design of the presented adaptation layers represent the main innovative contribution of this research work. The preliminary investigation on self-adaptive systems led to the formalization of a model for describing and organizing this kind of systems in a structured and systematic way, as well as for preliminary validating them. Moreover, the need for estimating complex architectures lifetime motivated the development of a lightweight framework for estimating the reliability function and the MTTF for architectures able to tolerate multiple failures, considering varying workloads. This framework is based on Monte Carlo simulations and random walks.

The obtained results proved the effectiveness of the proposed approach, obtaining remarkable improvements in terms of lifetime extension and energy consumption reduction, while meeting performance constraints.

MICROMACHINED ULTRASOUND TRANSDUCERS: FROM THEORETICAL MODELS TO OBJECT LOCALIZATION IN AIR

Alessandro Caspani - Supervisor: **Antonio Francesco Longoni**

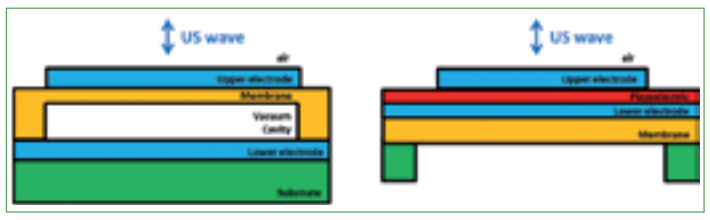
In this PhD work, MEMS technologies for modern ultrasonic transducers have been studied. As part of a collaboration between Politecnico di Milano and VTT - Technical Research Centre of Finland, the aim of this work is to obtain extensive knowledge on the prospects and limiting factors for these technologies to both replace current ultrasonic transducers, fabricated with piezoelectric crystals, and to explore the possibility to further extend their field of application.
The theoretical analysis and the supporting experiments reported in this work show a promising future for Micromachined Ultrasonic Transducers, based both on capacitive (CMUT) and piezoelectric (PMUT) transduction, as represented in Figure 1. The main features that stand out in these devices are: low-power operations, cheap cost of production, ease of interfacing with electronics, high density arrays capabilities.
The mechanical structure has a mass-spring-damper behavior and therefore can be represented with an RLC electrical circuit. Further expanding the scheme, three different physical domains can be separated by means of ideal transformer: electrical, mechanical and acoustical.
Analyzing the transfer function thus obtained from the electrical to the acoustical ports and vice versa two Figures of Merit are related to physical parameters of the device: the first is the emissivity, defined as the ratio of the emitted acoustic pressure and the applied AC voltage,

$$E(j\omega) = \frac{\bar{P}_{ac}(j\omega)}{V_a(j\omega)} = Z_a \eta \frac{j\omega}{k_m + j\omega b + \omega^2 m}$$

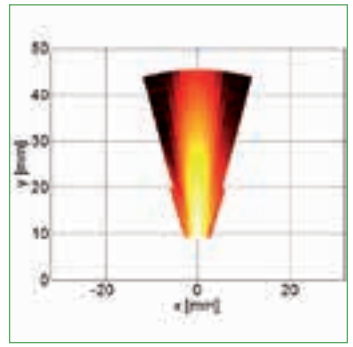
and the second is the sensitivity, defined as the ratio of the produced current and the pressure impinging on the membrane,

$$S(j\omega) = \frac{I_{ac}(j\omega)}{\bar{P}_{ac}(j\omega)} = \eta A_m \frac{j\omega}{k_m + j\omega b + \omega^2 m}.$$

Where ω is the pulsation, η the electromechanical coupling factor, Z_a the acoustic impedance of air, A_m the area of the membrane, k the stiffness of the membrane, b the damping coefficient and m the effective mass.
All the parameters of the model that are subject to process variation and whose precise knowledge is of paramount importance for the coupling to the electronics are obtained by means of an electromechanical characterization.
A wide selection of devices have been at disposal, with different dimensions, materials and working principles; as a consequence to optimize the ultrasonic testing system towards a specific application the most reliable and promising devices have been used in the transmission experiment. The experimentation system consists in a set-up for emission and detection of ultrasonic waves from two separated transducers placed in front of each other, with the possibility to change their relative positions with high accuracy. A preliminary study of signal shapes suggests to use a burst signal of few tens of cycles at the resonance frequency of the transducers for the voltage driving the emitter.
As a first result the emission efficiency and the detection sensitivity were measured at resonance to be *E = 48 Pa/V* and *S = 389 nA/Pa*, in accordance with the predictions.
The low-noise electronics designed has proven to have a lower noise floor than the transducer, allowing to measure the mechanical Brownian noise that sets the ultimate limit to the maximum distance for the ultrasound to be detected. This limit together with the minimum distance, which is related to the saturation of the sensing electronics, sets the operative range of the devices (~ 5 cm).
Another limit is imposed by the high directionality of the multi-membrane transducers, that, due to interference between each element's signal, has been measured to be 14° for the selected devices. The emission profile of the CMUTs is shown in Figure 2.



1. Principle scheme of a CMUT (on the left) and of a PMUT (on the right).



2. Emission profile of a CMUT transducer, where the color is the intensity of the emitted pressure from the source, placed in the position (0,0).

Lower frequencies PMUTs are foreseen to overcome these limitation and to work at longer distances, but at the moment of the writing the production technology has yet to reach the reliability and repeatability required.
Two proof of concepts of the envisioned applications have been provided with the same acoustic characterization set-up, exploiting an array of receivers, aligned side by side: (1) an ultrasonic source were correctly detected, with relative errors below 9 % in a range between 1 cm and 4.5 cm; the reduced range with respect to the one achievable by the single transducer is due to the angle between the position of the source and the normal to the surface membrane; and (2) a sound image (like in a sonar), were reproduced detecting the presence of a source and reproducing its shape; the limited contrast that these second example has shown is a consequence of the simple custom algorithm (as this was not the aim of the work), which is poor compared to industrial patented imaging algorithm.
The results discussed have been object of two works, one regarding the effect of different materials and radii and other demonstrating the localization performances in air, presented at IUS 2013 and Eurosensors 2014 respectively.

ACHIEVING QUALITY IN CROWDSOURCING THROUGH TASK DESIGN AND ASSIGNMENT

Ilio Catallo - Advisor: **Piero Fraternali**

Crowdsourcing, i.e., the assembling of strangers to accomplish a task, has the potential to revolutionize the way people work on the web. The promotion of crowdsourcing initiatives allow companies to easily collect and compound contributions in a distributed fashion, while letting individuals work and earn without the need for a physical working place or pre-existing employment contracts. Thanks to its high flexibility, crowdsourcing is gaining a more prominent role in both the industry and academia, and it has been estimated that companies have the potential to crowdsource more than 300 billion USD of work worldwide. As the number of organizations embracing crowdsourcing is increasing, crowd workers are likely to become a limited resource. An important issue is therefore to understand how to obtain, retain and persuade a crowd to contribute. In this work, we are especially interested in understanding which mechanisms are effective for eliciting high quality contributions from the crowd. Our approach is twofold. On the one hand, we focus on how a careful task design can help improve quality of contributions. We present a characterization of the design space of crowdsourcing tasks, and we then contrast the capabilities offered by the commercially-available platforms against the proposed characterization. On the other hand, we turn our attention to the problem of assigning tasks to crowd workers. In this respect, carefully considering workers' accuracy has already proved to be the key enabler for increasing task quality. We therefore propose a task assignment policy to support the assignment of tasks in relation to crowd workers' abilities, under the assumption that workers may exhibit varying accuracy depending on their workload. We validate our findings through an extensive experimental phase. Specifically, we conduct experiments with the aim of verifying which task design dimensions affect the quality of the outcome. Moreover, we offer empirical evidence of the existence of a fatigue/learning phenomenon among workers, and we extensively validate the proposed task assignment procedure against both synthetic and real data.

EXPLORING ARCHITECTURAL SUPPORT FOR APPLICATIONS WITH IRREGULAR MEMORY PATTERNS ON DISTRIBUTED MANYCORE SYSTEMS

Marco Ceriani - Supervisor: **Gianluca Palermo**

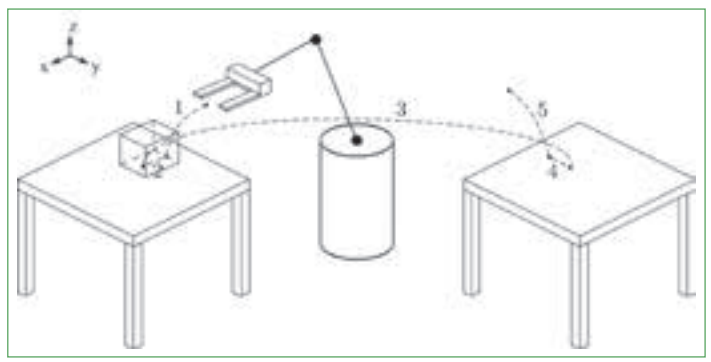
The interest on large scale knowledge discovery applications is rapidly increasing in both industry and scientific research. Example of such applications are found in the fields of social network analysis, data mining, protein sequences analysis and in the study of interactions in biological and other complex systems. The data analyzed by these applications are naturally represented by dynamic and irregular data structures such as graphs or unbalanced trees. The algorithms used, therefore, show an irregular behaviour in both control and data patterns, which have a degrading impact on the performance. One of the most limiting factors is the very poor spatial and temporal locality of memory accesses caused by the unordered traversals of the data structure. Many analysis algorithms require a partially ordered visit of the entire graph, for example in breadth first order, or random visits to the neighbourhood of each node. This kind of visits generates irregular patterns which cannot be predicted by a memory prefetcher, and because of the low temporally locality of the data most of memory accesses hits the main memory, making complex cache hierarchies essentially useless.
In addition, because of the small world phenomenon, data obtained from measuring social and economic activities often produces scale-free graphs, that is graphs whose degree distribution approximates a power law. These graphs have an ultra small diameter, which means that the maximum number of steps required to reach the furthest node is extremely small with respect to the number of nodes in the graph. This high interconnectivity of the structures makes it difficult to optimize the partitioning of large data sets on distributed memory machines. Therefore this class of algorithms do not perform well on current architectures for High Performance Computing (HPC), which are designed to exploit locality of accesses and regularities in the control flow. The only family of system architectures specifically designed for efficient execution of graph-based applications has been proposed by Cray with the MTA and XMT supercomputers. But they are highly expensive because they are composed of mostly custom components, which prevents exploiting economies of scale. Also, because of the necessary trade-offs, they achieve reduced performance on more regular workloads based on linear algebra and matrix operations, which are used in most of the engineering and scientific programs that require HPC technologies.
The objective of the research has been to study how to extend commodity distributed architectures with the double aim of improving the performance of applications with irregular memory patterns, and simplifying the programming model.
The approach followed is the design of an architecture template based on many-core processors, which extends commodity many core architectures introducing a small number of custom components. The new components enrich the processor functionality by providing features useful for running large-scale graph-based applications, but they do not require a modification of the core internal architecture or the memory hierarchy. Hence, the resulting architecture can perform equally well with regular and irregular applications, by enabling or disabling the additional components.
A set of four features are the keys of the architecture design. The first is a global and distributed address space, which allows the use of the Shared-Memory (SM) Single Program Multiple Data (SIMD) programming model. The second is the probabilistic reduction of dynamic hotspots, through a non-linear scrambled mapping of global memory addresses to the nodes of the distributed system. The third feature is the automatic exploitation of lightweight multi-threading to transparently tolerate the long latency of remote memory requests. Finally, the support for fine-grained synchronization, thanks to a hardware implementation of locking routines that allows the threads to lock single words of the global address space.
The feasibility and effectiveness of the approach has been evaluated with a prototype composed of 4 FPGA devices connected by an ad-hoc network. The prototype includes the custom-designed components, which are added to an architecture based on off-the-shelf cores and communication subsystem. These components offer the four fore-mentioned features with a minimal impact on the chip architecture. In spite of the small size of the prototype, the performance scaling of typical irregular kernels proved the effectiveness of the approach. In addition, the FPGA prototype allowed to evaluate the technical issues related to the implementation of the proposed architecture, suggesting the technical details required for supporting other commodity processors.
The performance data obtained from the prototype have also be used to formulate an analytical model, which identify the system bottlenecks and can be used for dimensioning a large scale distributed system.
The development of FPGA prototypes allows to reduce the design and evaluation cycle of hardware architecture with respect to full ASIC designs. However, they still require a considerable effort and amount of time, which repeats and increases for each incremental update. Therefore the research has moved on with the creation of a lightweight system simulator, in order to evaluate additional features using high level performance models instead of a full HDL implementation. The simulator neglects modeling the details of cache and memory hierarchy, which are irrelevant for applications which lack data locality, thus improving the simulation speed. On the other hand, it models the extended memory behavior introduced by the custom hardware components: the automatic identification of remote accesses that triggers a context switch, the scrambled address space and the fine-grained locking routines. The simulator allowed to evaluate the impact of additional architectural features, such as the support for atomic operations on the global address space. The use of remote atomic operations, in place of lock routines, allowed to significantly reduce the synchronization overhead and exposed larger amounts of parallelism enhancing the effectiveness of the many-core system.

A TASK-CONSISTENT SAFETY SYSTEM FOR CLOSE COOPERATION BETWEEN HUMAN WORKERS AND DUAL ARM ROBOTIC MANIPULATORS

Nicola Maria Ceriani - Supervisor: **Prof. Paolo Rocco**

Industrial manipulators represent a significant element for the automation of some industry sectors, such as cars manufacturing, or for machines tending and parts movement. However, their diffusion in productive settings such as consumer electronics industry is hampered by an insufficient flexibility, or by an excessive cost. Human robot interaction is a promising solution to such a problem, as cooperation between robots and workers could greatly increase robots flexibility, and, at the same time, the adoption of manipulators that are safe for human robot interaction would reduce the costs related to environment structuring.
The deployment of industrial robots in human robot collaboration scenarios poses new challenges for robot manufacturers: guaranteeing safety for human operators cooperating with robots, while achieving productivity in unstructured environments. Robots should appear friendly to workers, avoid collisions and reduce the risk of consequent injuries. At the same time, the pursue of safety must not diminish robots productivity, nor should it disrupt the possibility of task completion or generate a risk of damages for the manipulator or the production setup.

Industrial robots specifically designed for human robot interaction are becoming available in the market. However, industrial robot controllers currently lack the features needed to ensure safe and productive human robot interaction.
This thesis aims at extending an industrial controller functionalities with a collision avoidance system, in order to allow the robot operation in unstructured environments and in close cooperation with humans. This research therefore contributes to the use of industrial robots in new production scenarios and proposes a safety system which can be implemented adopting existing technologies, shortening the gap between research and application.
As a first step a classification for constraints composing a preplanned task, is proposed, which defines the relevance for task execution of constraints, and consequently identifies the possibilities for their relaxation. Once relevance of constraints has been identified, a preplanned trajectory can be modified and adapted to the current conditions of the environment, still preserving the possibility of its successful completion.
Then, a strategy for task consistent collision avoidance is proposed, which is based on the previously presented classification and executes evasive motions exploiting relaxed constraints. For this purpose, an assessment of danger generated by the robot on obstacles is adopted: constraints of higher relevance are relaxed as the level of danger increases. Then, exploiting the above mentioned danger assessment, evasive velocities are computed, that are used to avoid the detected obstacles.
Finally, a state machine that allows to automatically design the collision avoidance strategy from the constraints classification is proposed.
In order to integrate the collision avoidance system with an industrial controller, a system for the communication between them is proposed. Such a system allows the robot programmer to exploit the added collision avoidance functionalities using the standard robot programming interface. The communication system demonstrates the possibility of extending a standard industrial controller with capabilities for the adaptation to unforeseen events.
Two different implementations of the collision avoidance strategy are then proposed. First, null space projection is used in order to execute evasive actions consistently with task constraints. Such an application of the collision avoidance strategy is experimentally validated on a dual arm force controlled assembly task. For this purpose, a distributed distance sensor prototype is designed and created, to endow a dual arm ABB FRIDA robot with obstacle sensing. The collision avoidance strategy effectiveness is demonstrated by the capability of evading from a human entering the robot workspace through the modification of the preplanned robot task.
Then, collision avoidance is formulated as an optimization problem. Kinematic limitations of the robot and unilateral operational space constraints can be effectively taken into account with such an approach, increasing the robustness and the effectiveness of the overall system. Two versions of the system are proposed, with an increasingly accurate management of kinematic limitations of the robot, and a deeper exploitation of the robot capabilities for evasion. The first version is experimentally validated on a dual arm pick and place task, while the second one is validated only through simulation.



1. A manipulation task is divided in elementary operations in order to classify its constraints.

2. The FRIDA dual arm robot with the distributed distance sensor, adopted with the first collision avoidance strategy.

3. The second collision avoidance strategy is applied on a dual arm task.

HUMANS IN THE LOOP – OPTIMIZATION OF ACTIVE AND PASSIVE CROWDSOURCING

Eleonora Ciceri - Supervisor: **Piero Fraternali**

In the last years, social media have attracted millions of users and have been integrated in people's daily practices. They enable users to create and share content or to participate in social networking.
User-generated content, i.e., the various forms of media assets publicly available and created by end-users, is published every day on the Web and mostly in social media at a massive scale, either in the form of textual documents (e.g., blog articles, posts on social networks, comments and discussion) or in the form of multimedia items (e.g., images and videos). Most user-generated content is about personal lives and facts about users. However, users often publish more structured and complex information.
Crowdsourcing has gained increasing importance in the last years. The term crowdsourcing generally refers to the outsourcing of a non-automatable task to people. The growth of the time spent online has led to a growth of interest in crowdsourcing. Several works have been developed, either making users actively participate in the resolution of tasks or exploiting data they generate and publish over the Web. We refer to these approaches as, respectively, active crowdsourcing (i.e., active participation of motivated users in task execution) and passive crowdsourcing (i.e., exploitation of user-generated content to extract useful information).
On the one hand, active crowdsourcing is the process of outsourcing tasks to a large group of people, called workers. In this scenario, human workers are asked to perform very specific tasks (called crowd tasks), which usually are easy to be solved by humans but hard to be solved by machines.
In the context of active crowdsourcing, only tasks difficult to be performed by a machine are submitted as crowd tasks. They are often based on uncertain data, since these data can hardly be processed by computers, due to their unstructured nature. Unfortunately, an appropriate modeling of the impact of a crowd task answer on uncertain data is yet to be defined. Moreover, similarly to the use of machine resources, which cost, also human computational resources are not freely available in any amount, and may provide erroneous answers. Consequently, an approach for the selection of the best candidate set of tasks to submit to the crowd under some fixed constraints (e.g., costs and time) needs to be devised, together with quality assurance procedures that guarantee an appropriate result quality level.
On the other hand, passive crowdsourcing denotes an alternative approach for leveraging the online activity of users for task resolution, which amounts to analyzing a huge amount of publicly available contents, to extract information about behaviors, interests and activities of the social media population. Researchers from different fields (e.g., social science, economy and marketing) analyze a variety of user-generated datasets to understand human behaviors, find new trends in society and possibly formulate adequate policies in response.
However, due to the uncontrolled nature of users' participation on the Web, the huge mass of available data contains replicated information, as well as low quality or irrelevant content. Moreover, content is often replicated maliciously: users copy content created by others (and often subject to copyright laws), rename it and pretend they are the authors of the corresponding original content.
In this Thesis, we propose methods to overcome these problems, both in the active and passive crowdsourcing field, with the objective of maximizing the quality of results, under the assumption of budget and time constraints.

SWITCHING, RELIABILITY AND NOVEL FUNCTIONALITIES IN PHASE CHANGE DEVICES

Nicola Ciocchini - Supervisor: **Daniele Ielmini**

The information revolution has totally changed our human life and society. The widespread diffusion of internet and mobile technologies had an impact on the world that is probably more dramatic than the invention of steam engines and cars at the begin of XX century. This revolution has been possible thanks to the big efforts put by electronic companies on the development of integrated circuits, according to Moore's law. The MOSFET transistor, at the basis of computation in electronic processors, was shrunk from the 20 μm channel of 1975 to the 14 nm channel of 2014: There is no other technology in history that was able to accomplish such a dramatic improvement in such a short time. In the Turing machine approach, the ability to store information (memory function) is as important as the ability to process the information (logic function). For this reason, the development of logic devices (transistors) came along with the development of memory devices. Modern PCs and mobile devices make use of several types of memory, which differ in terms of speed, cost and data retention time. Among these, the non-volatile memory must retain data for years also when the power supply is switched off. The dominant non-volatile memory technology in the past decades up to the present day has been the Flash memory, in which the bit of information is stored as an electric charge in the floating gate of a MOSFET device. Nowadays, the Flash technology is facing several issue related to scaling, among which the most important are random telegraph noise, electrostatic control of floating gate and variability. For this reason, memory companies are trying to find alternative solutions to Flash memory. One of the most promising technology, which has already reached the industrial maturity, is the phase change memory (PCM). The PCM is a particular type of resistive memory, where the reversible phase transition of the active chalcogenide material, usually Ge2Sb2Te5 (GST), is used to store the logic bit of information. The two stable states in the memory correspond to the high-resistance amorphous phase (reset state) and the low resistance crystalline phase (set state). PCM devices have been scaled to the 20nm size, while low-power and nanosecond-switching operation has been demonstrated. On the other hand, the ultimate scaling of the PCM is still unclear, due to the impact of random telegraph noise, crystallization, and resistance drift. Nowadays, a deeper knowledge of the PCM physics is strongly requested to drive the development of the PCM technology in the years to come. This motivates the need for research activities, such as the ones described in this doctoral dissertation.

The introductory chapter of this thesis provides an overview of the current non-volatile memory (NVM) scenario, subdividing the possible technology evolutions within an evolutionary scenario and a paradigm shift. The phase-change technology is then introduced, dealing with its history, the basic operation and the elementary physical description. This chapter reviews the current state-of-the-art in the physical comprehension of sub-threshold conduction, threshold switching, crystallization and structural relaxation, providing the basic elements needed for the comprehension of the following four chapters. Finally, the current perspectives of the PCM technology are discussed, with a quick glance on the so-called PCMS architecture, which is expected to solve the current PCM limitations in terms of size scaling by stacking the memory element and a cell selector made of another chalcogenide material.

The second chapter is devoted to the study of structural relaxation-related phenomena in the amorphous phase of phase change memory devices. The chapter is particularly focused on the characterization and modeling of threshold voltage drift induced by structural relaxation in amorphous Ge2Sb2Te5. This effect leads to threshold voltage increase with time after program operation, which must be carefully controlled to avoid program/read failure in the memory device, especially in PCMS, where both the selector and the memory devices are affected by drift. Then, we show how it is possible to accelerate threshold voltage drift by mean of electrical pulses in the subthreshold region, which could represent an important tool from the application point view to limit the drift effect in the memory device. Finally, the study is extended to the modeling of resistance drift in the crystalline state of Ge-rich Ge-Sb-Te alloys for embedded non-volatile memory applications (ePCM). We show evidence of resistance drift and decay, which are attributed in our model to structural relaxation at the grain boundaries of the poly-crystalline state, and to grain boundaries coalescence respectively.

The third chapter of this thesis deals with a detailed study of the retention capability in PCM on a large statistical scale. Such studies are fundamental in order to allow large arrays to properly satisfy the data retention requirements. A wide experimental characterization is then provided in the temperature range below 180°C, presenting a detailed study of the cell-to-cell and cycle-to-cycle variability. The overall variability is interpreted through a compact Monte-Carlo model, able to explain both the cell-to-cell and the cycle-to-cycle variability contributions in terms of a pure gaussian spread in the activation energy for crystallization. Finally, the cycle-to-cycle variability is analyzed more in depth, allowing to subdivide the retention characteristics into three separate families namely: i) analog variability, ii) digital (binary) variability and iii) pseudo-repeatable characteristics.

The fourth chapter is dedicated to the study of crystallization kinetics in PCM. Our work shows evidence of non-Arrhenius crystallization in GST directly in PCM devices, by comparing the thermally induced crystallization (thermal regime) with the electrically induced crystallization (pulsed regime). The non-Arrhenius crystallization, leading to different activation energies in the Arrhenius plot of crystallization time in the two regimes, is attributed to the fragile nature of GST glass and to the broke of Stoke-Einstein relation above glass transition. We propose a new experimental technique to study electrically induced crystallization down to the holding current. In this way, we were able to extend our study of crystallization kinetics, and to characterize set transition in a wide time range from 50 ns to 10 μs. Then, we model crystallization in PCM by a finite element approach, which is based on filamentary crystallization after threshold switching and on non-Arrhenius crystallization kinetics. Finally, we show evidence of electrically induced crystallization in the subthreshold regime, by performing continuous current stress experiments at low current $\approx$ 1 μA and for relatively long times, in the range of 10^3 s.

The fifth chapter deals with a possible alternative application of PCM devices. We exploit the storage ability of PCM, together with additive crystallization and threshold switching, to perform boolean logic operation. In our approach, the PCM device is used as a state machine (memristor), where the state of the device can be changed by mean of electrical pulses applied to the cell. In our work we are able to accomplish a complete set of boolean logic operations, namely the NOT, the NAND and the NOR operation. When compared to standard CMOS logic, the PCM logic offers the advantages of logic-in-memory, of reconfigurable logic and zero static power dissipation, while is shows worse performance in terms of dynamic power consumption, switching time and endurance. This work paves the way for a new field of application for PCM, which together with neuromorphic computation makes this technology attractive for alternative way to compute information in the big data era of the present days.

A GENERAL SENSOR-FUSION AND PARAMETERS SELF-CALIBRATION FRAMEWORK WITH APPLICATIONS IN MOBILE ROBOTICS

Davide Antonio Cucci - Supervisor: **Matteo Matteucci**

Robots are mechanical or virtual artificial agents able to perform given tasks with a certain degree of autonomy. These tasks always involve interaction with the environment, whatever it is our familiar physical world or some virtual scenario.

In principle, most of these tasks could be performed relatively easily if only the robot knew certain quantities such as its own position, the position of its goal, the current distance from walls, if the planned path towards the goal is free from obstacles, and where possible obstacles are located. Unfortunately, these variables are seldom directly observable in practice. Moreover, even in scenarios where the operating conditions, such as the light conditions, or the site map, can be controlled or jointly designed with the robotic system, there will always be inescapable degrees of uncertainty in the robot and environment state.

In order to bound the uncertainty in their knowledge, most of the modern robots employ sensors and maintain an internal model of the state of the world; this model is updated according to observation evidence and it is then employed to make decisions about how to accomplish the assigned tasks. The set of sensors available for the robot is always chosen depending on the operating environment and on the required degree of autonomy. In general, as the assigned tasks grow in complexity, the set of variables that have to be observed increases and multiple sensors are required. As heterogeneous sensors observe different aspects of the reality, redundancy in perception results in an increased fault tolerance and robustness with respect to unforeseen situations.

As an example, let us consider the case of robotic systems for space exploration, such as the Curiosity rover, and part of the Mars Science Laboratory mission. Because of the delay in communication between Earth and Mars, the robot can not be teleoperated, and waiting for human instructions in each unforeseen situation is clearly impractical. Thus, the robot needs a high degree of autonomy, at least for elementary tasks such as heading towards given positions and obstacle avoidance. To this end, several sensors are employed; leaving apart the scientific experiments and general purpose elements such as the mastcam, a multi-spectra, high definition camera, the rover has two pairs of navcams, to acquire stereoscopic 3-D images, plus four pairs of hazcams, which are used for autonomous hazard avoidance and safe positioning of the robot arm, for a total of twelve cameras employed in navigation, plus an inertial measurement unit.

As new, noisy, possibly contradictory, evidence comes from multiple, heterogeneous, sensors, processing has to be applied in order to fuse the available information and update the robot internal model of the world. This is model is often called belief, and modern robots also include an explicit characterization of the uncertainty regarding variables which are critical with respect to their tasks. The problem of how these internal believes can be consistently updated as new observations become available has been subject of active research in the last fifty years, and it is still ongoing. Many techniques have been proposed and effectively employed in several applications. Notable examples are the Extended Kalman Filters, or, in general, recursive Bayes filters, and, more recently, graph-based optimization techniques.

However, hardware sensors, or pre-processing to be applied on raw data, often involve calibration or tuning parameters that turn out to be critical to build internal robot believes. For instance, on a mobile robot with a single camera, it is impossible to estimate the robot velocity from successive frames unless we know the orientation of the camera with respect to the robot base. Other examples are gains and biases in inertial measurement units, ferromagnetic properties of the robot affecting magnetometer readings, intrinsic matrices and depth distortion pattern of a RGB-D cameras, to name a few. To ensure that sound and consistent state estimation can be achieved, it is often required to determine these parameters with a high degree of accuracy. However, it is often difficult to to determine these by directly inspecting the robot (think about the case of 3-DoF orientations), while others are simply not directly observable, e.g., the matrices of intrinsic camera parameters.

A number of ad-hoc solutions has been proposed in the literature to handle accurate calibration of very specific sensor configurations (and they are still subject of active investigation). These techniques often rely on artificial environment structures, such as checkerboards in camera calibration, or on the availability of external information, not produced by the set of sensors being calibrated, such as position ground truth. Unfortunately, relevant parameters might change over time, such as biases in gyroscope sensors, which depend on environment temperature, motion, and on a number of other factors. In this cases, offline, ad-hoc, calibration procedures, and environment structures, can not be employed and typical solutions require the robot state space to be augmented to include estimates for calibration parameters. Few work has been done on the self-calibration of an arbitrary set of sensor, i.e., the problem of determining sensor calibration parameters employing only the information produced by the sensors themselves, possibly, during normal robot operation.

As robotic systems face new and more advanced tasks, system developers and researchers are required to handle very complex sensor-fusion and parameter calibration problems. Despite the wide variety of solutions available in the literature, platform dependent specifications make them not directly applicable, or require adaptations, enhancement, or substantial extensions. The lack of off-the-shelf, flexible solutions which are deployable with minor effort undermine the availability of baseline solutions to compare new approaches against and often requires researchers to develop from scratch even very simple sensor-fusion algorithms, "reinventing the wheel" and scarifying reusability.

In this work we introduce ROAMFREE, a general, open-source, framework for multi-sensor fusion and parameter self-calibration in mobile robotics. In its development, we employ and extend mathematical and software engineering techniques to ensure that the resulting framework can be easily specialized to handle specific cases, and some of its component replaced without any change to the overall system architecture. A comprehensive logical sensors library allows to abstract from the actual hardware and processing while preserving model accuracy thanks to a rich set of calibration parameters (e.g., biases, gains, distortion matrices and geometric placement characterization).

A modular formulation of the information fusion problem has been obtained based on state-of-the-art factor-graph inference techniques; it allows to handle arbitrary number of multi-rate sensors and to adapt to virtually any kind of mobile robot platforms, such as Ackerman steering vehicles, quadrotor unmanned aerial vehicles, omni-directional mobile robots. Different solvers are available to target both high-rate online pose tracking tasks and offline accurate trajectory smoothing and parameter self-calibration.

An extensive evaluation of the resulting framework has been performed on different mobile robots. ROAMFREE has already proved its flexibility and out-of-the-box deployment in several, real-world, information fusion and sensor self-calibration problems.

STUDY OF A COLLABORATIVE REPOSITORY OF SEMANTIC METADATA AND MODELS FOR REGIONAL ENVIRONMENTAL DATASETS' MULTIVARIATE TRANSFORMATIONS

Daniele De Rigo - Supervisor: **Giorgio Guariso -** Tutor: **Andrea Bonarini**

This work introduces a semantic modelling procedure to ease array-based multivariate transformations of public environmental data, along with the architecture of a collaborative repository of modelling meta-information based on the procedure. The procedure, *Semantic Array Programming* (SemAP), is intended as a lightweight paradigm to support integrated natural resources modelling and management (INRMM), in the context of wide-scale transdisciplinary modelling for environment (WSTMe, here tested from catchment up to regional and continental scale).
It is a common experience among computational scientists, to codify even short algorithms – if no out-of-the-box solutions are available – with remarkably longer implementations.
Computational science algorithms not rarely deal with large amounts of data with a precise (despite sometime nontrivial) semantic structure. If so, data may be organised in multiple groups with homogeneous semantic content. Examples of such groups are matrices, time series, tuples, graphs or more generic multi-dimensional arrays. Geospatial problems may often associate geographic information to particular arrays: for example, spatial regular grids of data are frequently represented by georeferenced matrices.
Domain specific frameworks may offer a convenient option for dealing with standard problems within a given sectoral domain. Object oriented approaches might enable information to be represented and transformed in sophisticate flexible ways. The objects of a monolithic model are typically straightforward to propagate and effective in transferring structured information with default behaviours/assumptions. However, this communication is more difficult to achieve for non-monolithic models using several programming languages and tools, with multiple teams involved and possibly no single expert able to cope with the overall integration complexity.
Within a particular discipline, a particular research team, or specialised modelling approach, a significant part of the overall information on the semantics of data and data-transformation models (D-TM) may be taken for granted. This means that a core base of knowledge might safely remain unexpressed among experts in the same domain. Unfortunately, this is not the case whenever that particular domain of knowledge has to interact with other domains, perhaps quite far from it. Namely, when a set of practices and knowledge – shared by a certain research community – has to be relativized from *universal set* of the research activities up to become a simple specialized *module* within a transdisciplinary context, the *common sense* evaporates. As a consequence, it should be communicated in a simple – but also compact and unambiguous – cross-disciplinary way.
Array Programming (AP) might support part of this task. AP originated for reducing the gap between mathematical formulation and code implementation by introducing very concise operators and coding patterns to deal with variables potentially composed by billions of elements and considered as atomic (with correspondingly terse manipulation). AP data structures can offer a support

- already widespread (given the extensive use of AP languages, e.g. MATLAB/GNU Octave, GNU R, Python with Numpy/Scipy, IDL), and
- noticeably less arbitrary/restrictive than a particular choice (within a virtually infinite set of possibilities) of objects to be shared among multiple and highly heterogeneous modules.

However, this support is still poorly exploited.
The AP data structures are very general: multi-dimensional arrays where the value of some elements may be *infinite* or *not-a-number* (IEEE 754 standard) or even complex-valued (e.g. two-dimensional vectors – the velocity of wind or other fluids, the spatial gradient of several quantities). From this *potentially* overwhelming generality it follows the basic idea of Semantic Array Programming: limiting this generality with array-based semantic constraints. The second key idea of SemAP is to encourage modularisation of data-transformations so as to easily propagate the semantic support to lower-level sub-D-TMs – which might prove helpful even to better explore software uncertainty. Modularisation may be seen as an abstract semantic constraint in order for the array of components of a given algorithm (i.e. their sub-D-TMs) to be made explicit in the algorithm implementation.
SemAP has been formalised and expanded to address geospatial problems by means of a problem-driven approach focused on the broad heterogeneity in the European continent. Real world case studies illustrate collaborative applications to WSTMe problems in Europe. The case studies are essential to build the reference repertoire to serve as a guidance for supporting the community of scientists involved in applying the modelling procedure, which is meant to drive a *collaborative, peer-reviewed repository of metadata and data-transformation models* related to web-available environmental datasets. The transdisciplinary collaborations in developing the case studies consolidated the growing research community interested in complementing existing approaches with the SemAP paradigm, especially considering geospatial applications (*Geospatial Semantic Array Programming*, GeoSemAP).
The case studies constitute the premise for the future real world presentation and exploitation of the repository.
First, a pan-European application is introduced for estimating soil erosion by water. SemAP is applied to extend a well-established environmental model (RUSLE), both to its general multiplicative structure (whose factors are extended exploiting the array semantics of the problem at model-scale) and to a specific critical module (where array-based multiplicity is introduced for mitigating extrapolation errors at module-scale with a robust ensemble approach).
The second case study deals with the temporal dynamics of a complex modelling and management problem under deep uncertainty. A SemAP-enhanced modelling architecture is introduced for large wildfire behaviour prediction, assessment and control, focusing on the multiple heterogeneous sources of uncertainty. A data-driven robust modelling explores the array of multiple fuel models, meteorological disturbances and fire control strategies.
Despite the local spatial scale of wildfires, their impacts may far exceed this scale (off site impacts): the architecture is explicitly designed in order for large events in the European continent to become susceptible of real-time rapid (i.e. approximated) assessment.
The third application characterises landscape patterns to support biodiversity policy-making with a multiplicity of approaches (indices).
An integrated multicriteria characterisation of connectivity and fragmentation is demonstrated for the forest focal habitat. An array of known connectivity indices is generalised in a new SemAP-implemented family to emphasize their implicit common structure while also suggesting new indices. Other index families are analysed with the help of a GeoSemAP workflow. A SemAP-based nonlinear statistical analysis (brownian distance correlation) shows the least correlated indices.
The first three case-studies deal with self-standing topics. The fourth one highlights the flexible reusability of specific SemAP techniques as specialised modules. Relative Distance Similarity illustrates the application of robust ensemble methods.
Finally, statistics are presented on the number of unique authors who contributed to SemAP-supported publications along with unique authors citing these works. Despite the intrinsic variability, these initial data show how the proposed paradigm is currently under active expansion within the community of potentially interested researchers.

MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM, DYNAMIC AND NON-DYNAMIC EMULATORS IN THE DESIGN OF OPTIMAL POLICIES FOR WATER RESOURCES MANAGEMENT

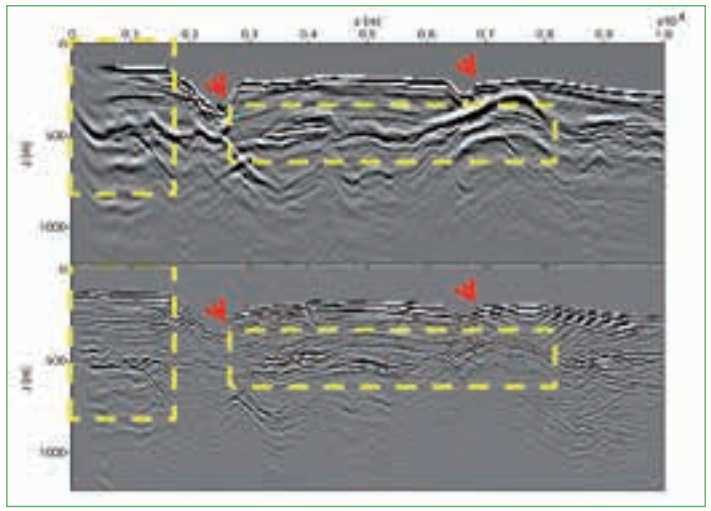
Quang Dinh - Supervisor: **Rodolfo Soncini-Sessa**

Dams and reservoirs are essential to satisfy human demand of water for a plurality of uses: irrigation, industrial and domestic supply, hydropower production, flood mitigation and recreation. The management of these multi-purposes systems is difficult since it has to deal with the relevant conflicts between those interests. The definition of a satisfactory and stable compromise requires the design of regulation policies of the reservoirs, which every day objectively establish the release to be operated from each reservoir, taking care of all the interests. The design of such policies requires to set up and solve a complex Multi-Objectives Control Problem, in which the interest of the Parties are quantified by indicators, the values of which are estimated with large, physically-based models that describe the effects of the water distribution in the downstream user systems (canals, hydropower plants, irrigation districts and municipalities).
However, even more difficult and challenging problems are posed by dam management. The construction of large dams has a strong impacts on the hydrological regime of and the movement of sediments in the regulated river. Dams affect the hydrological regime primarily through changes in timing, magnitude and frequency of high and low flows; the peak discharge is generally and intentionally significantly reduced. The trapping of sediment in the reservoirs causes sediment deficit downstream, which in turn triggers not only erosion on the river banks but also incision in the river bed. This process can extend over hundreds of kilometers and last for decades.
Starting from the pioneering Harvard Water Project in 1962, the research community has set a large effort on building more and more complex models to estimate the effects of different planning and management decisions. A vast amount of studies have been carried out on these effects and nowadays complex physically-based models are available to forecast and quantize them. Nowadays, very detailed and precise models can be easily set up, with which the evaluation of the effects, even subtle as the geomorphological ones, is possible. At the same time the art of defining and solving complex Mathematical or control Problems for policy design has reached an adult stage. However how to merge large physically-based models within a Design Problem is still an open issue. In the last years a possible solution approach was offered: the Emulation Modeling techniques, which gives the possibility of substituting the high-dimensional physically-based, distributed-parameters models with low-dimensional lumped-parameter ones. These techniques have been proposed in the Technological Sectors (like Aerospace Industry) and have up to now found just a series of application in the water world. The goal of this thesis is to explore the application of Dynamic and Non-dynamic Emulation Modeling techniques to design optimal policies of water management, considering the most important objectives of the considered case study.
Keywords: Reservoir operation, Multi-objective optimization, Response Surface, Dynamic Emulation, hydro-morphological model

SEISMIC PROCESSING AND IMAGING OF MULTIPLE REFLECTIONS

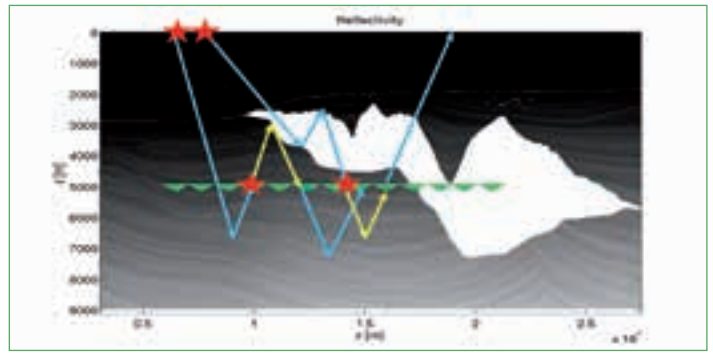
Carlo Fortini - Supervisor: **Giuseppe Drufuca**

Conventional seismic imaging algorithms are based on single-scattering hypothesis. The reverberations that generate during wavefield propagation in the subsurface (multiple reflections) are usually considered as unwanted noise. Even though the most energetic reverberations are the so called surface-related multiples, in areas of high structural complexity we usually record also strong internal and intra-salt multiples. The most common approach to handle with the presence of multiple reflections in the acquired seismic data is to try to eliminate them. However, the multiply scattered recorded events had interacted with the subsurface discontinuities, thus carrying useful information about it. When properly imaged, they can enhance the seismic image and may provide additional illumination in those areas poorly imaged by conventional approaches. Recently, researches on multiples have shifted their focus on the exploitation on what has been often considered only as noise. Instead of considering the reverberations only as noise, it is possible to use them as a source of valuable information. Both the most common migration algorithms, Wave Equation Migration (WEM) and Reverse Time Migration (RTM), can be modified in order to include multiple events in the migration procedure; the image obtained from multiples can be used to complement the primary image. In the first chapter of my thesis, I show the mathematical derivation on which the migration of multiples is based. I show with examples on both synthetic and field dataset, the main benefits and drawbacks of the proposed methodology. The main issue in using the multiple reflections is the cross-talk noise: the migration procedure returns the correct image plus some artifacts related to the cross-talk of primaries of one reflector and multiples belonging to some other interface. In chapter three and four I present two techniques that can be used to identify and construct the model of the cross-talk events. One method is based on the distinctive features that primary and cross-talk events show in the pre-stack domains, and relies on an advanced interpolation methodology. With second methodology presented, I show that the cross-talk terms can be eliminated during the migration process by modifying the imaging condition. The new imaging condition operates in an higher dimensional space with respect to the conventional cross-correlation imaging condition: the source and receiver wavefields are cross-correlated not only in time but also based on their local coherency in the time-space domain.



1. Comparison between the image obtained with primaries and with multiples. The red arrows point to some of the locations where the multiples migration is inferior to the conventional one. The areas highlighted with the yellow boxes, are those where the migration of multiples provide the better results with respect to the image of the primaries.

The last two chapter of my thesis deal with the internal multiples. In the first chapter I used the Born scattering theory to show how to exploit the information contained in the surface-related multiples when performing the seismic imaging. In doing so, however, I didn't take into account the internal multiple events. In areas of high structural complexity, though, for instance in presence of salt bodies, strong internal and intra-salt multiples are usually recorded. When properly imaged, these events can enhance the seismic image and provide additional illumination in those areas poorly imaged by conventional approaches. I propose a technique based on non-linear seismic interferometry, that allows to separately exploit the information coming from the internal multiples recorded by a conventional seismic acquisition. The original wavefield is used as input to reconstruct a new virtual seismic experiment with sources and receivers placed underneath the formations mainly responsible of the multiply scattered events. This new dataset is constructed using both the linear and non-linear components of the Green's function, thus accounting for the internal multiple reflections. Although multiples have been proven to be useful signal, their prediction and attenuation are still fundamental tasks. The migration of surface-related multiples can, in fact, benefit from a separation between primaries and multiples. Moreover the identification of the reflector mainly responsible for the generation of the reverberations, is very useful. In the last chapter, I propose a method for estimating both surface-related and interbed multiple artifacts, using as input the migrated section and the subsurface velocity model. The technique relies on the use of seismic demigration, it is faster than the conventional multiples prediction methods, independent from the original acquisition geometry and it returns an image of the unwanted artifacts that can be compared directly with the subsurface image obtained through the migration of the recorded data. The method is suitable for the multiple artifacts estimation and also as an interpretation tool for the identification of the horizons responsible of the internal multiples generation.



2. The internal multiples are exploit to reconstruct primaries that provide additional illumination of the subsurface. The events of interest are indicated with the yellow arrows.

METHODOLOGIES AND TOOLS FOR GAMES WITH A PURPOSE DESIGN AND GAMIFIED APPLICATIONS

Luca Galli - Advisor: **Fraternali Piero**

Human computation is a research area that focuses on exploiting human intelligence to solve computational problems that are beyond the capacity of existing Artificial Intelligence (AI) algorithms. The growth of the Web and Social Networks provides a massive amount of persons that can be leveraged to perform complex tasks, but a fundamental issue in exploiting the contribution of crowds is how to engage the potential users for the specified purposes and how to ensure the quality of their contribution. To overcome the problem, a set of approaches have been developed; Games with a Purpose (GWAPs) are digital games in which the players' actions in the game contribute to a real-world purpose outside of the game, whether it be predicting protein structures or providing labels for images. The standard way to accomplish the same type of work is to "crowdsource" the work directly using a service like Mechanical Turk in which contributors are paid as workers. To address the lack of extrinsic motivation that plagues traditional human computation platforms, GWAP provide intrinsic motivation in the form of entertainment. Many GWAP have been developed since the release of the first instance, the ESP Game in 2003. But not all GWAP seem to have lived up to the initial hype of transforming millions of hours typically poured into traditional games into useful and productive work.
The problem that GWAP have faced since their inception is related to the fact that the very fundamental mechanisms on which they rely on, to guarantee the quality of the submitted results, have been considered as "Game Mechanics" while in reality they are simply validation mechanisms. For this reason, even the most famous GWAP were centered on experiences that aimed at maximizing the throughput of high quality submitted content instead of focusing on the entertainment dimension typical of other digital games, producing applications that were perceived as "non games" by their users.
As it happened with GWAP, gamification, the process of using game design techniques and game mechanics to enhance traditional applications, has been able to accomplish significative results but also catastrophic failures. Once again, this phenomenon has to be attributed to poor design due to the lack of guidelines and best practices to support the development.
The main reason is the inherent difficulty of the design of both GWAP and gamified applications, which resides in the tradeoff between purposiveness and playfulness: in a traditional application, the improper insertion of gaming elements may result artificial and thus not produce the desired engagement effects, while on the contrary spoiling the user's productivity, symmetrically in a GWAP the task to be solved may mismatch with the game mechanics, thus decreasing the "playability" of the game and failing to attract people and engage them in the execution of the task.
Another common challenge of human computation systems is data reliability. Humans are expected to be unreliable, especially in ludic environments where a playful interaction with the system to test its borders is expected. Therefore, players may generate false data either on purpose or for other reasons. Different strategies have evolved to deal with this issue but they are typically tailored just to the particular task they have been applied to.
As human computation tasks are by definition not efficiently solvable by an algorithm, it is necessary to find new means to handle this challenge. The lack of established GWAP design paradigms, the difficulties of player engagement and retention and the issues of choosing or defining the right validation techniques in order to obtain meaningful results are limiting the capabilities that these systems may offer.
The proposed framework investigates the design of game mechanics and motivation techniques in games in order to solve human computation tasks by providing a set of tools that will be used to ease the development of interactive media applications that have to be integrated within media refinement tasks fulfilled by players. The work has also, dually, investigated the methodologies and approaches for gamification, that is the injection of game-like features in traditional applications (e.g. software development, customer relationship management) to improve key performance indicators.

ENERGY-AWARE TRAFFIC ENGINEERING FOR WIRED IP NETWORKS

Luca Giovanni Gianoli - Research director: **Prof. Brunilde Sansò**
Research co-director: **Prof. Antonio Capone**

In the last decade, the global widespread of ICT and telecommunication networks was sped up by the necessity of improving the efficiency of other sectors of the economy, while, at the same time, reducing their energy footprint. However, the massive development of ICT infrastructures and services has produced a significant increase of the energy foot-print of the ICT sector itself. To give an idea of the order of magnitude that we are talking about, in 2010, the yearly energy consumption of the world most important Internet Service Providers, e.g., AT&T and China Mobile, was over 11 TWh per year. Note that this is equivalent to the annual electricity production of a mid-sized nuclear power plant, or to almost half of the amount of electricity sold abroad by Hydro-Québec in 2011. Making telecommunications networks greener, in addition to reducing greenhouse gas emissions, may have significant economic impacts: for instance, for large companies like AT&T or Google, a modest 3% reduction of the electricity bill would result in several millions dollars of savings.
In our Ph.D. research project we addressed the problem of improving the energy efficiency of wired IP networks. The choice of wired IP networks among all the other possible domains, e.g., wireless networks, cloud or data center networks, sensor networks and so on, was motivated by two joint considerations which made us think that there were both need and room for novel significant contributions in this specific field:
(i) In 2011, while energy-awareness applied to other types of networks such as wireless had been already quite thoroughly explored, a very limited literature was available on how making wired IP networks energy efficient. This was quite surprising, since it was clear that wired IP networks too had a non negligible impact on the overall energy efficiency of ISPs.
(ii) Due to both hardware and management limitations, wired IP networks were identified as highly energy inefficient. Their consumption is always at the maximum level independently of network utilization. It was estimated that consumption reductions in the order of 50% would have been potentially achievable by optimizing the way wired IP networks are managed.
We addressed the problem of energy-awareness in IP networks in a very comprehensive way, aiming at dealing with all the important aspects related to both network management and energy-efficiency. We first identified the main strategy to pursue to make the consumption of IP networks proportional to the incoming traffic load, i.e., by putting to sleep the redundant network devices while guaranteeing that the active resources were able, thanks to an optimized configuration of the network routing, to guarantee the correct functioning of the network. The set of methods we developed can be identified as techniques for "sleep-based energy-aware network management" (SEANM). The sleeping-strategy was then applied and adapted to different network contexts with the aim of providing a comprehensive set of management tools to be used in any IP network according to the considered network configuration.
The strength of our research work lays in four main aspects:
(i) The depth and scope of our work, which, differently from what can be found in literature, covered the whole set of issues and topics which were relevant to energy-aware network management. We considered networks operated with different configuration settings and protocols, we studied the trade-off between network resilience and energy-awareness, we handled the problem of coping with network traffic uncertainty by implementing robust optimization techniques.
(ii) We filled a major gap in the literature by providing an accurate mathematical modeling for all the relevant SEANM problems.
(iii) We were the only one in literature to develop a real network management platform to implement and test the proposed approaches in realistic network environments.
(iv) We outperformed the results obtained by our competitors in those fields where state-of-the-art work were available. In all the addressed scenarios we were able to achieve energy savings from 20% to 60% according to the problem data and constraints. We were the first to formalize a way to model the allocation of network bandwidth in presence of elastic (TCP) traffic, providing thus the tools to correctly perform energy-aware network management in presence of TCP flows too.
Here follow a more detailed summary of the work done and of the the main results and produced along the whole Ph.D. project:
i) We developed a novel centralized off-line optimization algorithm (MILP-EWO) to save up to 60% of energy consumption in IP networks operated with the most popular shortest-path routing protocol, i.e., OSPF. The idea was to efficiently adjust the administrative link weights used to define the shortest paths among sources and destinations so as to exploit only the necessary network elements and consequently put to sleep the redundant ones. Proper constraints were respected to keep network congestion under control.
ii) We integrated MILP-EWO into a novel dynamic and centralized network management framework aiming at exploiting both off-line and online optimization to automatically adjust the link weight configuration. Tests carried on by means of emulated network environments proved the applicability and stability of the approach. To practically implement this extension of MILP-EWO, we developed a new open-source network management framework, i.e., JNetMan, offering a set of APIs to be used by network administrators to easily implement management policies based on SNMP commands.
iii) We developed both exact and heuristic methods for centralized SEANM in IP networks operated with flow-based routing protocols such as MPLS. We considered a multi-period scenario according to which a single day was split among six different macro periods characterized by a quite constant level of traffic. Limitations on the routing variability across consecutive time-periods and constraints to both guarantee quality of service and preserve device life-time were included. The proposed methods, which were applied to perform both off-line and on-line optimization, allowed to reduce the daily network energy consumption from 40% up to 60%.
iv) We extended the SEANM approaches for MPLS based network to explicitly guarantee resilience to single link failures and robustness to unpredictable traffic variations. Our goal was to study the trade-off between energy consumption and network survivability, while explicitly quantifying the energy cost of the latter. Tests showed that daily savings around 30% can be still achieved despite of the additional limitations on resiliency and robustness. We were able to affirm that the activation of redundant resources to cope with unexpected events costs up to 25% in terms of network consumption. To manage uncertain traffic demands we applied well known Robust Optimization techniques, such as cardinality constrained budget uncertainty.
v) We proposed a novel bi-level optimization model to correctly manage the allocation of elastic traffic demands, such as those carried by TCP. Our modeling framework exploited the concept of max-min-fairness and proportional fairness to correctly approximate the amount of bandwidth allocated by the network to each specific flow according to the capacity of the crossed link and the presence of other concurrent flows. We developed a SEANM application to show that, in presence of elastic demands, the new modeling framework allows to reduce power consumption when traffic conditions made typical SEANM approaches ineffective.
vi) Finally, we considered metro and backbone networks operated with Carrier Grade Ethernet and proposed a novel SEANM method to balance energy consumption and network congestion.

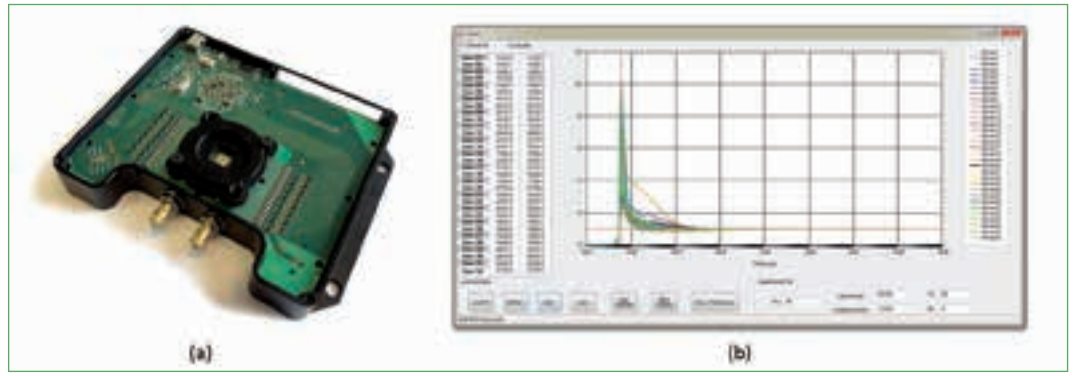
DEVELOPMENT OF MULTI-CHANNEL FPGA BASED CORRELATORS FOR PARALLEL FLUORESCENCE CORRELATION SPECTROSCOPY

Sixia Gong - Advisor: **Massimo Ghioni**

Fluorescence correlation spectroscopy (FCS) is a well-established technique to study binding interactions or the diffusion of fluorescently labeled biomolecules in-vitro and in-vivo. FCS is commonly implemented by using a confocal microscope to detect the fluctuations of fluorescence intensity arising from changes in the number of molecules diffusing through a small (~femtoliter) observation volume. The autocorrelation function (ACF) of fluorescence intensity fluctuations can yield physical as well as photochemical information (molecule size and concentration, blinking or binding/unbinding rates) about the diffusing molecules. To monitor the fluorescence fluctuations, FCS measurements need to be performed at nanomolar concentrations with typical acquisition times on the order of a few seconds to several minutes. However, faster acquisitions of FCS data are desirable in two cases: in high-content screening approaches, many molecules on reaction at different locations require simultaneous interrogation; also when observing fast evolving dynamic systems, diffusion parameters change as a function of time. It is thus important to develop FCS methods that enable simultaneous measurements at different locations within a living cell. Parallel FCS acquisition is then developed, with the help of multi-pixel detectors and multi-spot excitation generation technique, which maps each excitation spot onto every target pixel of the detector. Furthermore, simultaneous data acquisition and processing is needed, demanding for multi-input high efficiency correlators. Thereby this thesis describes design of multi-channel correlators for high throughput FCS experiments.

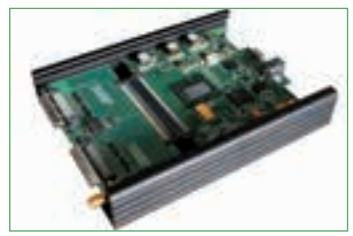
Various correlators have been developed, some are commercially available. The hardware correlators have been traditionally employed for real-time calculation of correlation functions over a certain dynamic range. Since in typical FCS measurements the correlation function spans over several decades of lag times, linear channel spacing is impractical and the multi-tau algorithm is used. Real-time, multi-tau correlators were formerly implemented on custom high-speed digital signal-processing hardware, either application-specific integrated circuits (ASICs) or, lately, field-programmable gate arrays (FPGAs). Nowadays, FPGAs can be manufactured in 28 nm CMOS processes and have reached integration densities that allow cost efficient implementation of even complex and resource intensive DSP algorithms. They offer lower non-recurring engineering costs and faster time to market than more customized approaches such as full-custom VLSI or ASIC design. On the other hand, software correlators are also available; they have fast design cycle, flexible structure and can provide offline operations. However for online computations, they could lack efficiency due to the high CPU utilization rate which prevents their application in high-throughput FCS experiments. As a result, FPGA based correlators providing multiple channels, high dynamic lag time range and online operations, are very promising for highly paralleled FCS experiments.

A single-channel FPGA based correlator was firstly designed employing the multi-tau algorithm. It features a maximum lag time of 150 ms while minimum time bin being 10 ns. This correlator is adopted to characterize afterpulsing effect of Single Photon Avalanche Diodes (SPADs), whose results are verified by a commercial correlator. So as to meet state-of-art standards, lag time range of the FPGA based correlator was then extended upwards to 80 s with minimum time bin being 5 ns. This long lag time correlator is divided into two parts to maintain real time display. Apart from the FPGA based correlators, a PC based software correlator exploiting an improved multi-tau scheme was designed for offline analysis. The signal trace for the software correlator is recorded in *photon mode* which counts the time interval between two pulses. Combining the FPGA based correlator for online correlation computation, with *photon mode* recording of the signal traces in FPGA, and PC interface integrated with software correlator for intercommunication and offline analysis, a complete single-channel correlator was developed.



1. (a) Photon detection module integrated with FPGA based correlator; (b) PC interface for intercommunication with the module and real time display of correlograms.

Based on the successful design of single channel correlators, a 32-channel correlator is then implemented which is intended to be directly contained in a photon detection module mounted with a 32 × 1 SPAD array. The module has a same FPGA as the one used for the single-channel design. In order to fully utilize the limited FPGA resources, the replication scheme for the 32-channel correlator is redesigned. The photon detection module together with the multi-channel correlator as a compact module (see Fig. 1) can be applied in FCS experiments providing direct signal detection and analysis path. However, some inherent features of SPAD arrays, namely afterpulsing and optical crosstalk effects, may introduce distortions in the measurement of correlation functions. These limitations are investigated to assess their impact on the module and evaluate possible workarounds.
In order to further upgrade the 32-channel correlator and enable correlation computation between modules, a standalone cross-correlator module has been developed. It employs larger FPGA, faster data transfer interface, able to hold 64-channel FPGA based cross-correlator with maximum lag time around 1 min. The module can connect with two external photon detection modules, receiving up to 128 channels of photon counting signals. The correlator system can execute instant signal-trace recording and fast online FPGA based correlation computation with real time display of all the correlograms. The signal track from each SPAD pixel can be stored in PC for offline analysis. According both to literature and commercial products, few of them are able to accommodate high number of input channels while maintaining large lag time range with low minimum time bin. Thus this correlator module would be of great interest for high-throughput FCS experiments.



2. 64-channel cross-correlator module

DEVICES AND TECHNOLOGIES FOR LARGE SCALE INTEGRATED PHOTONICS CONTROL

Stefano Grillanda - Supervisor: **Andrea Melloni**

Thanks to its ability in the generation, manipulation, and detection of light on-chip integrated photonics has been imposing as an enabling technology in a broad variety of fields of application, such as telecom, optical interconnection, bio-sensing, and quantum photonics. However, it is widely believed that the applications that will really benefit from the exploitation of integrated photonic technologies are those that require the aggregation of many components into complex systems-on-chip for the delivery of advanced functionalities. Indeed, much technological effort has been dedicated to scaling device dimensions down to the ultimate physical limit, so that now photonic platforms like silicon are mature enough to squeeze thousands of components into small chips. Yet, the realization of large-scale integration circuits performing complex tasks is still a challenge, and the result is that the perspective is still on the device rather than on the system.
In photonics, as in the case of electronics, device miniaturization is not a direct synonymous of device integration. In fact, integrated circuits cannot function properly without adequate tools to dynamically steer and hold each embedded device to the desired working point, counteracting functional drifts due to fluctuations in the environment, aging effects, mutual crosstalk, and fault events. Indeed, as the scale of integration increases the aforementioned effects become critical, and consequently monitoring, control and stabilization of components is mandatory. At the same time, while aggregating several devices, the power consumption required by tuning and control operations should not increase as well.
This thesis aims to develop devices and technologies to fill the existing gaps that prevent photonic integrated circuits to move from a single-device level to a new system-on-chip paradigm. In this work the first non-invasive light detector was demonstrated in silicon photonics waveguides. This transparent light monitor, that is the ContactLess Integrated Photonic Probe (CLIPP), exploits natural absorption effects of silicon waveguides to monitor the status of a circuit without affecting its operation and without wasting any additional photon with respect to those naturally lost by the waveguide. Thanks to its non-invasive nature and inherent CMOS compatibility, many CLIPPs can be placed in a photonic circuit, thus enabling multipoint monitoring of complex devices and circuits aggregating several components.
Furthermore, low-power transparent actuators, based on the integration of photosensitive chalcogenide glasses with the silicon platform, were realized to provide post-fabrication permanent trimming functionalities, both for compensation of fabrication tolerances and circuit reconfiguration. Athermal and trimmable silicon waveguides were developed, in order to enable simultaneously passive thermal stabilization and post-fabrication trimming of silicon circuits. Also, resiliency of these devices to high-power induced thermal effects is shown.
Active stabilization and feedback control of thermally actuated silicon resonators was demonstrated by exploiting an error signal provided by the CLIPP monitor integrated inside the microrings. Advanced control functionalities such as wavelength tuning, locking and swapping were demonstrated.
Finally, advanced functionalities and concepts, such as transfer function recovery, reconfigurability, adaptability were addressed in high-order filters and delay lines composed of several coupled ring resonators, and by even more complex integrated circuits, such as variable symbol-rate differential phase-shift keying receivers.

1. Silicon photonics chip hosting several photonic devices and circuits whose operation is non-invasively monitored and controlled by means of CLIPP monitors and CMOS electronics.

CALIBRATING A VARYING CAMERA FROM SILHOUETTES AND BACKGROUND

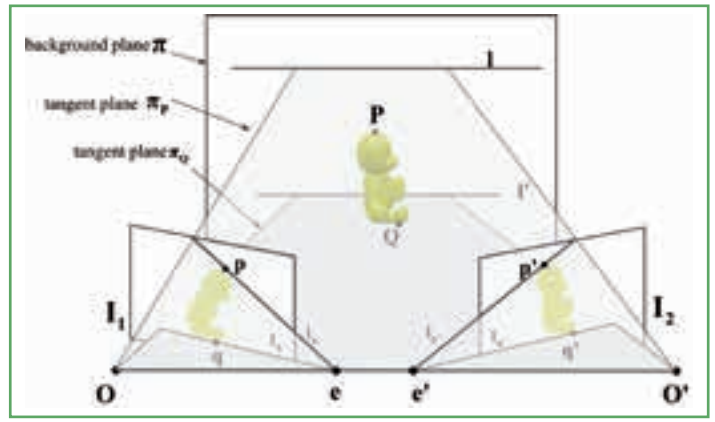
Dong Han - Supervisor: **Vincenzo Caglioti**

The visual modeling of a smooth textureless object is an interesting Computer Vision research problem whose applications rang from prosthetics to custom cloth, shoe manufacturing and film industry. The smoothness of the object makes it difficult to identify image correspondences, due to the continuously varying contour generator as the camera viewpoint moves. In addition, using consumer cameras (e.g., those provided with tablets or smart phones), often the auto-focus property of the device is active letting the intrinsic camera parameters vary between acquisitions. Therefore, traditional camera calibration methods cannot be applied.
In this thesis, we consider the scenario (Fig. 1) where a smooth 3D object of unknown shape is observed, against a background plane with unknown patterns, by an uncalibrated, moving camera with varying intrinsic parameters. Starting from the acquired images, we address the problem of recovering i) the value of the camera intrinsic parameters for each image, ii) the camera motion and, iii) the 3D reconstruction of some points on the object surface.

We propose a new framework that could calibrate a fully varying camera from background and silhouettes of smooth objects, not requiring more than two frontier points per view-pair. Although it is possible to calibrate a camera from only plane or silhouette, these methods suffers from limited usable information and thus their calibration are restricted by camera motion, camera type or the object shape: a plane can provide up to 2 constraints on the camera parameters, therefore only-plane based calibration methods cannot deal with more than 1 varying parameter; only-silhouette based calibration requires, in each view pair, at least 7 frontier points to exist, which is unrealistic for ordinary surfaces. Our algorithm is based on both silhouette and plane and it can be applied to calibrate a fully varying camera from ordinary surfaces without restricting either the camera motion or type.

1. camera calibration scenario

The proposed algorithm is divided into 3 main steps.
The epipole positions in each image pair are first estimated by finding two epipolar lines that are tangent to the object silhouette with the help of the plane induced homography.
Consider a smooth object being observed against a background plane by two cameras(Fig.2). The apparent contour (edge of the silhouette) of the object is the projection of the contour generator, which is a 3D curve consisted of points whose tangent plane goes through the camera center. The frontier point is the intersection between the two contour generators, where the tangent plane passes through both camera centers (the tangent plane is an epipolar plane). The tangent plane intersects background plane at a line and the trace of this line on the image planes of the two cameras are a pair of corresponding epipolar tangents, the two of which are related by the plane induced homography. If we find two frontiers and respectively their associated two pairs of corresponding epipolar tangents, the epipoles are determined as the intersections of the two pairs. The process of searching the frontier points is formulated as an optimization problem where we minimize the distance between the tangent line in the second image mapped by the plane induced homography and the convex hull of the apparent contour. Two global minima of the objective function correspond to the two outermost frontier points.



2. Geometric solution for recovering the epipolar geometry

The projective reconstruction of the image sequence is then robustly computed from the estimated epipolar geometry.
A view triangle is the graph representation of the geometries in a camera triplet where the nodes represent the cameras and the links between them are the fundamental matrices. A minimal representation of the camera matrices in a view triangle is used, where only inter-image homographies and the epipoles are involved. The multiple view projective reconstruction is built up incrementally using view triangles as the building block. Each time a new view is to be added to the set of calibrated views, an optimal view triangle is built and solved that has the smallest Sampson error. The calibrated set is then updated by adding the new node. Projective bundle adjustment is also performed to prevent error accumulation. The above process is repeated until all the views are added to the reconstruction.

Finally, a flexible self-calibration algorithm using the absolute dual quadric is employed to determine the camera intrinsic parameters and to upgrade the projective reconstruction to metric. Unit aspect ratio and zero skew are used as constraints in self-calibration and no other constraints are adopted. Euclidean bundle adjustment is performed at the end to improve stability and calibration accuracy.

The proposed algorithm is tested using both synthetic and real images. For the synthetic images, a scene is generated as shown in the figure, including a color foreground object, a moving camera and a background plane with gray-level unknown texture. The moving camera with varying intrinsic parameters is simulated using a set of spatially separated cameras distributed around the foreground object. To estimate the plane induced homography, we randomly generate a sufficient number of 3D points on the background plane and their images are corrupted with various level of Gaussian noise. The average relative error on focal length and principal point with respect to standard deviation (std. dev.) of Gaussian noise are obtained as the proof for the accuracy and robustness of the proposed algorithm. We show good calibration accuracy has been achieved with both relative errors are about 5% when the std. dev. of Gaussian noise is around 1 pixel.
For real image experiments, the images are taken with a mobile phone camera with autofocus effect. Two categories of experiments are carried out: 1. the intrinsic parameters of the camera are varying due to auto-focus effect; 2. the focal length is manually adjusted by zooming in and out. For both cases, good calibration accuracies are obtained in terms of re-projection error.

THE EMERGENCE OF DIVERSITY IN THE ADAPTIVE DYNAMICS FRAMEWORK: THEORY AND APPLICATIONS

Pietro Landi - Supervisor: **Fabio Dercole**

Innovation and competition processes are often identifiable in science. They are responsible for evolutionary dynamics driven by innovative changes in the characteristics of individual agents and by competitive interactions that promote better performing ones. Genetic mutations and natural selection play this role in biology, but the potential applicability of the evolution paradigm can be extended to social, economic, information sciences and engineering. Quantitative approaches to evolutionary dynamics were born from genetics and economic game theory. While biologists traditionally consider evolutionary change separated from the demography of the interacting populations, game theorists study the relative diffusion of a given set of alternative strategies and the robustness of the corresponding equilibria with respect to invasion from potential dissident. By contrast, the more recent approach of Adaptive Dynamics (AD) takes explicitly into account both the evolutionary and the demographic change and characterizes both the evolutionary equilibria and transients and non-stationary regimes. AD represents a flexible framework, based on the hypothesis of rare and small mutations, for the formal description of evolution of the characteristics of the system in terms of ordinary differential equations. Diversity increases in the system each time competition between innovative and resident strategies gives rise to their coexistence (evolutionary branching), and reduces when evolution brings groups of agents to extinction. Evolutionary branching is particularly interesting: in appropriate conditions, innovative agents can coexist with resident ones and their strategies, initially very similar, can then diverge generating two resident forms with different characteristics. The evolution of this enlarged system can still bring to the situation in which evolutionary branching is possible for one or both forms of agents present in the system. Thus, this succession of evolutionary branchings brings simple systems (with few resident forms) toward more complex and diversified configurations. The study of the possible branching scenarios is then very interesting in biology (where it gives an interpretation of the diversification of species from a common ancestor), but also in social sciences, economics, technology, engineering, etcetera. Moreover, some theoretical aspects of branching are still unstudied. For example, mathematical conditions under which branching occurs are expressed as sign conditions on appropriate second derivatives of the competition model, but theoretical results in critical cases in which such derivatives annihilate are not yet available. Although mathematically non generic, these situations are quite common in applications, in which particular symmetries of the model bring some derivatives to annihilate systematically. In conclusion, the main goal of the thesis is to focus both on the analysis of theoretical aspects of evolutionary branching in the framework of AD and on the development of models to interpret diversification phenomena in the above mentioned fields of science. The thesis is organized as follows. Chapter 1 is an introduction on the theory of evolution, starting from its history, passing through its basic elements (mutation and selection), and closing with the mathematical approaches to the study of the evolutionary dynamics. The concept of evolutionary diversification and evolutionary extinction are also intuitively introduced. Chapter 2 is dedicated to the Adaptive Dynamics approach, the resident-mutant competition model, the computation of the invasion fitness, and the AD canonical equation, that models the expected long-term evolution of the phenotypic traits of the coevolving community. In chapter 3 we focus on the emergence of diversity in the AD framework, that is, evolutionary branching. We classify the evolutionary equilibria with respect to their convergence and evolutionary stability, recovering the classical branching conditions, i.e., the two mathematical conditions in terms of second derivatives of the invasion fitness under which the system becomes dimorphic and experience disruptive selection, thus increasing its diversity. Chapter 4 is devoted to the study of the branching bifurcation, namely, the transition from evolutionary stability to evolutionary instability along with the change in a model parameter. This bifurcation occurs when the branching condition ruling evolutionary stability changes sign. To study such critical case, a particular third order approximation of the invasion fitness must be computed, and a novel property of the resident-mutant competition model must be exploited in order to obtain simple and general results. The case in which the other branching condition is critical is more complicated and is left for future research, but our theoretical approach is general and remains valid, also for further degenerate cases (e.g., when both the branching conditions are critical). In chapter 5 we develop a general methodology to study the evolution of biodiversity in eco-evolutionary two-species communities, with an application to prey-predator interactions. We then use such methodology in two fields of science different from biology. In chapter 6 we analyze the possibility that the interplay of natural and artificial selection due to fishing could lead to disruptive selection on exploited fish stocks. Finally, in chapter 7 we study the evolution of fashion purely driven by social interactions, with particular focus on the emergence of style diversity, and find out that different styles can successively emerge starting from a single style society. Chapter 8 discusses and summarizes the achievements of the work and close up the thesis with suggestions on extensions and future research.

MODEL PREDICTIVE CONTROL OF ENERGY EFFICIENT BUILDINGS IN SMART MICROGRIDS

Giancarlo Mantovani - Advisor: **Prof. Luca Ferrarini**

Summary of the work

This dissertation addresses the problem of thermal energy control in the context of large-sized commercial buildings characterized by wide open spaces and containing several thermal zones – i.e. shopping centers and malls, convention and congress centers, theatres, airports, train stations, large office buildings and so on. With respect to the residential sector, buildings in this category present specific features that make them particularly suitable for the achievement of energy reduction through the implementation of active efficiency strategies. In fact, they make an extensive use of Building Automation and Control Systems (BACS) technologies, whose potential is however largely unexploited in practice. Second, they have assets suitable for demand-response (namely comfort reduction consequent to a temporary energy price reduction), like thermal inertia, which decouples thermal and electrical consumption. Finally, adjacent thermal zones located in open spaces with high ceilings are widely coupled and often affected by a large amount of internal gains due to lighting systems, appliances and occupancy. This often leads to (1) vertical temperature stratification, which is one of the main sources of discomfort in large commercial buildings, and to (2) thermal zones overheating, especially during winter where the cooling power is limited. From a control-oriented perspective, this means that HVAC actuators are operated nearby the saturation boundaries, making the regulation problem very critical. In this context, this research work focuses on the application of control and optimization techniques voted at the reduction of both energy consumptions and vertical temperature stratification employing a model-based approach. Suitable grey-box class of models can be tuned using building physical and use information (white-box), while uncertain parameters are tuned from available measurement data (black-box). The structure obtained for the multizone model is detailed enough to catch all the relevant dynamics affecting energy and indoor comfort performances, and simple enough to be employed to tune temperature regulators. The modeling technique discussed is applied to an existing commercial building. Real data collected by the on-site BACS is used for tuning model parameters and validating the obtained model over a long period of time (winter season). The building multizone model is then used to design and both classic and advanced thermal energy controllers, whose aim is to achieve comfort improvements in term of vertical stratification reduction and energy saving. On the classic side, PI temperature decentralized control for each zone is designed. Then, the dissertation focuses on advanced building energy control based on Model Predictive Control (MPC). The implemented MPC regulator addresses notable steps forward with respect to the state-of-the-art, such as mathematical formalisms which allow to deal with the non-linear behavior of thermal actuators in common Heating, Ventilation and Air Conditioning plants (HVAC), to achieve a suitable degree of robustness for operating under real and uncertain conditions and to be able to shape and curtail the load according the specific context of operation. In addition, particular emphasis is put on model robustness with respect to disturbance inputs (such as internal gains, solar radiation, non-manipulable HVAC loads, occupancy), whose value in the future is hardly predictable. For this reason the building MPC is equipped with a Kalman filter which improves system robustness with respect to unknown disturbances and model errors.
The implementation of advanced control strategies for thermal energy control shows the potential to make a further step forward, which targets the integration of smart buildings in a more complex scheme that involves many other actors. For this reason, the second part of this work aims at studying the management and control of microgrids which include a large smart commercial building, generation from Renewable Energy Sources (RES) and a battery as electrical Energy Storage System (ESS). The microgrid system analysis, modeling, design and simulation is addressed in detail still paying attention to real world operating conditions, system non-linearities and the possibility to pursue different control objectives using the same control architecture. Specifically, the solution here proposed for the control of the smart microgrid above described is the adoption of a distributed control framework, where the previously presented building MPC operates the HVAC plants in a flexible way (to better follow RES production or to deal with transmission problems) and interacts with a second MPC which controls the ESS power flow, so as to minimize the overall operating costs, to maximize the exploitation of RES and to serve building needs in terms of thermal comfort, ensuring sustainability with minimum side effects. An intense insight in real system and components and also a significant experimental testing on a real microgrid have been carried out in order to evaluate the capabilities of this architecture in a reliable manner, pointing out advantages and limitations of an actual on-field implementation.

Main results obtained

1. Along with all relevant phenomena, the building thermal modeling framework presented takes into account the vertical temperature stratification among adjacent thermal zones. Model validity is enforced by tuning and validation phases carried out over real measurement data of an existing shopping center. The final results show an accuracy of about 0.5°C in simulating the zone temperatures and a monthly energy consumption error below 1%.
2. The thermal model is simple enough to be used for the design of both classic and advanced controllers for the control of vertical temperature stratification. Novelty is provided in the mathematical formulation of the MPC controller, which separates positive and negative values of thermal power in order to account for different operating costs of heating and cooling. Simulation results show that the MPC decreases monthly energy consumptions of 4.5%, without any change to the HVAC plants and a reduction of more than 60% in stratification with cooling actuators in place.
3. The developed controllers is able to deal with real world limitations of HVAC plants. Particular emphasis is put on strategies for the consistent compensation of actuators' non-linearities directly into the MPC, employing variable hard constraints specific of the current operating conditions and updated at each time step. The result is an increase of the heating and cooling capacity of HVAC actuators up to 25%, with respect to the case with fixed hard constraints.
4. The distributed energy control paradigm here proposed is an innovative strategy to deal with the control of smart microgrids composed by smart buildings, wind or PV RES and a battery as ESS. Two separated controllers interact in order to minimize the overall energy expenditure, taking into account different pricing schemas for buying and selling electrical energy, penalties due to load imbalance, battery non-linear efficiency, load limited flexibility and non-perfect renewable power prediction. The final results show that over a period of one month the presented architecture saves costs up to 63% with respect to the same scenario managed without an ESS.
5. The system is extensively tested both in simulation and on an appropriate experimental setting, controlling a microgrid with a real vanadium battery as energy storage system, an existing 10kW wind turbine and a resistive load to simulate the building consumption (Syslab microgrid at RISØ DTU, Denmark). The control architecture proves to be flexible enough to prosecute very different tasks, while keeping a standard and scalable mathematical formalism.

TECHNIQUES FOR HIGH-EFFICIENCY DIGITAL FREQUENCY SYNTHESIS

Giovanni Marucci - Advisor: **Prof. Carlo Samori**

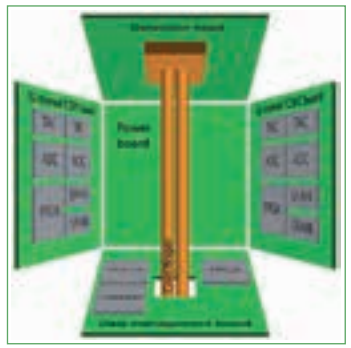
The evolution of modern wireless standards poses stringent noise specifications to the design of frequency synthesizers for high-data-rate communication systems, limiting tolerable jitter and spurs level. Moreover, such performance must be provided at low power consumption and small area, in order to meet the requirements for mobile applications and battery powered systems at low cost and high integration level. Therefore, the key challenge in frequency generation is the design of high-efficiency synthesizers (i.e. with low jitter at low power), and it is generally tackled relying on fractional-N phase-locked loops (PLLs) with LC oscillators, due to their better noise/power compromise with respect to ring oscillators. However, the analog content of LC-based PLLs do not benefit from technology scaling and prevents their design from easily fitting into a typical digital design flow. In this framework, we present a high-efficiency frequency synthesizer based on a fractional-N digital PLL with a ring oscillator. Minimum jitter is obtained developing a comprehensive analysis of noise in digital PLLs, specifically focusing on architectures with single-bit phase detector, which are the most promising in terms of efficiency. However, the coarse quantization of phase error makes these systems prone to the generation of limit cycles appearing as unwanted spurs in the spectrum. The random noise contributed by building blocks can eliminate those spurs acting as dithering signal, without the addition of extra noise. A closed-form expression of the total output jitter as a function of loop parameters and noise sources is developed, which suggests a design strategy for noise minimization. Yet the lowest achievable value for the jitter ultimately depends on the noise of the oscillator, which is considerably high in the case for ring oscillators. An effective way to significantly reduce it, without increasing appreciably power consumption, is to rely on the concept of injection locking, but unfortunately its application has been so far bounded to integer-N synthesis, preventing the introduction of inductorless frequency synthesizers into standardized wireless systems. With this aim, we propose a technique to enable fine fractional-N resolution in injection-locked PLLs that allowed us to reach the best power/noise trade-off among published fractional-N digital frequency synthesizers without integrated inductors.

DEVELOPMENT OF HIGH PERFORMANCE MULTICHANNEL SYSTEMS FOR TIME-CORRELATED SINGLE-PHOTON COUNTING APPLICATIONS

Luca Miari - Supervisor: **Ivan Rech**

In recent years, the interest of many research fields in non-invasive optical analysis has rapidly grown. In particular, studies performed in biology and chemistry have found increasing benefits from the development of systems able to perform single-photon measurements, since they push the sensitivity of the analysis to ultra-low intensity levels and to ultra-fast evolving signals. These sensitivities are achieved thanks to the employment of single-photon detectors and timing acquisition chains. The Time-Correlated Single-Photon Counting is one of the leading techniques on which many others rely on: Fluorescence Lifetime Imaging Microscopy (FLIM), Förster Resonance Energy Transfer (FRET), and Fluorescence Correlation Spectroscopy (FCS) are just a non-exhaustive list of them.
The TCSPC technique consists in the repetitive laser stimulation of a sample, recording each time the delay between the stimulating laser pulse and the pulse coming from the photon emitted by the sample and captured by the detector. After several photons it is possible to build a histogram that represents the original waveform of the light signal, since the probability distribution of the photon delays corresponds to the intensity of the light signal.

The necessity of a repetitive measurement inherently asks for a high count rate instrument: the higher, the shorter the total measurement time. Besides, other specifications concern the linearity, which is a fundamental feature to define the shape of the signal without distortions, and the time resolution, that characterizes the ability of the system to resolve very fast signals. However, applications are pushing the instruments to expand their number of channels, since this will reduce the total measurement time, opening also the way to many new kind of analysis. Time-resolved imaging and simultaneous spectrally resolved analysis are just examples of these new possibilties.
Regretfully, a strong trade-off exists between performance and number of channels in TCSPC systems: applications are forced to choose between instruments with many channels but poor performance or single channel instruments with state-of-the-art performance. Aim of the researchers is the development of high performance multichannel TCSPC instruments that will break this trade-off.
In particular, this thesis work is devoted to find solutions to the system part of the instrument: the management of the incoming data, their on-board storage, and their transfer towards a remote computer. Initially, the work has focused on the application of a 48-channel Detection System to a setup design to perform single-molecule FRET analysis. The system, already designed, was featuring a 12×4 matrix of Single Photon Avalanche Diodes (SPADs) and a Field Programmable Gate Array (FPGA) as its logic core. Its firmware, instead, has been designed to perform a time-stamping measurement on the incoming photons, in order to provide to the software raw data to compute the FRET efficiency.
A 100MHz clock has been used to stamp the incoming data, and forty-eight parallel First-In First-Out memories (FIFO) have been implemented as buffers between incoming data and downloaded ones. By optimizing the transfer process, a download rate of 20MBps has been achieved, which means an average count rate of 100kcps on every SPAD, enough to perform the analysis.
This work has been useful to evaluate on the field the necessity of multichannel instruments, thus a 1024-channel high performance system has been ideated in the research group. The increase of the number of channels carries many problems: the detector matrix and the acquisition logic are just some of them, as well as the dimensions of the system and its power consumption. Nevertheless, the technological limit today present is the data rate that can be managed on-board and transferred towards the computer. Fast protocols available for compact instruments do not provide a transfer rate far greater than 10Gbps, which means that the number of acquisition channels that can be properly handled is a few tens. A 64-channel acquisition system has been chosen as the target; a router will multiplex all the detection channels on these acquisition chains. To further reduce on-board complexity, but also dimensions and power dissipation, two twin 32-channel Acquisition Boards have been designed. They will exploit the TAC-ADC structure (acronym for Time-to-Amplitude Converter and Analog-to-Digital Converter) since it is the one that provides the best performance, especially in terms of linearity. The FPGA that receives the ADC data is in charge for creating the histograms, but an external on-board memory is necessary to store all of them. Provided the complexity of the system and its novelty in the research group, a Demoboard has been designed to test the performance of the full structure.
Once the data have to be sent out from the system, a section that gathers the two streams and implements a fast communication channel towards the computer is necessary.
After the study of the different protocols available on the market, two of them have been chosen to be implemented, in particular the SuperSpeed USB 3.0 and the 10-Gigabit Ethernet over optical fiber. They have been successfully tested on evaluation boards and then implemented on a specific Data Management Board, that mounts another FPGA to handle the logic operations.
To validate the performance of these new kind of systems, specific multichannel instruments are required. A multichannel pulse generator is necessary to provide the input signal to all the TCSPC chains, but none of the today commercially available one satisfies all the specifications. Thus, an 8-channel Pulse Generator has been designed as a test instrument for these systems. A 2-channel Module is the base block of the generator: after choosing the input signal between an on-board reference and an external trigger, a delayer loop delays the reference edge and a fast output transistor stage generates the output. This structure reaches a time resolution of 6ps when no delay is added, worsening to 20ps for a 1µs delay. Both results are remarkable and this validates the employed architecture, as well as the whole instrument.



1. Rendering of the 1024-channel system, final purpose of the work where this project has been inserted

STUDY OF A DAY-0 FRONT-END SOLUTION FOR THE DSSC DETECTOR OF THE EUROPEAN X-RAY FREE ELECTRON LASER

Bayan Nasri - Supervisor: **Prof. Carlo E. Fiorini -** Tutor: **Prof. Angelo Geraci**

In this Doctoral dissertation, the design, realization and testing of a new front-end for the DSSC project has been discussed. The Day-0 solution as an alternative approach to DEPFET sensor scheme has been demonstrated through both analytical evaluation and experimental characterization. The proposed front-end takes advantage of the simple implementation inside the same ASIC chip and produce the compression behavior by using a resistor (fixed value in the first approach and variable one in the improved version) in series with the transistor. In this thesis, new Day-0 solution was discussed in a way to keep the same filter as it is in DEPFET sensor scheme.

For the European XFEL, there are several imagers under development for the X-ray photon detections, including single-point, 1-D and 2-D detectors. As one of the 2-D imagers, the DSSC project aims to provide wide energy coverage with single photon detection capability thanks to its very low noise. To achieve this result a novel detector structure with a compressive characteristic based on the DEPFET concept is under development by the DSSC consortium. The high gain for small collected charge and the compression for large signals will provide both desired features of single photon detection capability and wide dynamic range.

In spite of its capability in high dynamic range and low noise performance, the manufacturing of the DEPFET sensor needs a sophisticated processing technology which requires relatively long manufacturing time. According to the mentioned complexity, it was evaluated to have an alternative sensor matrix, with its corresponding front-end, as Day-0 solution. A Day-0 solution is here intended as a solution characterized not by the best performance of the DEPFET, but available in a shorter time to allow first beam tests and experiments. The alternative sensor is made of mini Silicon Drift Detector (mini-SDD) and the compression behavior is obtained from the front-end on the readout ASIC and not by the silicon sensor, as in the DEPFET.

In the Day-0 solution, the DEPFET is removed and replaced still by a PMOSFET transistor, but now belonging to the ASIC chip and realized with the same IBM technology of the following stages of the ASIC. The PMOSFET is designed in a way to provide low noise and a compression characteristic close to the one of the DEPFET-based solution. In fact, this alternative front-end input stage should still provide a non-linear amplification of the detected signal. Accordingly, the replacement of the DEPFET with this stage should keep as much as possible equal the following architecture adopted for the processing electronics.

The proof-of-principle of the Day-0 solution has been verified in a first prototype. Using the prototype with a Silicon Drift Detector (SDD), X-ray measurements have been carried out to assess the electronics noise performances which are in line with the theoretical expectations. After the confirmation of the performance, a conservative design of the Day-0 solution has been implemented in the F1 chip as the first full size matrix chip. The F1 chip comprises two different front-ends suitable for two different sensor types: it can be mated to either a DEPFET sensor matrix which itself compresses the input signal or to an SDD sensor as the Day-0 solution.

An improved version of the Day-0 front-end was investigated to improve the dynamic range of the front-end and its noise performance. In this design, the nonlinear transconductance of the amplifying PMOS has been enhanced by statically switchable parallel devices and a non-linear capacitor has been added. Moreover, the resistor in the first version of the Day-0 front-end is replaced by a NMOSFET as a variable resistor. In the improved version of the Day-0 front-end, more than 2800 photons at 1keV can be allocated within the dynamic range of the ADC based on the compression behavior of the front-end.

In conclusion, the Day-0 solution as an alternative approach to DEPFET sensor in the DSSC project has been demonstrated through both analytical evaluation and experimental characterization. Although the DEPFET approach has a very good noise performance, but in the improved version of the Day-0 front-end, the noise performance has been improved as close as possible to the DEPFET approach. In the frame rate of 4.5MHz, the total noise of the DEPFET approach is about 42 el r.m.s while it is about 66 el r.m.s in the Day-0 solution, and in the 0.9MHz frame rate the electronics noise is about 12 el r.m.s for the DEPFET approach and 23 el r.m.s for the Day-0 solution. The dynamic range of the improved version of the Day-0 front-end can be matched with the dynamic range of the ADC in 1keV photons energy as the target photon energy one. Also in other photon energy ranges, the dynamic range of the Day-0 front-end can be tuned due to its flexibility of compression behavior introduced by extra parallel branches.

ANALYSIS AND DESIGN OF ENERGY-ORIENTED DRIVING ASSISTANCE SYSTEMS

Carlo Ongini - Supervisor: **Prof. Sergio M. Savaresi**

The growth of oil price, the congestion of urban areas and the increased sensibility for environmental problems brought scientists and companies to invest significant resources on improving the energy efficiency of the transport sector. Nowadays, one of the major environmental problems is air pollution deriving from the transport sector, and road transport alone is expected to be the largest contributor to anthropogenic climate forcing in 2020.
The devising of more efficient vehicles (e.g., improved control systems, aerodynamics, powertrain and engine design), the usage of alternative energy sources and fuels (e.g., electric and hybrid vehicles), and the deployment of Intelligent Transportation Systems (ITS) applications are solutions for the decarbonization of the sector. Intelligent technologies are therefore playing an increasingly important role in the drive for green innovation. Especially, the development of energy-oriented Driving Assistance Systems opens the perspective towards a new model of ITS, which can be referred as *Green ITS*, and aims at conceiving applications and services specifically designed to reduce energy consumption and polluting emissions.
Vehicle's energy consumption depends upon different influential factors, as vehicle characteristics (e.g., weight, engine power, aerodynamic drag), network condition (e.g., traffic flow, congestion), road characteristics (e.g., slope and topology), external conditions (e.g., weather and temperature) and driver behavior which impacts on vehicle dynamics (e.g., speed, acceleration, gear choice). In particular, the driving style has a huge impact on vehicle's energy consumption. Various research results showed that improvements in the driving style provide direct savings from 5% up to 40% of the total energy expenses as well as reductions in air pollution.
Within this interesting and evolving context, this Thesis proposes a couple of green ITS applications with the aim of proposing innovative solutions tailored to fuel-consumption optimization through the improvement of the driver behavior. In particular, the main objective is to design energy-oriented, vehicle-independent and passive systems aimed at influencing the driver behavior and promoting a fuel-efficient driving.
First, we present a system able to assess in real-time the driving style. The system is fully integrated in a smartphone application that acquires the signals related to the vehicle dynamics (e.g., velocity and acceleration) and computes three power-related indexes. The smartphone application provides feedback to the driver in order to induce a change in her driving style, which in turns should enable the energy savings. The interaction between driver and system is achieved through different Human-Machine-Interfaces (HMIs). Differently from existing studies, we design a vehicle-independent system that requires just few parameters without any connection to the vehicle CAN-bus or OBD-interface.
Furthermore, the driving style application has been integrated in a real scenario. We developed a vehicle-to-infrastructure (V2I) architecture as a part of the Green Move project, an innovative Electric-Vehicle Sharing System. Thanks to this platform, the driving style application can be dynamically loaded and unloaded on vehicles of the vehicle-sharing service. The application computes a quantitative estimation of the driving style of a vehicle-sharing user, communicates this information to a control center, and displays it to the driver. Thus, feedback is provided to the user to nudge her to adopt a more economical driving style, whereas the control center can use this information to make appropriate enhancements to its service or to provide additional features.
Finally, in this Thesis we propose a method and the related road infrastructure aiming at reducing the energy consumption of vehicles driving along roads controlled by traffic lights. An algorithm has been conceived for computing energy-optimal speed profiles to drive along a road taking into account a set of constraints (e.g., speed limits, traffic light rules). A road-marking infrastructure continuously suggests the optimal velocity the driver should adopt in order to minimize fuel consumption.

Main Contributions

The main contributions of this Thesis can be summarized as follows:

- A novel method for the energy-oriented, quantitative estimation of the driving style is presented. An important innovative contribution is given by the definition of three power-related indexes for evaluating the instantaneous driver performances. To the best of author's knowledge, existing systems foremost focus on the qualitative classification of the driver behavior, usually limited to a finite set of discrete labels. Moreover, they are vehicle-dependent, in that they were developed for specific vehicle models. Thus, the approaches are hardly reusable and require dedicated electronic equipment. Differently from existing studies, we design a vehicle-independent system that requires just few parameters without any connection to the vehicle communication interfaces.
- An experimental validation of the driving-style systems is given. The effectiveness of the approach, and the savings enabled by the interaction with the driver are assessed with an experimental campaign carried out on urban and extra-urban routes by different drivers. Experimental results prove that the proposed driving style system reduce the vehicle consumption up to 30%. A questionnaire proposed to the volunteers evaluates the different user interfaces designed for the driving style application.
- The driving style system has been applied in a real scenario. We introduce a V2I/I2V system, in which the driving style application has been integrated. The V2I/I2V framework is the main component of the electric vehicle-sharing project named "Green Move", a three-years project co-financed by Regione Lombardia and developed in the city of Milan. Within the Green Move project, we developed a prototype platform that:
  - provides an electronic device and its hardware/software interface, which allows the system to interact with a heterogeneous fleet of electric vehicles in a uniform way;
  - relies on mobile devices to let users access and interact with the system (e.g. take possession/release a reserved vehicle, open/close its doors, enable/disable the drive);
  - offers an infrastructure to customize the software configuration of vehicles by pushing new services, thus realizing a platform which allows services and user-defined applications to be dynamically loaded and unloaded on vehicles.
- An energy-oriented roadside assistance system for signalized intersections is presented. We propose a novel algorithm that determines the energy-optimal speed to pass a road with signalized intersections. Based on the knowledge of Signal Phase and Timing (SPAT) and road information, an off-line velocity planning algorithm determines the optimal trajectory in terms of energy consumption. The method explicitly considers the consumed energy as figure of merit to be minimized and solves a specific optimal control problem that considers the traffic lights as constraints.

SUNDROPS: SEMANTIC AND DYNAMIC DATA IN A PERVASIVE SYSTEM

Emanuele Panigati - Supervisor: **Letizia Tanca**

Nowadays people are surrounded by a high quantity of data, coming in different and heterogeneus formats. However, humans cannot exploit the whole power of this data without appropriate aid of digital means. The *SuNDroPS system* (Semantic and dyNamic Data in a Pervasive System) aims at supporting (mobile) users with a context-aware approach, allowing them to consider only a small set of data, automatically selected by the system itself according to their current context and interests. Part of this data (e.g. user data, service data) is stored in traditional information systems – e.g. supported by RDBMSs – while a large part of them – dynamic data that come e.g. from sensors or system logs – need to be treated as data streams and dealt with in the appropriate way with while integrating them with the other data.
This problem requires research beyond the current state-of-the-art.
The SuNDroPS system stems from the Context-ADDICT (Context-Aware Data Design, Integration, Customization and Tailoring) system and adds new features to manage high loads of dynamic, sensor-coming data, seamlessly combining them with the more traditional information. Moreover, SuNDroPS includes the new data mining algorithm *MREClaT*, a Map-Reduce variant of the EClaT algorithm allowing the system to reduce the processing time required to perform the frequent itemset mining task. This feature adds a useful functionality allowing the extraction of previously unknown knowledge from the data flowing in the system.

Context-ADDICT Revisited

The main goal of *Context-ADDICT* and SuNDroPS is to create a middleware infrastructure to support the design and development of context-aware data-intensive applications. The focus is on mobile, possibly peer-to-peer applications, where the notion of context can be exploited to provide the user with a filtered view over the data, retrieving only the information relevant to the users in their current context. Context-ADDICT was conceived almost a decade ago, and was based on a methodology for context-aware system design where the context must be explicitly declared. With SuNDroPS, the "injection" of context is transparent to the user and to the application: the current user context can be inferred from sensor readings and the data corresponding to each context can be automatically assigned by mining historical data by means of the designed, efficient data mining algorithm.

SunDRoPS has several components, some of which inherited from the Context-ADDICT system: the core of the system is still built upon the core modules of Context-ADDICT; in addition, SuNDroPS interfaces two more components that allow to handle sensor data and WSNs (Wireless Sensors Networks): *TRex* and *PerLa*. These new modules enrich the set of datasources that can be managed by the system, allowing it to manage events streams, datastreams and WSNs. For all the data mining activities SunDRoPS provides the new data mining component, *MR-Miner*, that implements *MREClaT*. The output of MR-Miner is then used as input by the *Run-Time Context and Preference Manager*, allowing it to speed up the automatic derivation of context-aware data and preferences.
While the original Context-ADDICT platform deals only with static data stored in data sources, its new revisitation and extension has to manage dynamic data coming from pervasive sources (sensors and sensor networks, social networks), considering also that these data carry intrinsic temporal information that might be helpful in context management.
The main difference between Context-ADDICT and SunDRoPS lies in the sensor data handling: while the first is based on a static and predefined sets of context and each transition must be notified by the user herself (or by an application running on the user device), SunDRoPS instead can (at least partially) automatically determine which is the running context of the user and/or of the system, on the basis of the environmental gathered data.
The problem of seamlessy switch context is quite on the cutting edge of the technology nowadays. Lots of different services provides context-oriented components (primarily location-based context). The SunDRoPS system empowers this contextualization allowing to use more data other than the location ones, providing a framework that can be adopted in very different scenarios, both in indoor (e.g. museums) or outdoor (e.g. emergency situation management) applications.

PerLa and TRex

PerLa and TRex modules run both as standard query wrappers and as "environmental sensing" extensions of the system.
The system execution flows can be split into two main sub-flows:

- A pull-based flow in which the system answers the users' specific requests
- A push-based flow in which the system provides information to the user only evaluating the context, without having received any specific request or query.

In the first case, the system simply acts as the original Context-ADDICT system does – answering incoming requests –, while in the second case the system continously uses and analyzes the information flowing through the sensors and checks if it is necessary to send new information to the user.
All the modules operate using a push paradigm (there is no module pulling information from other ones, every module listens for new data or commands); this requires a bit more bandwith consumption but this ensures a more reactive behavior of the system.

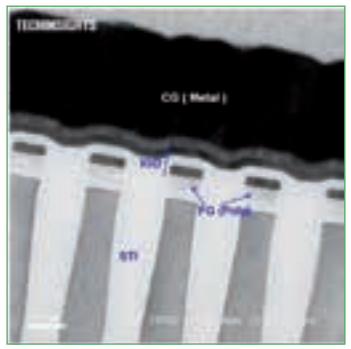
MR-Miner

Another problem is how to manage all the historical data about user preferences and past contexts in order to extract information to be used in the future for refining the information to be sent to the user. The *Run-Time Context and Preference Manager* module aims at performing this task.
The module operate an advanced personalization of the user current context choosing which are the information that best fit the user needs, analyzing the whole user transaction log. Since the log can be take very long time to be processed in order to speed up its analysis,only the most frequent operations that a user performs in a context in order keep track of hers habits are considered.
This frequent itemset mining task is eased by MR-miner that, applying the MREClaT algorithm that allows to analyze in parallel (adopting the MapReduce paradigm) the large amount of data flowing to the system, once stored in order to allow its historical analysis.

This work aims at building a real-life system based on the above technologies, plunging more deeply into some aspects which remain to be investigated like semantic data stream processing and cloud-based data mining algorithms. Prototypes of (parts of) the system have been produced, within the application domains of car-sharing services and support for citizens' mobility in the Green Move project for electrical car sharing,

IMPACT OF FEW-ELECTRON PHENOMENA ON THE OPERATION AND RELIABILITY OF NANOSCALE FLASH MEMORIES

Giovanni Paolucci - Supervisor: **Alessandro Sottocornola Spinelli**

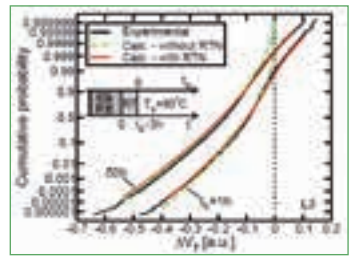


1. Cross section of a state-of-the-art NAND array (IMFT 20 nm technology) along the word-line direction.

The research work is focused on the emerging constraints to NAND Flash operation and reliability dictated by the granularity of electric charge in the gate stack. Progressive cell miniaturization entails that cell electrostatics and, in turn, cell threshold voltage (V_T), is ruled by an extremely reduced number of electrons, therefore pushing the Flash cell to work in a few-electron regime. This has two main consequences. First, the V_T displacement (ΔV_T) caused by the emission or the injection of a single charge in the gate stack (either the tunnel-oxide, the floating-gate or the interpoly dielectric) became so relevant that it can be clearly detected by using adequate experimental procedures. Second, the stochastic nature of the charging/discharging phenomena is way more evident than in the past: statistical fluctuations in the number of charges, in fact, play a significant role when such number reduces due to cell shrinking (Figure 1 provides an example of an extremely scaled Flash NAND technology). Also, when charges are trapped in the tunnel-oxide or in the IPD, additional quantities subject to statistical dispersion are the time constants of the the capture/release events and the ΔV_T following each of the events. The Ph.D. thesis extensively analyzes both of the previous points for two of the main phenomena ruling the behavior of modern Flash cells, *i.e.* charge detrapping and electron injection into the floating-gate via Fowler-Nordheim tunneling. In both cases the analysis started from a careful experimental characterization, which, thanks to the use of clever methodologies, allowed to determine for the first time the discrete, statistical nature of the phenomena. After that, starting from relatively simple assumptions on cell physics, new semi-analytic stochastic models were developed, based on a small number of free parameters. The aim of the models is to provide a useful tool to investigate how few-electron phenomena affect cell V_T and evaluate their role along the lifetime of a modern Flash device.

The impact of the research activity was recognized both by the semiconductor industry and the scientific community, as the main results achieved in this field were awarded at the IEEE International Physics Reliability Symposium (IRPS) both in 2013 and 2014.

The use of clever experimental procedures enabled the first direct observation of single-electron detrapping in NAND Flash arrays. Results revealed a significant statistical dispersion of the number of trapped charges, of the detrapping time and of the single-electron ΔV_T. Starting from these observations, the phenomenon was then re-examined from a discrete, statistical point of view, showing that in the most interesting case where the number of trapped electrons feeding the detrapping process is Poisson distributed among the cells, detrapping events are the result of a nonhomogeneous Poisson process and a simple and powerful formula allows the calculation of the full ΔV_T statistics without the need to resort to lengthy Monte Carlo simulations. Also, the important assumption was done that the detrapping time constant takes on a wide distribution on the logarithmic time axis, leading to the definition of an average spectral density of trapped electrons. The interaction of detrapping with random telegraph noise (RTN) was also highlighted, showing that the V_T instabilities resulting from the latter process cannot be ignored for a meaningful comparison with data, and that they can easily be included in the simulations.
The results, which can be applied also to different reliability phenomena involving charge trapping/detrapping in MOS devices, paved the way to the development of a comprehensive statistical model able to deal with V_T instabilities under whatever on-field usage of the memory array. To this aim, not only retention/bake conditions, but also the cycling conditions, *i.e.* the number of P/E cycles, together with the duration and the temperature of the idle periods in-between, were considered, exploiting the idea that the average spectral density of trapped electrons can be calculated according to the cycling pattern. The model can accurately reproduce the experimental data (an example is shown in Figure 2) and was used to discuss the accuracy of some testing schemes commonly adopted for the assessment of NAND device reliability.



2. Measured and calculated ΔV_T distributions of a cycled Flash NAND array during a high temperature bake emulating a data retention time stretch.

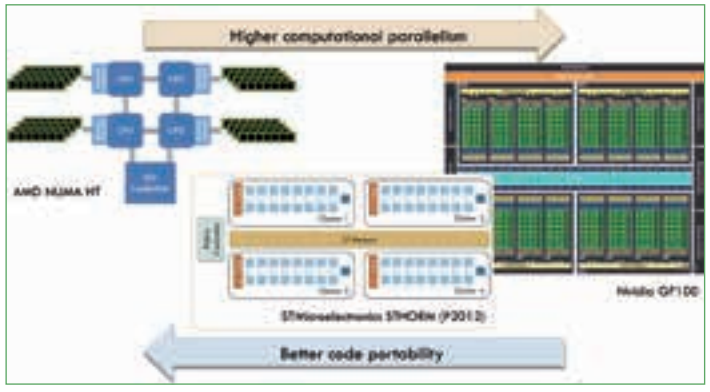
Single-electron charging of the floating-gate during Fowler-Nordheim programming of a mainstream Flash memory cell could be detected for the first time by using simple averaging techniques on a state-of-the-art NAND Flash array. The ΔV_T corresponding to an injection event is fixed and depends only on the capacitance of cell IPD, differently from the case of single-electron detrapping where the V_T shift is statistically distributed. Nevertheless, the injection process is of statistical nature, and the number of electrons injected during a programming pulse of amplitude V_S is approximately Poisson distributed for small values of V_S, while it it takes on a sub-Poissonian distribution at higher V_S, as shown by extensive experimental characterization. In either case, the spread associated to the number of injected electrons constitutes a relevant source of program noise (PN), setting the ultimate limit to the accuracy of the V_T placement of decananometer Flash cells.
In conclusion, the research activity provided the physical understanding, the modeling tools and the characterization techniques required to investigate the programming accuracy and the reliability of extremely scaled Flash memory technologies, highlighting the dominance of few-electron phenomena on cell operation. In this sense, the work provides a valid reference for further developing the Flash concept, pushing it towards the single-electron limit.

DESIGN SPACE EXPLORATION OF OPENCL APPLICATIONS ON HETEROGENEOUS PARALLEL PLATFORMS

Edoardo Paone - Supervisor: **Cristina Silvano**

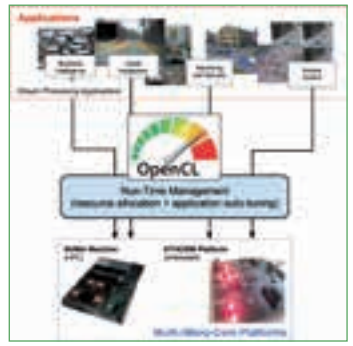
Multi-core processors are widely used in modern computing platforms, both in the embedded and High-Performance Computing (HPC) domains. Recently, also GPUs have been used for general purpose computing, to accelerate compute-intensive kernels of throughput-oriented applications. However, these platforms expose different programmability complexities: while general purpose multi-core CPUs provide good code portability on a small number of compute units, GPU architectures achieve much higher throughput but require more specialized application code (see Fig. 1). In between, there is a wide range of platforms with different trade-offs between programmability and computational parallelism, as well as GFLOPS/Watt efficiency.

This thesis targets the multi- and many-core platform domain, ranging from HPC to embedded systems. While these two worlds are still quite different in terms of computational power, recent years have witnessed a convergence of parallel architectures and programming paradigms. Today platform vendors are adopting OpenCL, a cross-platform API, to exploit the computational parallelism of modern accelerators while enabling functional portability of applications. For the application domain (see Fig. 2), we target stream processing applications, such as applications for smart cameras and augmented reality. The main contribution is about customization and optimization of OpenCL applications. It addresses one limitation of the OpenCL programming paradigm, namely the tight dependence of application performance on platform architectural details. Although the OpenCL API is cross-platform and generic enough to enable programming of different types of accelerators, customization of software parameters is necessary to achieve good performance when porting applications to the target platform. This problem is exacerbated by the intrinsic heterogeneity of modern computing platforms, which usually include several accelerators of different type.



1. Trade-off between programmability and computational parallelism for multi-core platforms.

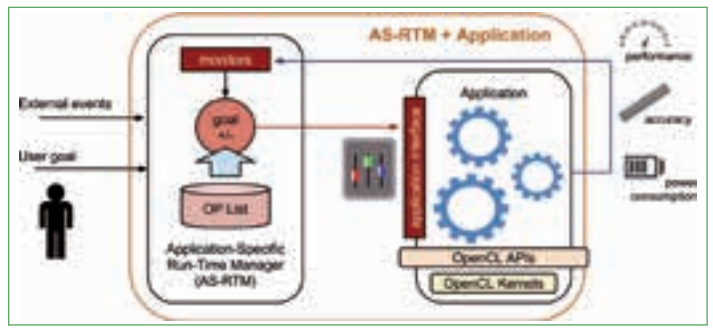
At this aim, an optimization methodology is proposed, based on customization of a parametric OpenCL application design. The methodology exploits Design Space Exploration (DSE) to identify the optimal solutions, with respect to multiple design objectives such as throughput or Quality of Service (QoS). The integration with the Multi-Objective System Tuner (MOST) framework allows to automate the DSE process and to implement advanced exploration strategies. On the one hand, the proposed techniques reduce the exploration time on simulation platforms while providing close-to-optimal solutions; on the other hand, they exploit platform-specific constraints to prune out unfeasible solutions from the design space.



2. Application and platform domain of the proposed methodology.

Another contribution of this thesis deals with resource sharing in multi-application scenarios. This type of parallelism – referred to as request-level parallelism – is enabled by the increasing number of cores integrated in the same chip. The proposed run-time management technique allows accounting for dynamic application requirements and workload variations, in order to optimize the overall average system performance. By applying techniques of software approximate computing, an application can be designed to expose tunable parameters that trade off the output quality with throughput. In our approach, DSE is used to identify the parameter configurations – called operating points – that provide the optimal trade-offs with respect to throughput, output quality and resource usage. In turn, this knowledge base allows implementing performance-aware scheduling and effective application auto-tuning, by reducing the decision space at run-time.

As shown in Fig. 3, each application is linked to a library that provides an Application-Specific Run-Time Manager (AS-RTM). The main purpose of the AS-RTM is to manage application adaptivity by monitoring the performance metrics and by tuning the application parameters at run-time. The AS-RTM is generic, but its behavior can be customized for each application by passing a different list of operating points. It also allows defining one or more application goals, which represent soft constraints on the performance metrics (e.g. the frame-rate or QoS). Experimental results show a better average performance with respect to a plain Linux configuration and, at the same time, a significant improvement of performance predictability.



3. Application adaptivity through the Application-Specific Run-Time Manager (AS-RTM).

The proposed design methodology and runtime software layer have been implemented and demonstrated on a real case study – an OpenCL stereo-matching application – targeting different industrial platforms. Some of the outcomes of this thesis have been used within the 2PARMA FP7 European project and implemented in official prototypes delivered to the project consortium.

A PROMPT GAMMA CAMERA FOR REAL-TIME RANGE CONTROL IN PROTON THERAPY

Irene Perali - Supervisor: **Carlo Fiorini**

Proton therapy is a form of radiation therapy that uses high-energy proton beams for cancer treatment. Differently from conventional radiation therapy, proton beams deliver their maximum energy within a defined range, thereby reducing adverse effects to adjacent healthy tissues. Proton beams open up new perspectives for the treatment of tumors in proximity to organs at risk, supposed the range of therapy particles is well under control. In clinical practice, range calculations are affected by uncertainties and safety margins are taken, preventing proton therapy from exploiting its maximum potential. Range assessment in vivo and in real-time is considered by clinical professionals a key for improving proton therapy. However, appropriate technical solutions are not yet available for the use in clinical routine.
Among different techniques for real-time range control, the thesis is focused on prompt gamma imaging, which is based on the fact the prompt gamma rays emitted by the target nuclei, after the interaction of the beam with tissue, are correlated to the beam penetration depth.
Imaging prompt gammas is challenging because they are emitted along a continuous energy spectrum up to 10 MeV and a detector with a count rate capability of tens of MHz is necessary. The objective of the thesis is to design a novel gamma detector, which satisfies these needs. The detector must be compatible to the introduction in a slit-camera system, where a knife-edge slit collimator selects prompt gammas emitted along the beam axis of the target, in order to produce a reverse 1D projection of the beam path on a scintillator crystal. The design of the camera system was first optimized with Monte Carlo simulations.
Preliminary measurements with the HiCAM camera, an Anger camera originally developed for SPECT applications, were performed operating the beam with currents lower than the ones adopted in clinical practice, demonstrating the feasibility of the slit-camera concept to reach millimeter accuracy in range determination. The measurements served the purpose of establishing the specifications for the new detector, compatible to the use at clinical beam currents. Several modifications were needed concerning the scintillator geometry, the photo-detectors, the electronics and the data acquisition.
Based on the knowledge on gamma ray detectors, we identified several technical options and investigated through calculations the feasibility of the different alternatives to find the best compromise in terms of performance, cost and simplicity.
A pixelated crystal configuration was preferred to a monolithic scintillator to reduce the count rate on a single channel. The crystal is segmented into 40 slabs with 4 mm pixel pitch, 30 mm thickness and 100 mm height. LYSO was chosen as best candidate due to its high density, fast decay constant an high light yield.
Silicon PhotoMultipliers (SiPMs) were selected for scintillation light collection because they offer compactness, good photon detection efficiency and fast response. We designed an optimized architecture of the systems, integrating all the components into a compact and practical instrument, easy to be introduced in the treatment room.
Due to the novelty of the adopted solution, the detector required a dedicated electronics. Custom electronics boards were designed to perform both spectra acquisition for a precise energy calibration and high efficiency photon counting for profiles reconstruction.
The best trade-off between energy resolution and counting efficiency was found during the design of the electronics boards.
We performed characterization tests of a reduced-size prototype to assess the design choices for the full-size camera.
Measurements confirmed that the slab configuration gives sufficient spatial and energy resolution, provided that the crystal surfaces are polished.
The full-size prototype was then characterized in terms of energy resolution at the energy peaks of ^{137}Cs and ^{60}Co (662, 1173 and 1335 keV, respectively). At these low energies we obtained an energy resolution of about 13 %, 9 % and 8 %.
The camera was tested in the West German Proton Therapy Center in Essen and in the Proton Therapy Center of Prague in order to evaluate its performance during beam irradiation, in different measurements conditions and with different targets. Both centers use the IBA C230 cyclotron, a isochronous cyclotron with a constant energy of 230 MeV. We first acquired profiles, by selecting the energy events in the 3-6 MeV energy range, during the irradiation of a homogeneous cylindrical PMMA target with therapeutic proton energies from 100 to 230 MeV and calculated that the number of protons needed to reach a 2σ precision of 4 mm in range retrieval is between 0.45×10^8 and 1.5×10^8 for the lowest and highest energies, respectively.
As observed from Monte Carlo simulations, the higher the beam energy, the higher the uncorrelated background, mostly due to neutrons. We verified the ability of the camera to detect millimeter shifts for high doses, by moving the target at steps of 1 mm to emulate a range shift. Profiles were compared to Monte Carlo simulations and showed good agreement. We noticed that simulations overestimate the correlated contribution.
We also observed the effect of a cavity in the target delivering a 2D map of pencil beams to a heterogeneous target. We observed that profiles corresponding to points of the map within the cavity are shifted with respect to the points in the full target.
Measurements at the WPE were useful to verify that the camera fulfills the requirements that we identified at the beginning of the project. Satisfactory accuracy in range retrieval was reached at clinical beam current for pencil beams with therapeutic doses.
At the Proton Therapy Center of Prague we were able to perform the first tests toward the clinical application of the instrument. We first compared profiles for targets with densities corresponding to bone and fat to profiles acquired with PMMA and water targets. After that, we applied for the first time a realistic treatment on an anthropomorphic phantom. Results are very encouraging because they demonstrate that the camera is able to monitor prompt gamma signal during a realistic PBS (Pencil Beam Scanning) delivery. Data processing and comparison with the simulations is ongoing. For the clinical use of the camera, there will be a set of simulated profiles for each spot of the treatment map and acquired profiles will be compared to simulated profiles in order to verify is a shift happened during the delivery of the treatment.
Finally, the very first measurements with the S2C2 accelerator, a compact superconducting synchrocyclotron, were performed to quantify the event rate of prompt gammas with the beam time structure specific for this accelerator. Results confirmed that the rate for a single slab is now 100 MHz and the current detector cannot sustain it. We should so consider several modifications in a novel design, such as a further pixelation of the crystal and the choice of a fully digital acquisition chain for pulse processing.
As stated in the first chapter, real-time range control in proton therapy would represent a major improvement in the delivery of this radiotherapy technique, which is already a valid method to defeat cancer. However, finding a technical solution that goes beyond scientific research and that can be clinically applied has represented the main challenge in this field. We hope that our efforts in this direction represent a milestone toward the reaching of this objective.

THE DESIGN OF EXERGAMING SYSTEMS FOR AUTONOMOUS REHABILITATION

Michele Pirovano - Supervisor: **Pier Luca Lanzi**

While the incidence of stroke rises worldwide, so do the costs of the subsequent intensive rehabilitation, setting off alarm bells that call for solutions to lower figures while preserving therapy efficacy. At-home autonomous rehabilitation appears as a promising solution, reducing costs for health providers and patients alike.
The trend of exergaming, i.e. exercising through video games, may represent the key to the success of autonomous rehabilitation. However, rehabilitation at home demands careful consideration, as all the requirements of a correct rehabilitation therapy must be addressed even in the absence of a therapist.
The aim of this research is to study the feasibility of at-home autonomous rehabilitation through exergaming. To do so, we explore the state-of-the-art of the exergaming field and devise guidelines to design effective and motivating exergames. We provide a novel definition of exergaming and we explore the design of its double nature, as exercise, and as game, and we provide guidelines for both. We create a methodology for the structured design and development of exergames for rehabilitation that leverages our new definition. We start from a given exercise, as defined by therapists, and we structure and detail it to produce a virtual rendition of the exercise that can be inserted imto a virtual environment. We then add these elements that partain to gaming that do not interfere with the underlying exercise mechanics to create an exergame through iterative prototyping, and we add the modules that allow automatic on-line supervision to obtain the final autonomous therapeutic exergame.
We design and develop a complete game engine for rehabilitation, built upon the Panda3D open source game engine, that integrates exergames and high-usability interfaces with autonomous supervision enabled by computational intelligence. The engine includes an abstraction of exergames that enables easy creation of new exergames for our system, knowledge-based on-line automatic monitoring through fuzzy systems, automatic on-line adaptation through Bayesian techniques, a layer that performs input abstraction and supports several different devices such as the Microsoft Kinect sensor and the Nintendo Wii Balance Board, and clear and meaningful feedback through high-accessibility and consistent interfaces and a Virtual Therapist Avatar. We also support asynchronous configuration and assessment by a remote therapist through recording and configuration functionalities.
We discuss extrinsic game elements that can be leveraged to increase the motivational factor of a rehabilitation exergame without undermining the benefits of the therapy, and we thus introduce and detail ad-hoc scoring mechanisms, algorithms to provide variations in content, and we detail long-term motivation mechanisms that leverage procedural content generation and interactive evolution methods.
We follow our guidelines to develop a set of nine games for posture and balance rehabilitation of post-stroke elderly patients. We conclude with results from several studies performed using our games, including a three-month pilot test with the complete system, proving the benefits of our solution.

STUDY, DESIGN, AND EVALUATION OF EXPLORATION STRATEGIES FOR AUTONOMOUS MOBILE ROBOTS

Alberto Quattrini Li - Supervisor: **Prof. Francesco Amigoni**

Autonomous mobile robotics has seen a wide spread development in recent years, especially for tasks that are difficult, dangerous, or simply boring for humans. Relevant examples include planetary exploration and search and rescue. There are several challenges that designers face during the development of systems of autonomous mobile robots, from low level issues, i.e., sensors, actuators, etc., to high level issues, i.e., control, navigation, etc. One of the most important aspects that affects autonomous mobile robots performance is the set of techniques that allow them to decide the next location to reach (*navigation strategies*) by possibly coordinating among themselves (*coordination method*), according to their current knowledge about the world they operate in, in order to autonomously carry out the assigned tasks.

To introduce the idea of navigation strategies it is useful to start from considering the (huge) literature about path planning, which shows that the core for most of the currently employed methods has been developed in mid-1990s.

In these approaches, a user specifies the goal and the robots can decide by itself how to go there. However, in several cases the goal might not be known a priori or the user cannot interact in real-time with the robots, and so these methods cannot be plainly used. What is needed is the development of navigation strategies that allow mobile robots to autonomously decide their next target locations, besides how to go to a specific target. Furthermore, the use of multiple robots can make the execution of the task more efficient, if they smartly coordinate among themselves. This dissertation focuses on exploration, in which one or more robots execute the following steps in order to discover and map the features of an unknown environment: (a) perceive the surrounding environment, (b) integrate perceived data in a map representing environment known so far, (c) decide where to go next and who goes where, (d) go to the destination locations chosen. Specifically, step (c) is the focus of the dissertation, namely the selection of interesting locations (exploration strategies) and their assignment to robots (coordination methods).

In spite of the importance of the exploration problem, general techniques that allow mobile robots to be fully autonomous are not mature yet.

First of all, exploration strategies are usually defined following two rather different approaches. On the one hand, they are defined in practical contexts of real (or realistically simulated) robots and are empirically assessed by testing them in some environments. On the other hand, exploration strategies are defined in theoretical settings. In these approaches, proposed methods are assessed using theoretical tools like worst-case bounds and competitive ratio in some classes of environments. However, sometimes assumptions are not fully realistic (e.g., infinite line-of-sight visibility).

Further, most of the exploration strategies and coordination methods proposed in literature base their decisions only on the current metric map, which represents the spatial features of the environment, like the position of obstacles. In the last years, several methods have been proposed to build semantic maps that associate semantic labels (e.g., 'corridor' or 'room') to portion of the underlying metric map. Despite the great effort in constructing semantic maps, the study of their use for exploration is still rather limited.

Finally, a lively debate on good experimental methodologies is currently ongoing in the autonomous robotics community, as they have not reached yet a maturity level comparable to that of other disciplines. The relative comparison currently made between different methods without an optimal reference makes it difficult to assess how much room for improvement an exploration system has. A more complete evaluation should involve an absolute comparison between the performance of (online) exploration strategies and the optimal (offline) performance in the test environments, based on the competitive ratio. Another issue is the difficulty in reproducing experiments as parameters are usually not reported in the descriptions of experiments, and thus it is not clear what factors impact the performance of exploration.

Given this background, in the general context of the multirobot exploration problem, the objective of this dissertation is threefold:

To contribute to bridge the gap between theory and practice for exploration strategies.

I contribute to define the problem of calculating the optimal off-line exploration paths under some realistic assumptions – i.e., robot with time-discrete and limited perception and environment represented as a grid. I analyze the relation between such discretization and its continuous counterpart and formulate the discrete problem as a search problem. Thus, I develop the first algorithm to find the (approximated) optimal exploration path. Simulation results show the viability of our approach for realistic environments.

Moreover, I contribute to strengthen the experimental results obtained with real (and realistically simulated) exploring robots, by theoretically analyzing in environments modeled as graphs worst and average cases. Specifically, I derive bounds on traveled distance by a robot employing some exploration strategies that consider distance and information gain as criteria to evaluate frontiers. The obtained theoretical results show that, in the worst case, taking into account also information gain in selecting the next destination location does not provide any advantage over considering only distance, while it does in the average case on graphs modeling realistic indoor environments.

To improve exploration strategies and coordination methods.

I define exploration strategies and coordination methods that embed information coming from semantic maps. This allows to privilege some specific areas of the environment. For example, if robots know that an area of an environment is labeled as 'corridor', then that area should be privileged and more than one robot should be allocated to it, so that the exploration of the environment is speeded up, as rooms are typically attached to corridors. I experimentally show that there is a significant improvement about the exploration of relevant and total areas of indoor environments within a given time interval, when a priori information about the relevant areas of the environment is available.

To improve the experimental assessment of multirobot exploration systems.

The method I propose for computing the optimal exploration path could serve as a tool for complementing the evaluation of on-line exploration strategies for autonomous mobile robots by having an optimal reference with which is possible to compute the competitive ratio of a given exploration strategy in a given environment.

Further, I experimentally evaluate (in simulation) the impact of some controllable factors on exploration (different perception/decision timings, and exploration strategies vs. coordination methods), providing some insights that could be useful for a roboticist that has to set these parameters.

In addition, I show how some of the artificial intelligence techniques used in this dissertation can be used for exploring a belief state space in the context of pursuit-evasion games, in which a pursuer attempts to capture an adversarial evader that tries, in turn, to actively escape, when they both have a line-of-sight sensor model.

The contributions provided in this dissertation could foster the achievement of the long-term goal towards the theoretical and practical definition and the evaluation of exploration strategies and coordination methods for increasing autonomy of mobile robots.

A TIME SERIES SYNTHESIZER OF TROPOSPHERIC IMPAIRMENTS AFFECTING SATELLITE LINKS DEVELOPED IN THE FRAMEWORK OF THE ALPHASAT EXPERIMENT

Laura Resteghini - Supervisor: **Carlo Capsoni**

The increasing diffusion of broadband internet connection and the development of Ultra High Definition TV (UHTV) requires the use of satellite communication systems able to provide large frequency bandwidth by making use of radio frequency carriers (Extremely High Frequency - EHF) up to Q/V Band. Some satellite-based service providers, such as EUTELSAT Company, start using satellite operating in Ka Band despite, at such these frequencies, the system has to cope with very strong attenuation introduced by the tropospheric constituents. Indeed, the physic of the channel is mostly influenced by the working frequency of the transmitting system, and the gathered attenuation, which affects the signal, is increasing with frequency. In this respect, satellite system can exploit the use of Propagation Impairment Mitigation Techniques (PIMT) to reduce the negative effects of fading due to atmosphere. The development and design of these systems must be supported by the use of tropospheric channel models and synthesizers able to characterize the time varying channel not only in statistical terms. For satellite systems operating at these frequencies the main source of signal degradation is represented by the rainfall events. However, even if rain strongly affects the signal power, it is quite limited in space and time. Moving towards the use of frequencies from 50 to 70 GHz even the so-called "clear-sky" attenuation due to atmospheric component such as gases, water vapor and oxygen as well as clouds become relevant, especially for systems with low power margin.
In this work we focused on the study and development of a time series synthesizer for main tropospheric components (water vapor, clouds oxygen and rain) that mainly affect the transmission of signals in free space condition. In particular, the aim of this study is the development of a generator of time series of attenuation to simulate the variability in time of the tropospheric components separately and their combination in total attenuation. The basic idea is to start from real measurements collected during propagation campaigns to generate time series of attenuation for each atmospheric component reproducing the statistic of attenuation at a selected location for defined Satcom parameters (frequency, elevation angle). The measurement database taken into consideration is the one provided by the ITALSAT experiment over 7 years of measurements. The input database is accurately modified and scaled to better adapt to the site-specific climatology of the location. The data collected by different instruments (radiometer, beacon receiver and raingauge) are used to generate a database of time series for water vapor, clouds and rain events. The synthesizer generates, for each atmospheric component separately, a time series of attenuation of measured data and finally a time series total attenuation obtained by the combination of all atmospheric effects (water vapor, oxygen, clouds and rain). In the first part of my work, we considered the non rainy attenuation components focusing in particular on the modelling of Integrated Liquid Water Content (ILWC) and Integrated Water Vapor Content (IWVC). One critical step of this retrieval is represented by the identification, discrimination and interpolation of the rainy periods in order to avoid the error in the estimate of ILWC and IWVC. The rainy periods identification is obtained considering three databases of measurements (raingauge, beacon receiver and radiometer) properly combined. We assumed 24 hours as the basic time frame for the time series of ILWC and IWVC. Daily time series are properly catalogued into a set of classes according to ILWC and IWVC maximum values. The site-specific statistics of ILWC and IWVC to be reproduced, provided either by measured data or by ITU-R models, are taken as input by two separated optimization procedures that return the number of daily time series to be selected from each ILWC and IWVC class. For the time series selection, we identify the daily time series of ILWC and IWVC that jointly satisfy the solution of the two optimizations. In the second part of the work we focused on the synthesis of rain attenuation events. The rain events collected in the database are catalogued in 10 classes according to the peak of attenuation they experience. A dedicated optimization takes as input the long term statistic of rain attenuation, provided either by measured data or by ITU-R models, and return the total amount of rainy time to be selected from each class. Then, the events of rain attenuation are randomly selected according to optimization result. Finally, the composition of all tropospheric components is achieved through a new algorithm based on the identification and classification of cloud types (Cloud Type Algorithm CTA). This new procedure guarantees an accurate superimposition of all events according to the physical properties of clouds. The validation of the model has been carried out on both first and second order statistics, showing the ability of the synthesizer in generating appropriate time series able to reproduce the input site-specific statistics. The model is valid for every link for site located in temperate regions with a range of frequencies is between 5 and 70 GHz and elevation angle between 5° and 90°.

CHARACTERIZATION AND MODELING OF CONDUCTION AND CRYSTALLIZATION STATISTICS IN PHASE CHANGE MEMORIES

Maurizio Rizzi - Supervisor: **Daniele Ielmini**

The information revolution has driven a substantial transformation in both our personal lifestyle and in the work organization within the companies.
Such impressive changes have so far been possible thanks to continuous performance enhancements in the semiconductor field. A key enabling factor has been the availability of memory technologies able to store data of larger and larger size in a faster and faster way.
The leading actor of such a success story has been the Flash technology, which is based on the ability to change the threshold voltage (and accordingly the read current) of a MOS transistor by injecting charge in a dedicated floating gate layer.
Given the requests from the memory market of even higher performances, Flash memories are currently facing severe challenges. The size scaling, so far well predicted by the Moore's law, is expected to be harder and harder below the 20 nm technology node. For this reason the semiconductor companies are strongly looking for novel concepts able to sustain the improvements seen so far.
Among the novel technologies proposed as next generation memories, a privileged position is currently hold by the Phase Change Memory (PCM).
PCM has the memory capability thanks to the property of particular materials (e.g. chalcogenide alloys) to reversibly switch between an amorphous and a crystalline phase with markedly different electrical resistivity. Such principle allows obtaining non volatile memories with interesting read/program speed, good scalability perspectives and a remarkable cycling capability.
Nowadays, a deeper knowledge of the PCM physics is strongly requested to drive the development of the PCM technology in the years to come. This motivates the need for research activities, such as the ones described in this doctoral work.

The introductory chapter of this work provides an overview of the current non-volatile memory (NVM) scenario, subdividing the possible technology evolutions within an evolutionary scenario and a paradigm shift. The phase-change technology is then introduced, dealing with its history, the basic operation and the elementary physical description.
The first chapter reviews the current state-of-the-art in the physical comprehension of sub-threshold conduction, threshold switching and crystallization, providing the basic elements needed for the comprehension of the following three chapters. Finally, the current perspectives of the PCM technology are discussed, with a quick glance on the so-called PCMS architecture, which is expected to solve the current PCM limitations in terms of size scaling by stacking the memory element and a cell selector made of another chalcogenide material.

The second chapter is devoted to the study of the electrical conduction in the amorphous phase. Such studies are of particular importance, given that PCM is a resistance-based memory. For this reason, a deeper knowledge of the conduction properties allows to optimize the resistance window between the two programmed states and to carefully optimize the cell geometry.
Amorphous materials present the peculiar property of a non-homogeneous conduction, because of the disordered nature of such materials. This poses some critical issues, since the prediction of the resistance value as a function of the cell geometry (e.g. for the cell down-scaling) is non-trivial. To deal with such problem, this chapter introduces a novel model for conduction, based on the so-called energy landscape. Such an approach allows taking into account the disordered nature of chalcogenide materials, providing some insights into the filamentary conduction. The model is then able to better describe some cell properties, such as the programming characteristics (resistance as a function of the programming current) and to better estimate the amorphous thickness from the I-V curves.
The energy landscape model is lastly extended to consider the carrier heating, at the basis of threshold switching. This phenomenon is crucial, since it enables the phase transition (by allowing a larger current to flow into the cell) and it limits the maximum read voltage. The proposed switching model is then validated as a function of the ambient temperature and of the amorphous size.
The present study, conducted as a function of the amorphous cap size for two nano-scaled technology nodes, is particularly relevant for the prediction of the conduction properties as a function of the cell geometry, possibly allowing to address the cell design in future PCM technologies.

One of the key properties of the PCM technology is its non volatility, which is guaranteed by the relatively long time needed for the relaxation of the (metastable) amorphous state into the (stable) crystalline one.
The third chapter of this work deals with a detailed study of the retention capability in PCM on a large statistical scale. Such studies are fundamental in order to allow large arrays to properly satisfy the data retention requirements.
A wide experimental characterization is then provided in the temperature range below 180°C, presenting a detailed study of the cell-to-cell and cycle-to-cycle variability. The overall variability is interpreted through a compact Monte-Carlo model, able to explain both the cell-to-cell and the cycle-to-cycle variability contributions in terms of a pure gaussian spread in the activation energy for crystallization.
Finally, the cycle-to-cycle variability is analyzed more in depth, allowing to subdivide the retention characteristics into three separate families namely: i) analog variability, ii) digital (binary) variability and iii) pseudo-repeatable characteristics.

The fourth chapter is dedicated to the statistical study of the program operation in PCM, or phase switching. The study is focused on the analysis of the programming characteristics at statistical level.
First of all the work describes the dependence of both the set (or crystallization) and the reset (or amorphization) operations on the initial reset state, determined by the reset voltage. The dependence of the melt current and of the obtained resistance distributions on the amorphous size are then presented and discussed.
The collected data allow to compare the high-temperature, set region with the low-temperature, retention regime providing the following observations: i) the previously observed non-Arrhenius behavior of crystallization is confirmed at statistical level and ii) the set operation is shown to have a tighter distribution of the crystallization times compared to the retention case.
The experimental characterization of the set operation as a function of the reset voltage (or the amorphous size) provides some useful indications for the choice of the ideal reset conditions and allows predicting the set dependence on the cell down-scaling.
Finally, the work compares two set techniques, namely: i) square pulses with amplitude lower than the melt level and ii) triangular pulses with slow quench from the melt. Resulting resistance distributions are analyzed and compared as a function of the pulse energy, concluding that the triangular pulses allow to obtain tighter R distributions at a given pulse energy.
The studies of the array-level statistics in PCM are rarely reported in literature; for this reason the statistical studies presented in chapter III and chapter IV of this doctoral work are particularly interesting for the array-level optimization of future PCM products.

ECONOMIC MECHANISM FOR ONLINE PAY-PER-CLICK ADVERTISING: COMPLEXITY, ALGORITHMS AND LEARNING BY MARCO ROCCO

Marco Rocco - Supervisor: **Prof. Nicola Gatti**

Sponsored search auctions (SSAs) constitute one of the most successful applications of *microeconomic mechanisms*, producing a revenue of about $42.8 billions in the U.S. alone in 2014, dominating display ads, the second largest revenue source. In a SSA, a number of advertisers bid to have their *sponsored links* (from here on *ads*) displayed in some *slot* alongside the search results of a keyword. SSAs currently adopt a pay–per–click scheme, which requires positive payments from an advertiser only when its ad is clicked. Given an allocation of ads over the available slots, each ad is associated with a *click–through rate* (CTR) corresponding to the probability that such ad will be clicked by the user. CTRs are estimated by the auctioneer and play a crucial role in the definition of the auction, since they are used by the auctioneer to compute the optimal allocation (in expectation) and to compute the payments for each ad.
In microeconomic literature, SSAs have been formalized as a *mechanism design* problem, where the objective is to design an auction mechanism that incentivizes advertisers to bid their *truthful valuations* (needed for economic stability) and that assures both the advertisers and the auctioneer to have a *non–negative utility*.

A crucial issue in SSAs is the study of effective *models* of the user attention and their exploitation in the auction mechanism. A number of works showed that *externalities* play an important role in the user behaviour. On the other hand, externalities may make the problem of finding the optimal allocation intractable, even when approximated. The most widely adopted user model is the *Cascade Model*, in which a user is assumed to scan the ads sequentially from the top slot to the bottom slot with a probability to observe the subsequent slot that depends on the last observed ad (*ad–dependent externality*) and on its position (*position–dependent externality*) and with the remaining probability the user stops to observe the ads. The computational complexity of the SSA problem when the Cascade Model is adopted is unknown, but it is supposed to be NP-hard. Moreover, the Cascade Model presents *limitations* in the way the externalities are represented w.r.t. the real world. I worked to overcome some of them introducing *three new models*. I provided a detailed characterization of the computational complexity of two of the three models. Afterwards, for the models that are NP-hard, I proposed approximation algorithms.

In literature it is known that there could be a gap in the guarantee of the approximation algorithm between the optimal approximation algorithm that can be designed and the optimal one under the constraint of truthfulness. For this reason, I studied both the situations, the first in order to give a computational complexity characterization of the problem and the second in order to provide a mechanism design result.
Studying the third model, I also enlarged the set of applications considering also the environment of *Mobile geo–location advertising*. Specifically, this is an environment where mobile ads are targeted based on a user's location (e.g., streets or squares within a city or a district). This field has been identified as a key growth factor for the mobile market and has not been widely studied, in particular not from a mechanism design point of view, a crucial ingredient for its success. In my final dissertation, I proposed exact algorithms and then I identified subclasses of instances where the problem can be easily solved in polynomial time. Finally, I proposed polynomial time approximation algorithms that can be implemented in truthful mechanisms.

The final part of my work is devoted to situations where mechanism design and *online learning* mash up. Specifically, in practical applications, it is not true that all the parameters of the model designed for the SSAs are known. This fact opens the problem of studying stable mechanisms while estimating parameters during the repetitions of the auction. The problem is challenging since it represents one of the first examples where online learning theory and mechanism design are paired to obtain effective methods to learn under equilibrium constraints (notably the truthfulness property).
In literature there are papers that study this problem when a single slot is available and a specific parameter called quality is unknown. I extended the study to the case of multi-slot instances considering also situations where parameters different to the quality are unknown. In order to obtain truthful mechanisms, it is necessary to adopt Multi-Armed Bandit (MAB) algorithms that separate the exploration and the exploitation phases.
In the final dissertation, I provided bounds over the loss of these mechanisms w.r.t. to the one adopted when all the information is available, considering two different measures of loss (regret): regret in the revenue of the auctioneer, i.e., how much the auctioneer loses, and regret in the Social Welfare, i.e., how much the community loses in terms of values of the allocation.

In conclusion, from a more general point of view, I can summarise my work in the following way. I studied *online pay–per–click/visit auctions* (e.g. SSAs and Mobile geo–location advertising) along the *four main perspectives* that are crucial for the success of an economic mechanism:
1) *computational complexity of finding the best allocation*: studying the computational complexity of a problem is important to deeply understand the hardness of the problem. Specifically, this aspect could be crucial in the choice of the model to adopt for a specific application;
2) *truthfulness*: an economic mechanism often is composed of agents that interact. Agents that are usually rational (selfish). Rules are required in order to handle the interaction and to guide it to a *stable outcome*. Otherwise, the market could become unstable and unpredictable. For this reason, in order to guarantee the stability it necessary to design *truthful mechanisms*;
3) *exact and theoretically bounded approximation algorithms*: once the hardness of a problem is known, the problem has to be solved. This requires the design of *algorithms*. In the case the finding the optimal allocation is not an easy problem, a study of *approximation algorithms* is required, otherwise the problem cannot be solved in practical situations. The study of the computational complexity could guide the choice of which approximation algorithm has to be studied. At the same time the truthfulness requirement influences the design of the algorithm too;
4) *online learning*: in real world environments, often, we face situations of *lack of information*, i.e. we do not have all the information required to make the best decision. Thus, in order to apply, in practice, our algorithms and mechanisms it is also important to study ways to handle this uncertainty, e.g. adopting online learning tools. Of particular interest is the study of *worst–case bounds* over the loss (regret) due to the lack of information w.r.t. the ideal case in which all the information is known. The truthfulness requirement influences these bounds.
The ideal goal is the design of very expressive user models admitting very efficient allocation algorithms that can be used in truthful mechanisms with the minimum online learning regret. In the work, I showed that this is never the case in practice. Indeed, each user model provides a different trade–off in terms of expressiveness, economic stability, approximation bounds, and online regret bounds and therefore there is not the best user model for each scenario, but each scenario potentially requires a different model.

DECOMPOSITION METHODS FOR QUADRATIC ZERO-ONE PROGRAMMING

Borzou Rostami - Supervisor: **Prof. Federico Malucelli**

The Quadratic 0-1 program with linear constraints is a very general class of optimization problems and has a wide range of applications. This problem has the following form:

QP: min $x^TQx + c^Tx$
s.t. $Ax = b$
$x \in \{0,1\}$

where $A \in R^{m \times n}$, $Q \in R^{n \times n}$, $b \in R^m$, and $c \in R^n$.

Many combinatorial optimization problems admit natural formulations as quadratic 0-1 programming problems. The Quadratic Assignment Problem (QAP), Quadratic Traveling Salesman Problem (QTSP), Graph Partitioning, Quadratic Knapsack Problem (QKP), and Quadratic Minimum Spanning Tree Problem (QMSTP) are among the well-known particular cases of the QP which arise in a variety of real-world applications. All of these problems are known to be NP-hard in general and therefore the is also NP-hard.
Branch-and-bound algorithms are the most successful procedure for solving the. The branch-and-bound algorithm is based on decomposition of the original problem into a series of smaller subproblems, and then recursively solves each subproblem, and discards the non-optimal solutions by using the best obtained lower estimated bound. In fact this approach behaves well only if one gets tight lower bounds for the objective function. In general, the solution methods of finding a lower bound for the can be divided into two main groups: Reformulation-Relaxation approaches, and Reformulation-Decomposition approaches. Since in general the is a non-linear non convex problem, most of the proposed approaches try to reformulate the problem either as an equivalent Mixed Integer Linear Program (MILP) or as an equivalent quadratic 0-1 program and solve the resulting program by effective algorithms that take the problem structure into account. Based on the structure of the resulting problem different relaxation and/or decomposition methods may be applied to provide a lower bound.
This thesis consists of two parts: Part 1 deals with the solution methods for the general case of the QP. We study different reformulations and relaxation strategies based on linear and semidefinite programming. More precisely we start with classic linearization methods to obtain a lower bound, and then try to improve the reformulation so that its LP relaxation provides a stronger lower bound in a reasonable time. Moreover, we propose the Semidefinite programming as a different approach which can be used either for generate a strong relaxation of the or to provide a convex reformulation of the problem. Then we describe different reformulations of the based on an equivalent convex or non-convex quadratic 0-1 programming. After introducing the different -based reformulation strategies, we use various decomposition techniques (including Lagrangian decomposition) to obtain a strong lower bound. Part 2 is concerned with the some special cases of the related to quadratic version of some well-known combinatorial optimization problems. Among the most important classical combinatorial optimization problems, we study the quadratic assignment problem (QAP), quadratic minimum spanning tree problem (QMSTP), quadratic traveling salesman problem (QTSP) and finally, quadratic shortest path problem (QSP).
The Quadratic assignment problem (QAP) is one of the classical difficult combinatorial optimization. Due to its wide verity of applications and its resistance to solution strategies, numerous researchers have studied the QAP and proposed both heuristic and exact solution methods. We review different reformulations and lower bounding procedure for the problem and then dedicate our efforts to analysis and the interpretation of lower bounds. The Reformulation Linearization Technique (RLT) applied to the Quadratic Assignment Problem yields mixed 0–1 programming problems whose linear relaxations provide a strong bound on the objective value. Nevertheless, in the high level RLT representations the computation requires much effort. We propose a new compact reformulation for each level of the RLT representation exploiting the structure of the problem. Computational results on some benchmark instances indicate the potential of the new RLT representations as the level of the RLT increases. Moreover, we study two special cases of the QAP including the Adjacent QAP and QAP on reducible graphs. The Adjacent Quadratic Assignment Problem (AQAP) is a variant of the QAP where the cost coefficient matrix has a particular structure. Motivated by strong lower bounds obtained by applying RLT to the classical QAP, we propose two special RLT representations for the problem. The first is based on a ``flow" formulation whose linear relaxation can be solved very efficiently for large instances while the second one has significantly more variables and constraints, but possesses some desirable properties relative to the constraint set. For the QAP on reducible graph we give a Lagrangian decomposition based on splitting the variables and then dualizing the copy constraint so that the resulting problem can be decompose to two quadratic semi-assignment problems.
The QTSP is as a variant of the classical Traveling Salesman Problem (TSP) whose costs are associated with each two edge that are traversed in succession. We first present the problem statement and propose some linearized integer formulations for both symmetric and asymmetric version of the problem. In order to obtain a tight lower bound to the problem we provide a Linear Programming formulation for the general QTSP that has a variable for each cycle in the given graph. Since the number of cycles is exponential in the graph size, we propose a column generation approach. We compare the bounds resulting from this new formulation with those obtained by some linearization techniques. Computational results on some set of benchmarks used in the literature show that the column generation approach is very promising.
The Minimum Spanning Tree Problem (MSTP) is one of the most known combinatorial optimization problems. It concerns the determination of a minimum edge-cost subgraph spanning all the vertices of a given connected graph. The Quadratic Minimum Spanning Tree Problem (QMSTP) is a variant of the MST whose cost considers also the interaction between every pair of edges of the tree. We review different strategies found in the literature to compute a lower bound for the QMSTP and develop new bounds based on a reformulation scheme and some new mixed 0-1 linear formulations that result from a RLT. The new bounds take advantage of an efficient way to retrieve dual information from the MSTP reduced cost computation. We compare the new bounds with the other bounding procedures in terms of both overall strength and computational effort. Computational experiments indicate that the dual-ascent procedure applied to the new RLT formulation provides the best bounds at the price of increased computational effort, while the bound obtained using the reformulation scheme seems to tradeoff between the bound tightness and computational effort.
Finding the shortest path in a directed graph is one of the most important combinatorial optimization problems, having applications in a wide range of fields. In its basic version, however, the problem fails to represent situations in which the value of the objective function is determined not only by the choice of each single arc, but also by the combined presence of pairs of arcs in the solution. We model these situations as a Quadratic Shortest Path Problem, which calls for the minimization of a quadratic objective function subject to shortest-path constraints. We prove strong NP-hardness of the problem and analyze polynomially solvable special cases, obtained by restricting the distance of arc pairs in the graph that appear jointly in a quadratic monomial of the objective function. Based on this special case and problem structure, we devise fast lower bounding procedures for the general problem and show computationally that they clearly outperform other approaches proposed in the literature in terms of its strength.

WSN POWER-PERFORMANCE OPTIMIZATION IN MULTI-APPLICATION SCENARIOS

Luigi Rucco - Supervisor: **Carlo Brandolese**

Our work presents a multilayer approach to support independent heterogeneous applications from multiple diverse users in wireless sensor networks. The aim is to provide significant performance, resources availability and abstraction from low-level details, while reducing energy consumption to a level comparable to that of ultra-low power and resource-constrained platforms. The following fundamental goals are at the basis of our research.

Support to heterogeneous applications. The support to heterogeneous applications must be firstly provided in the allocation phase, to ensure that the functional goals of the users are satisfied, preserving the efficient operation of the network. Secondarily, at node level, the hardware and software platform must provide sufficient resources and isolation to allow the concurrent execution of software coming from diverse sources.

Standardization of applications development. The application development should be as close as possible to standard programming models and environments, so that clients can avoid taking into account low-level platform details and develop their applications as they do for general-purpose programs. This requires, for instance, the use of standard programming languages, libraries and system calls, a file system and linking/loading mechanisms, as well as POSIX-like threads and processes.

Network lifetime. The first two goals ask for a level support that ultra-low power and resource-constrained devices may not offer. This entails the need for more powerful platforms and a higher level of abstraction that lead to increased energy consumption. Since the lifetime of the network plays a primary role both for the clients and for the service provider, guaranteeing that the superior resources do not lead to a fast energy depletion is a fundamental goal for the whole framework.

Guaranteeing quality of service. The need for energy-efficiency should not degrade the performance in terms of quality of service, as specified by the clients. One possible metric can be, for example, the respect of the deadlines and the adherence of the execution to the specified sampling and processing periods. A co-optimization of energy and quality aspects is therefore needed to avoid an unbalanced behavior toward only one of these two critical aspects.

To fulfill the goals above described, we propose a top-down approach, which traverses all the hierarchical layers of a WSN: from the network level, where opportune software allocation models are addressed, through the node level, which provides support to the execution of multiple heterogeneous applications, till the non-functional optimization layer that sustains the lifetime of the nodes.

As regards the network level, starting from the encouraging results obtained in previous works, our effort has been devoted to overtake some of the limitations characterizing the initial centralized software allocation model. As an evolution of our previous work, we aimed at defining an ILP model—which eliminates some restrictive hypotheses and sets the golden reference for the allocation problem—along with a very fast heuristic, lightweight and accurate enough to be embedded in cluster-heads and more powerful gateway-nodes. In perspective, the heuristic would possibly allow an in-network distributed allocation of clients' applications.

Continuing our top-down analysis, after having defined a suitable allocation model, we moved at node level with the aim of providing more computational and memory resources, as well as a higher degree of abstraction and standardization. In particular, the creation of an easy and user-friendly development model for the applications has been a main contribution in this direction. This has been achieved by bringing some of the features of general-purpose systems, such as standard processes and threads, static and dynamic linking and loading, standard programming languages and APIs, into our framework. Special attention has been paid to allow the execution of more complex tasks and the reduction of radio transmissions by leveraging local processing, as well as the memory isolation of multiple processes running on the same node. With the same effort, we have moved toward the simplification of code development, through the adoption of standard libraries and system calls and the support to code mobility for dynamically reconfiguring nodes. These objectives have required a combined hardware/software co-design to guarantee suitable performance of the platform.

The next step in our top-down flow has been devoted to address the non-functional optimization level. The first issue that we have tackled is related to the heterogeneity of the sources that deploy their applications on the network, which may lead to very inhomogeneous periods and tolerable slacks. This can cause a fragmented duty-cycle on the nodes, eventually leading to fast energy depletion and operation inefficiency. At first, thus, we investigated the possibility of merging as much as possible the execution moments of the periodic tasks, in order to reduce the energy overheads of frequent awakenings of system and devices. To do so, we determined how to break up the periodic evolution of the system in small self-contained sub-problems, on which determining the minimal sequences of system awakenings and choosing the sequences with best quality can be computed with reasonable effort and without loss of global optimality. In this direction, we have devised an optimization approach capable of significantly enhancing the energy efficiency of the nodes while minimizing quality-of-service loss.

After the rationalization of the duty-cycle, the subsequent step within the non-functional optimization layer has regarded the adaptation of the node operation profile to the actual computational needs, to ensure that in no case the energy consumption exceeds the bare necessities of the applications. In particular, we started from a careful analysis on the specific characteristics of the sensing and the processing stages in wireless sensor networks. In classical approaches to WSNs programming, in fact, the phases related to data sensing and data processing are strictly mingled in the application layer: programs are usually developed as cycling loops in which measurements are performed on the considered physical parameter and, soon after, the retrieved data are sent toward the sink via the radio channel. From this paradigm ensues a general behavior of the network which can be very energy consuming. We have designed the operation of our framework starting from the assumption that locally analyzing data on the end device can strongly reduce the energy consumption of the network, because of the higher costs associated to data transmission over the radio channel with respect to local processing. From this standpoint, the isolation of the sensing phase from data processing and both from data transmission, introduces a higher level of flexibility in optimizing the three phases separately. By decoupling sensing from processing, extending idle periods at the deepest possible power states is also possible. This, however, may cause loosing the memory status when the microcontroller is powered-off or put in stand-by. This new problem mainly concerns the preservation of the status (registers, stack, heap) of the applications between two subsequent active periods and requires a power-efficient hibernation mechanism. Consequently, a formal hibernation model has been carefully defined to ensure that the best choice is always made between a complete system hibernation and a low-power sleep state.

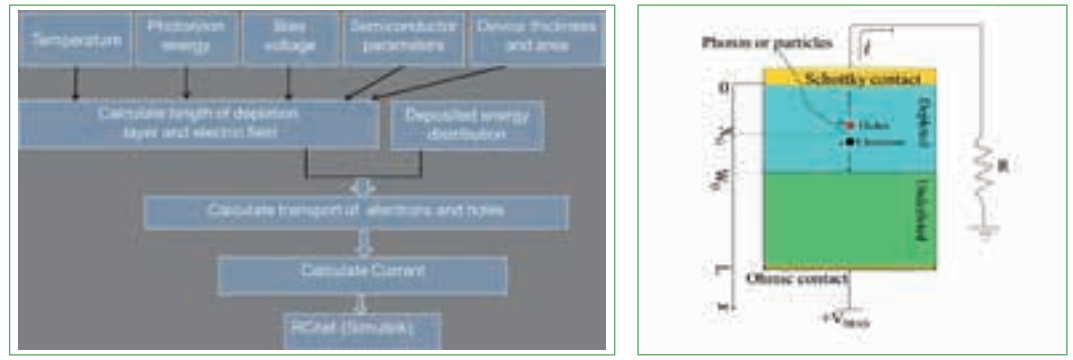
We extensively validated our models by developing a prototype hardware/software platform and, where needed, suitable simulation frameworks that stress the models at their limits. Results confirmed a behavior that fully complies with the general objectives of the work.

STUDY ON SILICON CARBIDE RADIATION DETECTORS FOR LASER-PLASMA RADIATION APPLICATIONS

Yongbiao Shi - Supervisor: **Giuseppe Bertuccio**

The topic of my thesis is devoted to the study, design, characterization and application of Silicon Carbide (SiC) detectors for photons and charged particles, like alphas, protons, ions, especially for laser-generated plasma radiation experiments. High purity and thick SiC epitaxial layers have been used to realize radiation detectors. The application of SiC detectors in laser-generated plasma physics has been done: in particular, the effects of the ultra-high intense radiation levels on the detector will be studied both theoretically and experimentally. The charge carrier transport in SiC has been studied and the timing performance of the device has been theoretically determined and experimentally demonstrated.
Laser-generated plasma experiments related to ion acceleration and nuclear reactions are worldwide carried out and many great results have been achieved in the last decades. Traditional radiation detectors used in laser-generated plasma experiments are showing their limits and new detectors with better performances are required. In the last few years, SiC radiation detectors have been so proposed to be employed in these experiments due to their great physical and electrical properties.

Under this framework, several SiC detectors have been characterized during my thesis work acquiring interesting data on their properties and performance in terms of response speed, time resolution and tolerance to plasma radiation. Time response of SiC detectors has been studied by means of both simulations as experiments.
A simulator based on Matlab and Simulink has been realized for this study. Figure 1 shows the block schematics of the simulator: parameters such like temperature, photon or ion/particle energies, detector geometry and bias conditions, doping of the SiC layer can be set.
The Simulations of the response of SiC detectors to photons have been carried out obtaining very fast signals with rise time and width of few hundreds of picoseconds, which are in a well agreement with the experimental data.
The response of SiC detectors to alphas and protons with different energies have been studied as well, considering the deposited energy distribution given by the available simulator SRIM 2013. Minimum rise time and pulse width of 0.3 ns and 0.7 ns, respectively, have been predicted for 5.5 MeV alpha particles. Similar results have been obtained with 2 MeV protons. A study devoted to the optimization of SiC detectors operating in time-of-flight configuration has been done by means of the designed simulator. The effect of the doping of the SiC epitaxial layer, of the detector geometry and bias conditions have been studied to establish the criteria for designing ultra-fast SiC devices for time of flight detectors.
The laser-plasma experiments were conducted in Prague Asterix Laser System (PALS) under the project of "High energy proton acceleration by thin hydrogenated-doped silicon dielectric targets using a sub-nanosecond laser" (HEPA) organized in collaboration with Dr. A. Picciotto from FBK-IRST, Italy. The project was aimed to maximize the proton/ion energy yield at target interaction of a sub-nanosecond laser PALS with hydrogen targets. The project also planned to compare the ion/proton energy with those achieved with advanced silicon, metallic and polymeric targets, keeping the target thickness constant. The goal of the project was to propose new semiconductor materials as possible new targets to be employed in laser-driven particle acceleration experiments instead of the standard metals or polymers. Our SiC detectors were used together with other

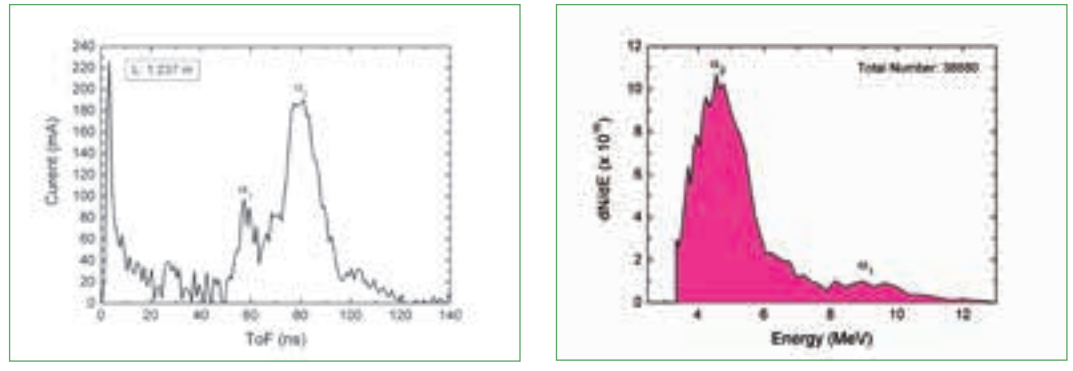


a b

1. (a) Simulator developed on the physical model of SiC detectors. (b) Simplified physical model of SiC detector performance.

nuclear radiation detectors to analyze the radiation emitted from the plasma. Great results have been achieved from our SiC detector from this project: nuclear fusion was firstly fund with so low density of laser boron interaction, high yield production of alpha particles was demonstrated using SiC detector by triggering the proton (p)-boron (11B) nuclear reaction (1)current signal pulse amplitude as high as 1.8 A has been acquired with our 5 mm^2 SiC detector, 0.8 ns risetime and 1 ns pulse width has been measured with our 1 mm^2 detector, 8 well resolved peaks have been detected within only 20 ns, and 2 peaks within less than 2.2 ns have been detected in this experiment. It has been proved that Interdigitated SiC detectors are more sensitive to slow ions than more conventional pad SiC detectors. By achieving high signal to noise ratio and nanosecond time resolution, the advantages of SiC detectors have been demonstrated over the traditional detectors used in laser experiments, like Faraday Cup. In addition, a method for calculating the number of particles from the signals delivered by time of flight semiconductor detectors has been developed, in the figure, α1 and α2 are identified by the CR-39 nuclear track detectors and our SiC detector allowed to determine their energy distribution with very high precision.

1. A. Picciotto *et al* "Boron–proton nuclear fusion enhancement induced in silicon targets by low-contrast pulsed laser," *Physical Review X*, 2014



a b

Figure 2. (a) Signals form shot 44023 acquired with circular pad SiC detector, the alpha signals start from 50 ns and end at 97 ns. (b) Calculated alpha number of Figure 2 (a), the total number is 38880

THERMAL AND ENERGY MANAGEMENT TECHNIQUES FOR MULTI-CORE AND MANY-CORE SYSTEMS

Federico Terraneo - Advisor: **Prof. William Fornaciari**

Introduction and description of the objectives

Today the semiconductor industry is facing increasing problems to continue delivering performance improvements at a pace that is now expected by its user base, as well as the general public. This is not yet caused by problems in achieving feature size reductions, but by the side effects and nonidealities caused by said scaling.

The failure of Dennard scaling in deep nanometer architectures results in an ever worsening power density increase, eventually leading to the dark silicon problem, where power and thermal constraints limit the number of transistors in a chip that can be switched at the maximum clock speed to an ever decreasing fraction. This problem, if not solved, could be a major roadblock in the evolutionary path from multi-core to many-core architectures.

Power efficiency alone is not enough to mitigate the dark silicon issue, as one of the major problems caused the power density increase is the need to effectively dissipate the generated heat away from the silicon die to prevent immediate failures as well as reliability issues caused by high operating temperatures. This problem will also be worsened in the future with the introduction of advanced chip technologies, such as 3D die-stacking.

Another important consideration is the high variability in the load experienced by multi-cores, caused by the potentially very different activities performed by the cores. In this perspective, effective dynamic thermal management solutions that can push the cores to their maximum performance subject to the constraint imposed by the need to remain within safe operating temperatures is a key aspect to achieve the best utilization of the computational capabilities of current, and especially future multi-cores.

Methodology

The thermal dynamics of both conventional 2D multi-cores as well as 3D die-stacked ones show two separate time scales: a slow one, in the order of seconds to a few minutes, and a fast one, in the millisecond range.

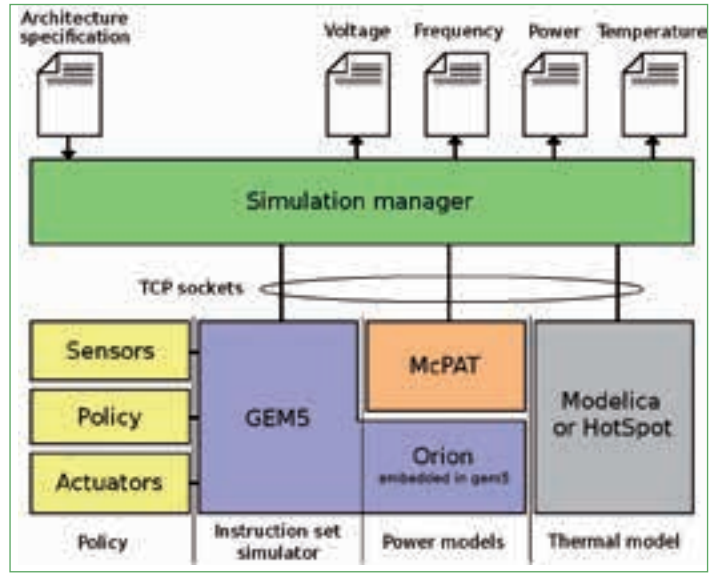
The fast dynamic is that of the silicon layer, that has a small thermal capacity, and owing to the non-negligible thermal resistance of the silicon bulk, can "swing" very rapidly with respect to the temperature of the heat sink. The slow thermal dynamic, is conversely that of the heat sink. To effectively control the temperature of future generation multi-cores, controlling the fast thermal dynamic becomes necessary.

However, traditional thermal policies would need to be operated at too fast rate, and interrupting the cores so frequently to run the policy would result in an unacceptable overhead.

The proposed policy relies on event-based control theory to couple the fast reaction time needed to counteract abrupt temperature changes with a low overhead. A hardware-software split is proposed where a hardware state machine generates events when the temperature of a core changes by a given threshold or a programmable timeout occurs, while a software policy implements the control algorithm. Since the policy is no longer executed periodically, the intervention rate is adaptive, depending on the variability of the core temperature.

This solution couples the performance benefits of a policy implemented entirely in hardware with the flexibility of a software policy.

A simulation flow has been developed in order to validate the proposed policies, composed of a cycle-accurate multi-core instruction-set simulator, a power model and a thermal model. The instruction-set simulator is based on GEM5, extended to support multiple voltage and frequency islands on the chip, down to a per core granularity. The simulator includes a dynamic model of a PLL and voltage regulator to account for the overhead of DVFS transitions. A model of the resynchronization logic for signals that cross frequency domains has also been implemented to account for the resynchronization overhead. The McPAT and Orion power models are used to produce power traces from the execution data produced by the GEM5 simulator. A flexible thermal model has been developed using the Modelica language, also supporting 3D die-stacked chips. The model is component-oriented meaning that each individual component is modeled separately with its own differential equations, maximizing flexibility.



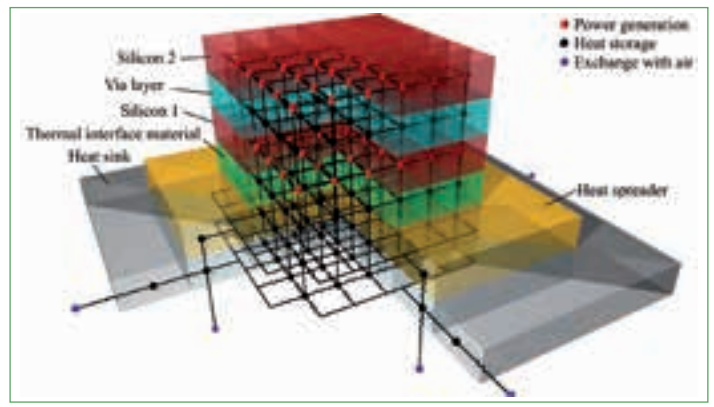
1. The developed simulation flow

Discussion of the results obtained

An innovative thermal control strategy was proposed, based on event-based control theory. This control scheme was tested using the developed simulator, by running standard benchmarks on a simulated 24 core 3D die stacked testbed chip, in order to assess its suitability for novel architectures exploiting 3D stacking. The selected set of benchmarks exercised the ability of the policy to withstand diverse workload requirements, such as both CPU bound and I/O bound applications. The proposed scheme outperforms state of the art fixed rate control strategies, evidencing the inherent advantages of a solution that dynamically and autonomously adapts the controller intervention rate to the application needs, without the need for application-specific tuning.

Moving to the technological side of the matter, the devised control solution is particularly suitable to be implemented in real multi-core architectures, thanks to its flexibility and negligible overhead. In fact, the implementation of the controller was proven to take a few tens of clock cycles.

The proposed event-based controller, was thus shown to achieve a level of performance comparable to a fully hardware solution while retaining the flexibility of a software implementation.



2. The simulated 3D multi-core processor

CONTROL OF A BRAKE BY WIRE ACTUATOR FOR RACING MOTORBIKES

Fabio Todeschini - Supervisor: **Sergio M. Savaresi**

During the last years in the automotive field - both from industrial and academic side - a significant effort has been done for the evolution of active safety and/or performance vehicle dynamic control strategies, as well as for the development of autonomous vehicles. The practical application of these systems is naturally subdued to the availability of on-board actuators that allow to regulate the desired control variable independently - or with certain degrees of freedom - from driver's willing. This need has led to the so called *drive-by-wire* paradigm, where the standard mechanical connection between driver and actuators is replaced by an electronic system, devoted to the regulation of the actuator according to driver or vehicle control logic requests.
Among others, the brake-by-wire (BBW) technology focuses on the design of an actuator capable of applying the desired braking torque to the vehicle's wheel. During the years, different technical solutions have been explored. The most successful ones are the electro-mechanical (EMB) and the electro-hydraulic (EHB) architectures. In the former, an electric motor directly exerts the requested torque on the wheel. A promising variant of such architecture is the Electric Wedge Brake (EWB), which exploits the DC motor torque to move a wedge where the braking pads are installed, rather than exerting a clamping force directly on the braking disk. The EHB - directly derived from the most spread vehicle brake architecture - employs a hydraulic system, activated by an electronically commanded motor/pump, to generate the requested braking force.
In this thesis a hybrid EMB/EHB solution is considered: a traditional hydraulic brake is employed and, using an electric motor mechanically connected to the master cylinder, the desired pressure on the braking pads is generated. With respect to the well known EMB and EHB solutions, the considered one has the advantage of keeping the usual vehicle hydraulic brake layout, adding just the electro-mechanical actuator, thus saving space, weight and cost. Furthermore, it gives flexibility about where to locate the BBW actuator and it does not increase the sprung mass of the vehicle.
This actuator is specifically designed by Brembo for high performance motorbikes.
The BBW actuator main goal is to provide the desired braking torque at the wheel; for this, the straightforward control variable should be the braking torque. However, due to the sensor reliability, cost and encumbrance, the system is equipped with a pressure sensor. Therefore, the actuator control problem consists in the tracking of a reference pressure that comes from the driver-vehicle interface (*i.e.* the braking pedal) or from other dynamic control strategies.
For such a BBW technology, the control problem, however, turns out to be quite challenging: firstly because of the typical non-linearities of the traditional hydraulic layout - *e.g.* presence of brake fluid reservoir, oil compressibility - along with those related to friction and temperature variations. Moreover, the required control performances are highly demanding, both in terms of bandwidth, both in terms of absence of overshoot.

The starting point of the control design is the system model: writing each actuator component equations, a physical based complete model is derived. This model represents very well the system experimental response and it is useful to adopt it as a simulator; however it is too complex to be employed for control design purposes. For this reason, besides the complete model, a control oriented one is derived. This model provides a sufficient system description for control purpose; its analysis gives important guidelines for the design of the control algorithm.

Once the system to be controlled is modeled, the control problem can be faced; in this thesis it is solved following two approaches.
A first one is focused on exploring innovative and advanced control techniques; these control techniques are studied and then adapted to the particular system, finally they are validated on a test bench. Pursuing this approach simplifies the testing phase: in fact, validating the control algorithms on a test bench allows us to discard the safety requirements, which are really strict and severe in this application. This is not possible when testing the control algorithms on a real system, *i.e.* on a racing motorbike.
The second approach is focused on solving the control problem adopting conventional control techniques; in this case the control algorithm are studied and designed based on the system model, then they are implemented and tested on a real motorbike. For this reason, in parallel with the control design, the fault detection algorithm design performs an important role.

In pursuing the first approach, firstly we deal with the friction, which has a dominant role in the BBW actuator. In this context the classical dithering compensation is compared with a more sophisticated model based compensation approach. In particular, the model based approach approximates the non-linear friction model with an innovative linear-in-the-parameters (LP) approximation, this is updated through an adaptation mechanism.
In parallel to the adaptive friction compensation technique, a sliding mode based controller guarantees the pressure tracking.

Then, anti-windup compensation technique is employed in order to build a compensator able to linearize the system intrinsic non-linearity. At this point, considering that in real systems control action has an upper physical saturation due to hardware limitation, an anti-windup compensation is derived. The employment of these two compensators permit to design a pressure control on a linear system, without considering the intrinsic system non-linearities. Moreover, this control algorithm does not discard the control action saturation in the design phase.

Finally, the Virtual Reference Feedback Tuning (VRFT) paradigm is experimentally validated on the test bench; in doing this we focused on a particular task that it could be necessary when designing the control algorithm: the design of a position controller. In this particular case of study we compared a canonical model based PID tuning with the VRFT approach, showing that the VRFT algorithm provides satisfactory performance. This approach makes the control design faster and avoids critical phases such as the identification one and the controller designed one.

Following the conventional control techniques and exploiting the peculiarities that the control oriented model highlights, we propose two different control strategies: an adaptive position-pressure switching control and an adaptive cascade position-pressure. These control strategies are simple, robust and easy to tune. Their design phase is performed on the test bench, then they are implemented and tested on a real motorbike, showing satisfactory performances. Moreover, due to the safety critical application, a fault detection algorithm suited for the particular case of study is presented.

A COGNITIVE FAULT DIAGNOSIS SYSTEM FOR SENSOR NETWORKS

Francesco Trovò - Supervisor: **Prof. Manuel Roveri**

Sensor networks represent a valuable technological solution to monitor and acquire data from an environment, a critical infrastructure or a cyber-physical system. These data are generally used as input for an application, which is able to react to changes occurring to the inspected system. Examples of these applications based on sensor networks, are those inspecting an environmental phenomenon (e.g., a river or a rock wall) protecting critical infrastructures or those monitoring the behaviour of a water distribution network. These monitoring systems are composed by a set of sensors, processing boards and a data transfer apparatus. During the operational life of this infrastructure, the sensor network continuously inspects the system status, by sending measurements to the central processing station, where the application runs. In fact, based on measurements coming from the sensor network, applications can be designed to take decisions (e.g., in the case a deviation from the usual working conditions is registered, an alarm is raised) and contextually react to the change (e.g., for critical infrastructures, alert the population about the threat). In this scenario, a model for the inspected phenomenon is usually unknown and the assumption of process stationarity may not hold, even though these assumptions would generally improve the decision abilities of the aforementioned applications. Thus, information coming from the sensor network is critical to monitor the underlying process, to check the status of the system and react according to its behaviour.

Sensor networks usually work in harsh conditions, which may induce faults, thermal drifts or ageing affects affecting both the embedded electronic boards and the sensors. In fact, they are affected by physical degradation (due to e.g., humidity, dust, chemicals and electromagnetic radiations), which may induce a gradual deviation of the measured value from the real one. Thus, it is of paramount importance to promptly detect and diagnose faults occurring in a specific unit, since they could affect the application layer, which operates based on the assumption that provided information is not corrupted. If the application does not take into account the possibility of a fault, it may take an incorrect decision, e.g., not alert the population when a threat is present or vice versa. Moreover, in the sensor network scenario the unexpected deviation from nominal conditions may be caused either by a fault or by a change in the inspected process. It is crucial to distinguish between these two situations: in the former case, direct maintenance should be performed, while the latter one requires a reaction to an environmental change, specifically chosen based on the considered application scenario. Traditional Fault Detection and Diagnosis Systems (FDDS) are systems specifically designed to detect and diagnose faults possibly occurring in complex systems. More specifically, the tasks of a FDDS are: detection of the fault, i.e., to promptly understand whether a deviation from the nominal state has occurred; isolation, i.e., to determine which unit is providing faulty measurements; identification, i.e., to capture the main characteristics (e.g., type, intensity) associated to the fault.

One of the main drawbacks of traditional approaches for FDDS is that they generally require a priori information on either the system in nominal conditions or the possible faults. Thus, their direct application to the sensor network scenario is far from being trivial, since, as pointed out before, we do not have information about the model of the system generating the data, or the possibly occurring faults.

In recent years, a novel and promising cognitive approach has been proposed to design FDDS. This novel generation of Cognitive Fault Detection and Diagnosis Systems (CFDDS) is able to automatically learn the nominal and the faulty states in an on-line manner, and is generally characterized by the ability to exploit temporal and spatial relationships present among the acquired data. Most of existing cognitive FDDS apply the cognitive approach only to a single aspect of the system, thus they still require at least to know partial information about the analysed process, or they are addressing the design of CFDDS for specific applications.

In this dissertation we propose a new CFDDS meant to operate on sensor networks. The proposed system is able to characterize the nominal conditions of the system, by relying on fault-free data coming from the sensor network. At first, the proposed system learns the dependency graph existing among datastreams, to select only relevant functional relationships. Based on them, the proposed CFDDS is able to perform fault detection, isolation and identification, without requiring a priori information on either the inspected process or the faulty states.

More specifically, to model the relationships constituting the causal dependency graph of the sensor network we rely on the concept of Granger causality, which allows to consider only those relationships providing meaningful information for fault detection and diagnosis. After learning the network dependency structure, fault detection and diagnosis is carried out in the space of estimated parameter vectors of linear time invariant models approximating the functional relationships included in the dependency graph. Deviations from the learned nominal concept are detected by means of a decrease of the log likelihood provided by the Hidden Markov Model modeling of the parameter vectors sequences. Following the detection phase, an isolation mechanism based on the logic partition of the dependency graph is able to distinguish between faults and change in the environment. Finally, if a fault has occurred, an identification procedure is executed to characterize different faults by relying on a newly developed evolving clustering-labeling technique in the space of the parameter vectors, which learns the fault dictionary in an on-line manner.

The innovative aspects of this cognitive framework for fault detection and diagnosis are:

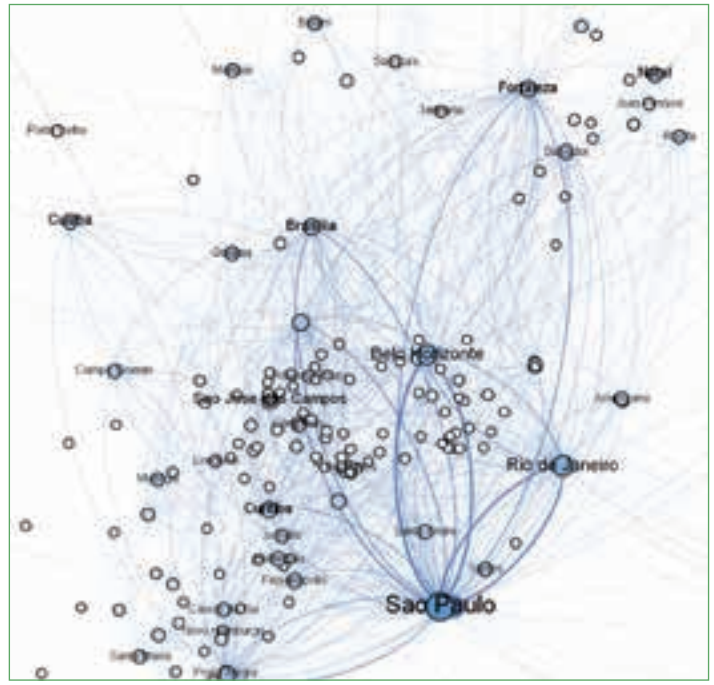
- the design of a CFDDS completely based on the cognitive approach, which is able to cover all the phases of fault detection and diagnosis in the sensor network scenario;
- the development of a set of integrated techniques, coming from statistics and machine learning fields, which rely on a theoretically sound framework developed by the system identification field;
- the ability to characterize the temporal and spatial relationship existing among data with the dependency graph, learned with the use of a statistical framework;
- the ability to characterize the nominal state of the system inspected by the sensor network through learning mechanisms based on the cognitive approach;
- the ability to learn the fault dictionary during the operational life of the system, without requiring a priori information about the possible faults.

SPATIAL ANALYSIS OF ONLINE DATA TO TRACK CITIES' SOCIO ECONOMIC INDICATORS AND URBAN LAND USE

Carmen Karina Vaca Ruiz - Supervisor: **Piero Fraternali**

Online geolocalized data is being massively produced as a result of both, the interactions on online social networks, and the content shared on the Internet that is annotated with geographical locations. This constitutes a rich source of information to characterize geographical places where either the people interacting reside or where the geo-tagged content is produced. Urban resources are allocated according to socio-economic indicators, and rapid urbanization in developing countries calls for updating those indicators in a timely fashion. The prohibitive costs of census data collection make that very difficult. To avoid allocating resources upon outdated indicators, one could partly update or complement them using digital data. In this dissertation we propose methods to estimate urban indicators as well as an unsupervised learning framework to discover dynamic areas of the city using the geotagged content published by either residents or visitors.
First we conduct and analysis of online attention patterns evolution in a content sharing platform. Evolution of online social networks is driven by the need of their members to share and consume content, resulting in a complex interplay between individual activity and attention received from others. To shed light on the matter, we look into the patterns of activity and popularity of users in the Yahoo Meme microblogging service. We observe that a combination of different type of social and content-producing activity is necessary to attract attention and the efficiency of users, namely the average attention received per piece of content published, for many users has a defined trend in its temporal footprint. The analysis of the user time series of efficiency shows different classes of users whose activity patterns give insights on the type of behavior that pays off best in terms of attention gathering.
Second, we analyze a random sample of interactions in the same service but focusing on content generated in Brazil and accurately predict the GDP and the social capital of 45 Brazilian cities. To make these predictions, we exploit the sociological concept of glocality, which says that economically successful cities tend to be involved in interactions that are both local and global at the same time. We indeed show that a city's glocality, measured with social media data, effectively signals the city's economic well-being. To this end, we aggregate the attention that the city's residents are able to attract on the platform at the level of the city and quantify it using a set of metrics that are put together in a linear model that accurately predict the GDP.
Finally, we propose an unsupervised learning framework to capture the composition of cities. To discover functional areas in a city, spatial discovery algorithms have been recently applied to social media (e.g., Foursquare) data: functional areas are often identified based on semantic annotations of places and human mobility patterns.
We propose a framework based on an objective function to maximize. By being integrated into any clustering algorithm, this function aims at finding and labeling areas such that an area's label is semantically related to the points in the area and to those in the area's neighborhood without being too general (e.g., the label 'clothing stores' is preferable to 'professional places'). We evaluate the framework with a hierarchical clustering algorithm upon Foursquare data in the cities of Barcelona, Milan, and London. We find that it is more effective than baseline methods in discovering functional areas. We complement that evaluation with a user study involving 111 participants in the three cities, and with an additional temporal segmentation of areas upon Flickr data. The results generated by our framework can benefit a variety of applications, including geo-marketing, urban planning, and social recommendations.

1. City graph depicting online attention exchange



2. Functional clusters identified in Barcelona

EXPLORING MOTION-BASED TOUCHLESS INTERACTION FOR AUTISTIC CHILDREN'S LEARNING

Matteo Valoriani - Supervisor: **Prof.ssa Franca Garzotto**

Autistic Spectrum Disorder (ASD) and autism are both general terms for a group of complex disorders of brain development, characterized by a triad of symptoms related to lack of social interaction, deficits in the acquisition and expression of language, and repetitive patterns of behavior often accompanied by sensorimotor impairments. For years, different techniques have been used to improve the quality of life of people who have various developmental dis- abilities. However, the use of technology continues to receive limited attention, despite the fact that it tends to be a high interest area for many of these children.

This work tries to present a solution in this broad and varied panorama developing innovative interactive technologies for autism that can be integrated with therapeutic and school activities and can be autonomously used by therapists and teachers to promote, through engagement: social interaction, communication capabilities and motor skills. Despite a general lack of interest in this field, in the last years we have seen an increasing number of technologies in the research literature and on the market focused on helping and educating children with autism. Existing products and prototypes support a variety of interaction modes and have been designed for different platforms and input devices, from conventional mouse or joysticks to (multi)touch gestures, speech-recognition devices, physical manipulation of digitally augmented objects, robots, head-mounted devices for autistic children. Still, very few studies have explored the potential of full-body interaction, and in particular of motion-based touchless interaction. In the research arena, motion-based touchless interaction has started to be explored for learning and therapeutic purposes. Still, most existing works consider the domain of "regular" children, while very little is known about how motion-based touchless interaction works for autistic children and if this paradigm can be successfully applied to these subjects.

This work wants to provide three different contributes to the current research on the subject: a theoretical one, i.e. a set of guidelines for the development of applications for children with cognitive and motor disabilities; a technological one, a flexible architecture to support the development of applications according to such guidelines and three examples of games for autistic children with low-moderate cognitive deficit, low-medium sensory-motor dysfunction, and motor autonomy, developed and tested on the field; a demonstration of the initial hypothesis that the touchless paradigm can be successfully applied in the treatment of autistic children. In the first part, this work will provide an overview of the current state of art in the field of nonconventional interaction paradigms for the treatment of autistic children, including virtual reality, tangibles, robots, mobile devices, discussing in particular touchless motion-based interaction with respect to autism.

In the second part, I will display through an empirical experimentation and using play therapy, the actual effectiveness of this type of interaction with children with autism spectrum disorder. The experiments show positive results, but a further analysis of the usability highlights that children ran into a number of problems in using commercial games currently on the market. On the basis of such observations, and thank to the collaboration with three therapeutic and educational centers which participated in the phases of development, design and evaluation, a structured set of generic and specific guidelines has been distilled, as presented in the second part of this work.

These guidelines informed

1. empirical experimentation

the design of three touchless games, based on Kinect, aimed at improving children's motor, cognitive and social skills, whose development informs the third part of this work. Considering every child's uniqueness, the games are strongly customized and "user friendly", in order to allow for therapeutic purposes, both at the centers' and in remote mode.

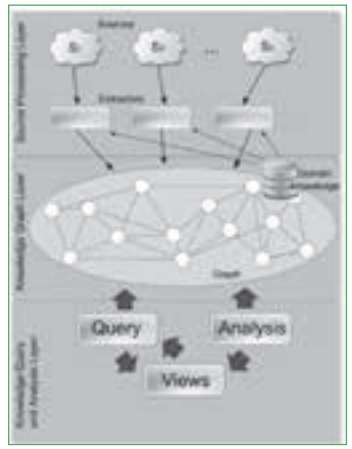
2. example of game developed using guidelines

As described in the fourth part of this thesis, the games have been evaluated in a controlled study remarking improvements for the areas of interest. The results of these empirical studies have confirmed and extended the outcomes of prior research, providing additional empirical evidence that touchless gaming does have a strong potential to improve autistic children's attention and motor- visual skills. Overall, the research sheds a light on the opportunities offered by full body touchless games for therapy and education of these special users. This thesis ends with the proposal of a research agenda in this field and an outline of future work

THE SECOND LIFE OF TELEVISION CONTENT

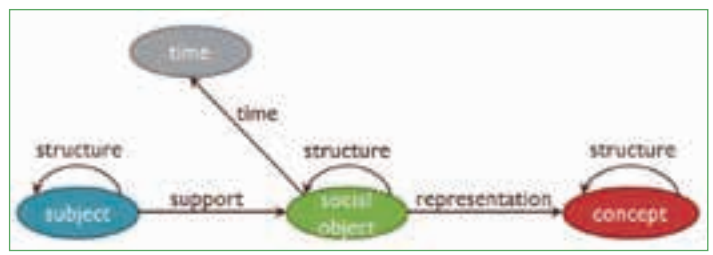
Luca Vignaroli - Supervisor: **Prof. Paolo Paolini -** Co-supervisor: **Prof. Letizia Tanca**

In recent years the way users watch television has radically changed. Also the home environment is changing since many smart users watch television while using a portable PC or a tablet as secondary screen more or less related to the broadcast programmes.
At the same time, social networks allow the final user to be immersed in a collaborative environment and to talk about television. TV users' social activities implicitly make connections between concepts by means of videos, news, comments, and posts. The strength of such connections may change as the perception of users on the Web changes over time. Moreover user-generated contents (UGC) are revolutionizing all phases of the content production value chain , in particular it can be observed that a very large number of UGCs include significant portions of content already broadcast by the TV networks.
In this context a number of Social TV applications are emerging, providing to the final user tools for social interaction while watching television or media content related to a particular TV program. If properly leveraged, these collaborative social environments can be seen as rich information data sources, indirectly returning to broadcasters and content producers some form of implicit feedback from the final users.
In this thesis I define a Framework for the integration of the heterogeneous and dynamic data coming from different knowledge sources (broadcasters' archives, online newspapers, blogs, web encyclopaedias, social media platforms, social networks, etc.). The framework uses a knowledge graph to model all the heterogeneous aspects of the information in homogeneous way. I instantiate it in the context of the investigations: the integration of the cultures of TV and Web defining a model for the integration of the heterogeneous data coming from the knowledge sources (broadcasters archives, EPGs, collected audience data, social networks, etc.) which play a role in what I call the "second life" of TV content, starting from its production phase, going through the on-air phase, and continuing with the on-line phase, during this phase the television content turns to be a "magnet" for users in the network attracting users, and it becomes a "Social Object".
Through a prototype called MeSoOnTV, the Media and Social-driven Ontology-based TV knowledge management system, that enables the integration of the heterogeneous data coming from multiple knowledge sources. Finally, a demonstrator has been developed in order to show some possible features offered by the described framework based on a meaningful dataset.
Figure 1 presents an overview of the integration framework. It consists of three main layers: a source processing layer, a knowledge graph layer and a knowledge query and analysis layer. The source processing layer has the role of collecting all the data which will be conveyed in the model. It accesses number of predefined web/social/ media sources and processes them in order to extract those information units which will be represented as nodes in the knowledge graph, as well as those information that support the existence of relationships (modelled as edges in the graph) among them.
The knowledge graph layer manages the knowledge graph, which is the core of the proposal. The graph contains essentially three types of nodes: social objects, subjects and concepts, and all social, representation and structural interactions among them.
The knowledge query and analysis layer consists in a set of components for querying, browsing and analyzing the knowledge graph. A query module extracts subgraphs from the knowledge graph based on user's requirements and constraints. Each extracted subgraph can be seen as a "view" over the complete knowledge graph, only containing nodes and edges potentially relevant to the user query. An analysis module, provides a set of analysis and data mining tool to obtain models and patterns from the knowledge graph. It can act directly on the knowledge graph, or it can handle the subgraphs extracted from the query module also in terms of matrices or tensors.
The core of the framework is the knowledge network. In particular, I'm interested in capturing the dynamic evolution in time of the graph by using temporal nodes associated to social objects and describing their lifecycle.
In order to summarize the final definition of the graph (figure 2) a brief description of main node types is given, the knowledge graph represents the result of public actions of users in social environments. Nodes represents Items. Edges represents relationships among Items. The knowledge graph has three main entities (node types): "Subjects" represent Users that act in some way, "Social Objects" that are the result of public acts and "Concepts" that are physical and ideal objects referred by



1. overview of the integration framework

subjects via their public actions. The possible relationships between differed items could be: a group of subjects that recognize a social value of an act supports the resulting social object, a social object represents a social instance of some concepts on a precise context and structural relationships express «part of» links between entities of the same type.
Social objects can evolve in time. I consider temporal representation of a social object towards a special type of concept called time concept. Each edge of the knowledge graph can be weighted, the weight expresses the strength of the relationship, together with the structure of the graph, they are used in the analysis phase to calculate various metrics, such as similarity between entities, the set of entities to consider.
Finally I introduce the demonstrator built to show some possible features of the framework and some possible wais to exploit the model on which the framework is based (figure 3). It enables the integration of the heterogeneous data coming from multiple knowledge sources. The system captures multiple aspects of the domain, from the semantic characterization of the TV content, to the social characterization and the social perception of a TV event, to the temporal evolution of the social perception. The demo scenario relies on real data gathered from YouTube and Twitter, related to several Italian TV talk shows on politics, broadcasted by RAI and other Italian operators.



2. the knowledge graph definition

3. a screen shoot of the demonstrator