



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 170 in total. The students are subject to an examination every year to evaluate the progress achieved in their research and course work.

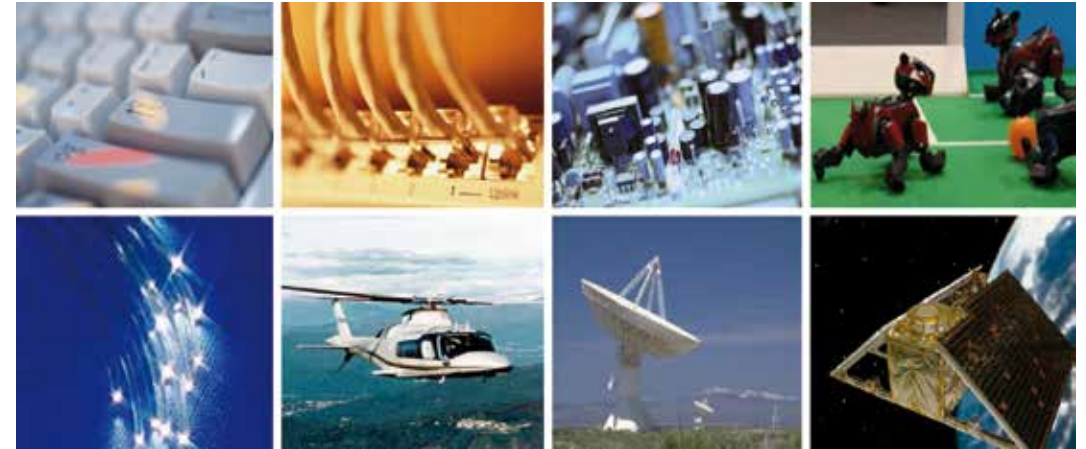
### Topics

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. Marco Lovera): 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. In 2008, the PhD "Alumni Association" was started, that organizes a successful yearly scientific and relational event (PHDAEY).

### 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, thanks to specific budget grants available in the Department, at least 10 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-supervision of a doctoral Thesis with Ecole

Polytechnique 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.

## ADVISORY BOARD

Livio Baldi (Numonyx S.p.a.)

Claudio Bertolini (HP Labs - USA)

Riccardo De Gaudenzi (European Space Agency)

Stefano De Panfilis (Engineering Ingegneria Informatica S.p.A.)

Giuseppe Fogliazza (MCM Spa)

Corrado Leita (Symantec Research Labs Europe)

Stefano Signani (Unicredit S.p.A.)

Maurizio Zuffada (ST Microelectronics S.r.l.)

### Prizes and awards

In 2013 the following awards have been obtained by PhD students:

**Chorafas Foundation Award:** Antonio Filieri, Giovanni Marzin

**Best Ph.D. in Information Technology Award:** Matteo Oldoni

**IBM Ph.D. Fellowship Award:** Valerio Panzica La Manna

# LOW-NOISE, LOW-POWER FRONT-END ASICS FOR HIGH-RESOLUTION X-RAY SPECTROSCOPY AND IMAGING IN SPACE WITH SILICON DRIFT DETECTORS

Mahdi Ahangarianabhari - Supervisor: Giuseppe Bertuccio

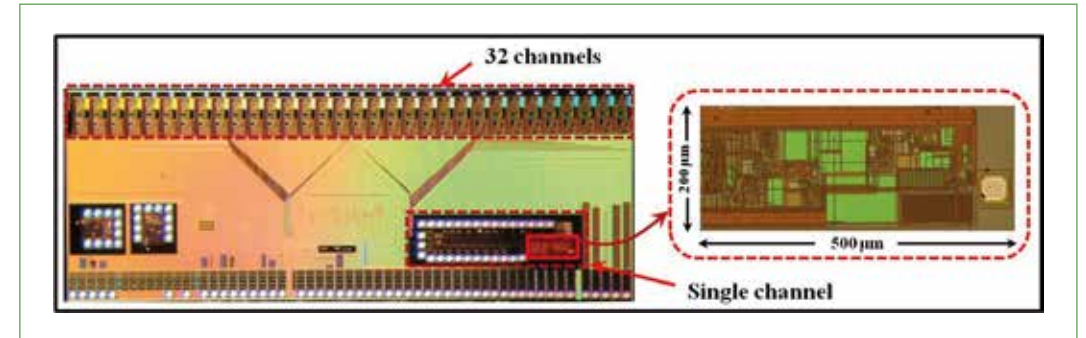
The work presented here has been carried out within a scientific collaboration between Politecnico di Milano (this doctoral thesis), University of Pavia, INFN-Trieste, INFN-IASF Bologna and INAF-IASF Rome, related to the assessment phase of LOFT (Large Observatory For X-ray Timing) mission. The LOFT mission is one of the four cosmic missions selected by European Space Agency (ESA) as a medium class space mission for high resolution X-ray observation of compact objects and ultra-dense matters in space. LOFT scientific consortium includes 141 Institutes in 22 Countries. This doctoral thesis is focused on the design, simulation and experimental characterization of VEGA: the low-noise, low-power CMOS ASIC (Application Specific Integrated Circuit) designed for processing signals from the X-ray Detectors on board of LOFT. VEGA is specifically optimized to readout signals from large area monolithic Silicon Drift Detectors (SDDs) constituting the Large Area Detector (LAD) of LOFT, devoted to high resolution X-ray spectroscopy and imaging in energy range from 500 eV to 60 keV. Each SDDs has an active area of 76 cm<sup>2</sup>, an anode capacitance of 350 fF, an anode current ranging from 0.7 pA to 10 pA. The most challenging task of this project is to design a very



1. LOFT: Large Observatory For X-ray Timing, one the cosmic missions is selected by ESA as a medium class space mission.

compact ADC-ready ASIC with a total power consumption of less than 500  $\mu\text{W}/\text{ch}$  and a noise level lower than 20 electrons r.m.s. in order to achieve an energy resolution better than 260 eV FWHM at 6 keV for the full system. The VEGA ASIC is comprised by an analog, a mixed-signal and a compact logic circuit to accomplish all the required functionalities for high resolution X-ray spectroscopy as required by the LOFT mission. The analog chain includes a very low noise charge sensitive preamplifier followed by a first order semi-Gaussian shaper and

a peak stretcher/sample and hold circuit. The mixed-signal section includes an amplitude and a peak discriminator circuits. The compact logic circuit is implemented in order to control the functionalities of the analog and the mixed-signal sections. Various functionalities are implemented in the VEGA ASIC including: 3-bit digitally selectable shaping time, 4-bit digitally selectable fine threshold, pile-up rejection, preamplifiers and discriminators disabling. VEGA is designed, manufactured and tested in single channel and 32 channels versions. The VEGA front-end channel has a total area occupation of 200  $\mu\text{m} \times 500 \mu\text{m}$  with a total power consumption of 480  $\mu\text{W}$  for the single channel and 420  $\mu\text{W}$  for the multi-channels versions. A minimum intrinsic Equivalent Noise Charge (ENC) of 12 electrons r.m.s., equivalent to 104 eV FWHM pulser line width is measured at 3.6  $\mu\text{s}$  shaping time at +27°C for the FEE circuit without detector. A minimum pulser line width of 140 eV FWHM, corresponding to ENC of 16 electrons r.m.s., is measured at 3  $\mu\text{s}$  shaping time at -30°C with <sup>55</sup>Fe radiation source for the single channel bonded to a 10 mm<sup>2</sup> cylindrical Silicon Drift Detector. The measured line width of <sup>55</sup>Fe Mn K $\alpha$  is 185 eV FWHM, well



2. Layout of the VEGA chip: including the single channel and the 32 channel versions. The Front End Electronics channel has dimensions of 200  $\mu\text{m} \times 500 \mu\text{m}$ .

below the 260 eV required by the specifications of the LOFT mission. The multi-channels version of VEGA has a minimum intrinsic ENC of 13  $\pm$  1 electrons r.m.s. at room temperature. The multi-channel VEGA bonded to the first prototype of the LOFT SDD with the active area of

about 76 cm<sup>2</sup> has shown 202 eV FWHM energy resolution on Mn K $\alpha$  energy line of <sup>55</sup>Fe. VEGA with its 12 electrons r.m.s. ENC (104 eV FWHM pulser line width) and its power consumption of 420  $\mu\text{W}/\text{ch}$  - 4 times lower than any other ever published equivalent ASIC

- opens a new state of the art in low-noise low-power CMOS integrated circuits for large area X-ray semiconductor detectors.

Table 1. The fully CMOS ADC-ready VEGA ASIC features.

Technology	AMS 0.35 $\mu\text{m}$ CMOS
Input charge	Electrons
Dynamic range	500 eV - 60 keV (Si-detectors)
Peaking times	1.6 $\mu\text{s}$ to 6.6 $\mu\text{s}$ (3 bits selectable)
ENC (intrinsic, T = +27°C)	12 electrons r.m.s. (without detector)
ENC (T = -30°C)	16 electrons r.m.s. (with 10 mm <sup>2</sup> detector)
Energy resolution ( <sup>55</sup> Fe Mn K $\alpha$ , T = -30°C)	185 eV FWHM (single channel, 10 mm <sup>2</sup> SDD) 202 eV FWHM (32 channels, LOFT LAD)
Power consumption	482 $\mu\text{W}$ (single channel version) 420 $\mu\text{W}/\text{ch}$ (32 channels version)



# MANAGEMENT OF MULTI-PURPOSE RESERVOIRS UNDER CLIMATE CHANGE: IMPACT ASSESSMENT AND ADAPTATION STRATEGIES

**Daniela Anghileri** - Supervisor: **Prof. Rodolfo Soncini-Sessa**

Climate change is considered as one of the major forces that will affect water availability in the future. Identifying possible response strategies is a complex task because future projections are deeply uncertain and hydro-climatic conditions are expected to evolve gradually in the next decades. Current water management practices may not be robust enough to cope with the climate change impacts on water supply, flood control, agriculture, energy and ecosystems. Methods and tools to assist water resource planners and decision makers are thus required. In this thesis we assessed how current and novel, adaptive approaches to water resource management can be used to cope with uncertain and nonstationary hydro-climatic conditions. We discussed the two mainstream approaches to assess climate change impact and to design adaptation strategies, i.e., the scenario-based approach and the vulnerability-based approach. We explored more in depth the vulnerability-based approach, which, in our opinion, is the most promising way to tackle the problems related to the sustainability of the water uses. To this end, we presented a set of simulation and optimization models, referring to System Analysis, that can be used to describe

the vulnerabilities and strengths of the water systems and to identify the room for improving the efficiency of water resource management. In particular, we integrated simulation models that are usually employed to estimate the impact of climate change and to define adaptation measures with water-value models, decision models, and optimization techniques, which allow to describe stakeholder expectations on water uses and the behaviour of the water system managers. We demonstrated those modelling and optimization tools through the application to three real-world case studies. We used the case study of Lake Como (Italy) to analyse the current water system institutional framework, and we demonstrated that re-framing the institutional setting toward a more cooperative and flexible one can reduce water system vulnerability to dry spells. We then assessed the potential of the management strategies, which proved to be the most efficient ones in the present hydro-climatic conditions, to deal with future water availability as projected by climate models. As a result of multiple uncertainties in the impact assessment procedure, the exact quantification of climate change impacts on water resources was not fully

possible. Nonetheless, the analysis proved to be useful in understanding the threats to future reservoir operations. For example, it demonstrated that re-optimization of the reservoir management policies can only partially compensate for the losses induced by climate change in the case of the lake management, but that it can not reduce the losses for the hydropower reservoirs located in the upper part of the catchment. This difference is the result of the combination of the future streamflow features and the physical characteristics of the reservoirs. The analysis thus produced knowledge that is not only relevant for the specific case study, but it can be exported also to other contexts. We used the case study of Lake Maggiore (Italy) to analyse how hydrological processes has changed in the last century and if these changes have affected water resources. We demonstrated that the water system has an inherent buffering capacity, which allows to naturally compensate hydrological variability. We then explored the possibility of further enhancing the resilience of the water system to hydrological changes by acting on the reservoir management strategy. Specifically, we tested adaptive management approaches as viable adaptation

measures to climate change. Adaptive management requires decisions to be continuously revised to react to changes in the decision making context. It is based on the insight that uncertainty can be reduced via system monitoring by gaining information useful from the management point of view. We proposed two formulations of the adaptive management approach which differ because of the pieces of information gained by the system monitoring, one looking into past hydrological records, the second foreseeing future streamflow. For the case study of Lake Maggiore, we developed a stochastic recursive optimization approach. We modeled the statistical properties of the inflows to the lake by using probability distribution functions estimated from historical records. These stochastic hydrological models were updated every year to account for new available records and progressively discarding the old ones. These models were then used to recursively design the reservoir operation so to possibly adapt to hydrological nonstationary conditions. The performances of the reservoir operation did not allow to validate the usefulness of the adaptive approach. Our hypothesis is that, on the one hand, the modelling

tools we used to represent the hydrological dynamics are overly simplified and probably inadequate for the purpose of the analysis, on the other hand, the methodology should be improved to better describe the relationship between conflicting objectives and to consider different hydrological models updating schemes. We tested the second type of adaptation management using Model Predictive Control and forecasts of future streamflow to improve water resource management in the Oroville-Thermalito complex (California), a mixed rain-snow dominated water system. We compared the value of two different realistic long term streamflow forecasting models in improving reservoir operation for water supply. We showed that, although reliable streamflow forecasts can theoretically improve reservoir operation, the realistic forecasting models currently available are not able to provide the pieces of information needed for mitigating droughts, i.e., reliable estimates of water volume in the long term. The main contributions of the thesis can be summarized in the following points: i) we exploited simulation and optimization techniques to re-frame the institutional setting where reservoirs are operated,

demonstrating that a shift toward a more cooperative and flexible setting can increase the overall efficiency of water resource management and can improve the resilience to unforeseen events, ii) we proposed an impact assessment procedure to assess the ability of water resource management practices to compensate future water stresses as projected by climate models, iii) we presented tools to assess the inherent adaptation capacity of water system to hydro-climatic changes which, on the one hand, allow to gain a deeper knowledge of the water system characteristics, and, on the other hand, can drive the identification of the most promising adaptation measures to further enhance the adaptation potential of water systems; iv) we addressed the topic of trend detection in environmental time series combining novel and traditional tools in order to simultaneously tackle the issue of seasonality and interannual variability, which usually characterize natural processes; v) we used stochastic recursive control and model predictive control to test if adaptive management is a viable adaptation measure to climate change.

## DEVELOPMENT OF HIGH PERFORMANCE ELECTRONICS FOR TIME-CORRELATED SINGLE-PHOTON COUNTING SYSTEMS

**Sebastiano Antonioli** - Supervisor: **Ivan Rech**

In recent years a growing interest has arisen in noninvasive optical analysis, in particular photon-counting measurements have gained wide acceptance in several research fields, because of their extremely high sensitivity. Pushed by the outstanding performance achieved by state-of-the-art single-photon detectors, such as single-photon avalanche diodes (SPADs) and photomultiplier tubes (PMTs), the time-correlated single-photon counting (TCSPC) technique is one of the preferable solutions if fast and weak light signals have to be measured.

A typical TCSPC system measures the delay between the arrival time of a photon emitted from the analyzed sample and a reference pulse given by a laser stimulating the sample. After the detection of several photons, a histogram is built up. The shape of the histogram represents the probability distribution of the photon detection times and, under the assumption that one photon, at maximum, is emitted in each cycle, this distribution corresponds to the light signal emitted by the stimulated sample. Two classical TCSPC measurement block architectures are reported in literature. The first one is based on a time-to-amplitude converter (TAC) followed by an A/D converter (ADC). The TAC

generates a voltage proportional to the time delay between the photon arrival time and the laser reference pulse whereas the ADC digitally converts the TAC output voltage. An alternative configuration implements a time-to-digital converter (TDC) that generates the digital value of the measured time delay exploiting the transit time of the timing signal within a chain of logic gates.

Over the last years these two architectures have been implemented in high performance single-channel TCSPC systems, but modern applications are increasingly demanding for instruments capable of carrying out multiple TCSPC measurements at the same time or as a function of different parameters. Multichannel TCSPC systems currently available on the market achieve high performance in terms of time resolution, differential non-linearity (DNL) and count rate; nevertheless, their structure is mainly a parallelization of a single-channel architecture. This approach is critical since area occupation and power consumption increase linearly with the number of channels. Also multichannel systems that can integrate thousands of single-photon detectors on the same chip have been developed but, since the area occupation

and the power dissipation of the single acquisition chain must be very low, the reached performance is far from the state of the art.

Aim of this thesis work is to design an inherently multichannel TCSPC system able to break the trade-off between performance and number of channels.

The exploited measurement block is the TAC/ADC configuration since it achieves the best-in-literature performance in terms of DNL (a high DNL value can lead to a distorted reconstruction of the analyzed light signal). The ADC gives the main contribution to the total DNL of the system since A/D converters currently on the market have a typical DNL value which is an order of magnitude higher than the value required for a TCSPC measurement (few percents of least significant bit - LSB). To obtain a DNL value that fits TCSPC requirements the dithering technique is implemented. For this reason, a random signal, generated through a D/A converter (DAC), is added to the TAC output voltage and is subsequently subtracted from the digital value resulting from the ADC conversion. This way a fixed TAC measurement is acquired using a random set of ADC channels, thus averaging the errors on the width of those channels.

During this work, several TCSPC systems have been realized; the first one allows to acquire a single TCSPC measurement: it exploits the TAC realized in another thesis work, a single channel ADC to sample the TAC measurement, an FPGA to build up the histogram and a USB transceiver to export the data. This board has been realized to validate the architecture of a TCSPC system including the TAC, together with the dithering technique implemented with a commercial DAC, and to characterize the converter itself in order to support the development of a new multichannel TAC. The achieved performance is very high, comparable with state-of-the-art instruments, in particular in terms of high conversion rate (up to 4 MHz), high time resolution ( $< 50$  ps FWHM, with the full-scale range fixed at 45 ns) and low DNL ( $< 4\%$  peak-peak of LSB over almost the entire range). Moreover, with this system, two important goals have been reached: very small dimensions ( $95 \times 40$  mm<sup>2</sup>) and very low power consumption (2.5 W). These characteristics have allowed to start the development of larger TCSPC systems.

Indeed, in the second designed board, eight acquisition channels have been implemented keeping the same small area occupation



1. Picture of the 32-channel TCSPC module.

of the single-channel with only 6 W of power consumption. Two 4-channel TAC arrays are employed in this system, each of them including four converters with variable full-scale range, a DAC that implements the dithering technique with area occupation as small as possible and four adder stages, used to adapt each converter output to the input dynamic range of a commercial 8-channel ADC. High performance has been obtained also with this system, in particular: very high time resolution (down to 20 ps FWHM selecting the 11 ns FSR), low DNL (6% peak-peak of LSB), high conversion rate (up to 5 MHz per channel) and low crosstalk among the channels (lower than 6% of the time bin width in the worst case). Finally, four 8-channel boards have been parallelized in a compact module to realize

a complete system able to manage 32 independent TCSPC channels with performance comparable with the state-of-the-art instruments, a total power consumption lower than 30 W and overall size equal to  $160 \times 125 \times 30$  mm<sup>3</sup>. Moreover, this module has been designed to receive the 32 photon-timing signals from a 32-channel time-resolved single-photon detection system, realized in a parallel thesis work. This detection head includes a  $32 \times 1$  SPAD linear array, developed in a custom technology, that achieves high photon detection efficiency, good time resolution, low dark count rate (if cooled) and low afterpulsing probability. Hence, the complete TCSPC system realized with these two modules features very high performance and it has been designed to fit the specifications of modern TCSPC applications.

## A METHODOLOGY TO IMPROVE THE ENERGY EFFICIENCY OF SOFTWARE

**Marco Bessi** - Supervisor: **Prof. Chiara Francalanci**

In the last ten years IT systems have grown very fast, with a consequent growth of the IT energy consumptions. This growth has raised a number of considerable issues. First of all, energy costs have dramatically increased and their impact on the overall IT infrastructural costs has become even more significant. Second, energy requirements represent one of the scalability issues of datacenters, since providers have difficulties at supplying data centers with all the required energy. Moreover, information technology contributes strongly to the green-house CO2 emissions. All these issues belong to a new field of study: Green IT.

Although software does not directly consume energy, it affects the energy consumption of IT equipment. Software applications indicate how information should be elaborated and to some extent guide the use of hardware. Consequently, software is indirectly responsible of energy consumption. Despite this, software engineering focuses on software performances and quality. Energy efficiency is not even included among software quality metrics. Most enterprises focus on the trade-off between cost and quality, while neglecting software energy efficiency. Therefore there is a

need to analyze, quantify and optimize application software from the point-of-view of the energy efficiency. The literature explains how optimizing software algorithms has a direct and potentially beneficial impact on energy efficiency. However, there is broad evidence of the practical hurdles involved in optimizing software algorithms, tied to the technical skills of programmers, to the role of domain knowledge, and to the need for massive code refactoring and related costs. This thesis focus on software energy consumption with the goal of providing a methodology to identify and reduce energy inefficiencies at limited costs.

The energy consumption of IT has attracted the attention of both academic and industrial communities in the last years. Green IT raises research challenges at different infrastructure levels, including the selection of hardware devices, data centers design, management and usage practices, and software design and development. In our previous researches is shown the importance of software energy efficiency, as the ultimate cause for the energy consumption of all other infrastructural levels. IT managers seem to believe that software per se has a very limited impact on the energy

consumption of real-word systems. Their claim is that although the applications have different designs, they do not differ significantly from one another in terms of hardware requirements and, thus, energy efficiency. It is generally accepted that the operating system can make the difference in energy consumption, not the application software itself. However, as a matter of fact these beliefs have never been empirically verified using a systematic scientific approach and there is no clear evidence as to whether application software has a tangible impact on energy consumption.

We propose a methodology to estimate the energy consumption of application software independent of the infrastructure on which the software is deployed. Our methodology makes a distinction between the usage of computational resources by an application and the unit energy consumption of computational resources. In this way, usage of computational resources and unit energy consumption can be measured independently. The usage of computational resources includes the usage of processor and input/output (I/O) channels to execute transactions. Both measures can be easily obtained with widely used testing platforms. Unit energy

consumption can be measured just once for each class of devices or extracted from their datasheets. Our methodology can support the definition of benchmarks of energy consumption for classes of functionally similar transactions or applications. These values can be used to define classes of efficiency within a given sample of applications or to compare the energy performance of similar applications. These benchmarks are used to identify energy inefficient software applications. In this thesis we present a methodology that applies memoization techniques to improve the energy efficiency of inefficient application software. Memoization is a programming technique that caches the results of a software program in memory. Memoization increases energy efficiency when storing and fetching a value consumes less energy than executing the corresponding software program. In general, processing a complex computation-intensive function requires more energy than caching results and reading them when needed, especially if the function is frequently called with the same input values. Furthermore, these benefits can be reaped without a need for code inspection and redesign, as they are the result of an intelligent management of memory resources at run time. Memoization has been extensively studied by the literature and has been applied in different contexts, including dynamic programming and incremental computation. However, applying memoization to improve the energy efficiency of application software raises

new research challenges. First, a restrictive definition of the concept of pure function is required to avoid manual inspection of code and enable large scale applicability. Pure functions are good candidates for memoization. Second, the benefits of memoization depend on the precision of results: the greater the precision, the larger the size of memory required to store all possible results of a function. While this issue is neglected in existing literature, in the context of application software precision represents a typical item of a Service Level Agreement (SLA) and can be tuned at run time. Our methodology obtains an energy-efficient management of memory resources in three steps: 1) pure functions are identified as good candidates for memoization, 2) a subset of initial candidate functions is selected to make sure that each selected function can improve energy efficiency when memoized, 3) available memory is allocated to different functions in order to minimize overall energy consumption based on a set of SLAs. Note that the second step includes an empirical performance model supporting the runtime estimate of the energy benefits from the memoization of each candidate function. The model includes two important parameters: a) the statistical distribution of the input values of each candidate function, b) the impact precision requirements on the amount of memory required for the memoization of the function. Both parameters are strongly affected by the application context and represent important determinants of the overall

feasibility of the methodology. We have developed a software suite to test our approach. The methodology is tested on a set of well-known financial functions. The experimental campaign is executed with automatically generated input workloads based on real parameters, including the range, mean, and standard deviation of input parameters. Both the code and the workload parameters have been provided by a large Italian corporation that have accepted to participate in the empirical testing of our methodology as a pilot case set. Empirical results show how our approach can provide significant energy and time savings at limited costs, with a considerably positive economic impact.

## REVERSE ENGINEERING OF VIDEO CONTENT FOR FORENSIC ANALYSES

Paolo Bestagini - Supervisor: Prof. Augusto Sarti

The recent development of multimedia devices and editing tools, together with the proliferation of video sharing web sites, has made the acquisition, alteration, and diffusion of video content relatively easy tasks. As a consequence, we find more and more video sequences available on the Internet, but each of them is potentially tampered with by anyone. It is then clear that the development of tools that enable the recovery of past history of video sequences in order to prove their origin and authenticity is more than an urgent necessity. In the last few years, many forensic techniques have been proposed to detect malicious modifications of multimedia data. However, many solutions are specifically tailored to the still image case. Video solutions have been proposed only more recently. For this reason, in this thesis, we focus on the challenging development of video forensic algorithms capable of detecting specific information about the past history of a video sequence, operating in a complete blind fashion. This means that digital content does not need to be signed beforehand, and the analysis is completely conducted having at our disposal only the final object, with no priors about its past.

This is possible due to the fact that any non-invertible operation leaves peculiar traces on the video. Therefore, coding, editing, and other operations are characterized by the footprints they leave behind. The analysis and detection of such footprints allow to reveal if a video has undergone a specific operation.

The goal of this thesis is to develop a series of detectors acting in a strictly blind fashion to infer pieces of information related to the past history of video objects. In particular we aim to propose algorithms that work in real world scenarios, rather than proposing a mathematical analysis of simplified models. This means that some methods may be based on empirical observations, however they are always validated by means of experiments on real video sequences.

The detectors we propose exploit traces left by non-invertible operations on video sequences. In doing so they can operate in a blind way on material that has not been signed or fingerprinted. The most common non-invertible operation that is applied to video sequences is coding. Indeed, it is very unpractical to save uncompressed sequence even at a modest resolution

due to the huge amount of space needed, and for this reason videos are always almost distributed as compressed. It is then natural to exploit traces left by codecs to reveal part of the history of a sequence. To this purpose, in this thesis we aim to answer a basic question: how many times a video has been encoded during its lifetime? Answering this question is of paramount importance. Indeed, the more the number of coding steps, the higher the probability that the sequence has been edited. Therefore it cannot be considered authentic. In order to detect how many times a video has been compressed, we developed an algorithm based on a statistical law, called Benford's law. More specifically, each compression step operates a quantisation operation on the video content (typically in the DCT domain). By studying the empirical distribution of the most significant digit of quantised DCT coefficients, it is possible to notice differences between videos encoded different number of times.

Another question we aim to answer is: which codecs have been used during the multiple compression steps? Answering this question may pave the way to the detection of the acquisition device. Indeed,

usually the first compression step is operated directly on the originating device. Since different vendors often make use of proprietary codec implementations, detecting a specific codec implementation is then characteristic of a vendor or of a device.

To this purpose, if we detect that a video has been encoded twice, we developed an algorithm capable of detecting the first codec used in this codec chain. This is done exploiting the idempotency property of video codecs. That is, when a video sequence is re-encoded using the same coding architecture and coding parameters, the input and output sequences are alike. To this purpose, we re-encode the double-coded sequence with different codecs and parameters and show how to detect when the correct codec is found.

In addition to coding footprints, also editing operations may leave peculiar traces. One of these operations we aim to detect in this thesis is frame temporal interpolation. This operation can be used for different purposes. An example is the fake frame rate fraud. That is pretending that a video has been shot with a high frame rate camera, while it has only been interpolated to achieve a higher frame rate. Another scenario is the following. Let us consider the case of two video sequences acquired with different frame rates (e.g., 25 and 30 frames per second). In order to splice them together, the frame rate must be equalized (i.e., both sequences must share the same frame rate). To do that, a frame

interpolation algorithm is used. This scenario might seem not so relevant. However, let us consider the case a malicious user wants to insert an object from the first video into the second one. It is clear that the portion of the first video copied and pasted into the second one must be temporally interpolated. Therefore, being able to locally (in time and space) detect frame interpolation is a way to detect possible tampering. The algorithm that we propose is based on a detector studied to detect spatial interpolation on still images. However, it is adapted to work in the time domain, also considering the case of motion-compensated frame interpolation (which is much more realistic).

Object insertion or removal from a scene can be operated in other different ways. In particular, we consider two additional scenarios. The first one is that of taking a picture, and repeating it in time covering a spatial region of some frames. We refer to this kind of attack as image copy-paste. The second one consists in repeating a spatial region from the same video (e.g., part of the background) to hide objects or persons at another time location. We refer to this kind of attack as video copy-move. To detect the first kind of forgery, we propose an algorithm that detects peculiar traces left in the frame residuals (computed as differences between adjacent frames) when images are pasted over time. This trace is clearly masked if the forged sequence is compressed after the tampering, however the algorithm still works up to a certain introduced coding

distortion. The second kind of attack (the one based on video rather than image copy-move) is detected exploiting an analysis of frame correlations. The detector is able to find if a time-space sub-region of a video has been replicated, and indicates the position of the forgery.

Even though the algorithms presented to this point are robust to compression, some operations can mask the presence of the forgeries and fool the detectors. These operations are known as anti-forensic operations. An easy yet effective anti-forensic algorithm is video re-capture. That is, a forged sequence is displayed on a monitor and re-captured using a camcorder. The re-captured sequence is visually similar to the displayed one, but many forgeries traces are deleted, due to the digital-to analog/analog-to-digital chain of operations. In order to tackle this scenario, we developed an algorithm able to detect video re-capture. This algorithm is based on the analysis of a peculiar ghosting artefact that characterise re-capture from LCD monitors. If a sequence is detected as re-captured, it clearly cannot be considered original.



# METHODS AND APPLICATIONS OF DISTRIBUTED AND DECENTRALIZED MODEL PREDICTIVE CONTROL

**Giulio Betti** - Supervisor: **Prof. Riccardo Scattolini**

In recent years, process plants, electrical, communication and traffic networks and manufacturing systems have been characterized by an increasing complexity. Usually, they result to be composed by a huge number of relatively small or medium-scale subsystems interacting via inputs, states or outputs. The design of a centralized controller for the whole large-scale system is often a difficult challenge due to possible limitations, for instance, in computing capabilities or communication bandwidth. Reliability and robustness of the overall control system could represent additional reasons for avoiding centralized controllers. According to these motivations, remarkable efforts have been put in the research field of hierarchical, decentralized and distributed control. Among all the possible solutions, particularly interesting appear to be those based on Model Predictive Control (MPC). Several reasons can be listed to support this claim: first of all, MPC is a multivariable optimal control technique and, among the advanced control methods, it is probably the only one really adopted by the industrial world and able to cope with operational constraints on inputs, states and outputs. Secondly, in the last years fundamental properties

such as stability of the closed-loop systems and robustness with respect to several classes of external disturbances have been proved for many different MPC formulations. Lastly, when controlling large-scale systems in a distributed or decentralized framework, the values of inputs, states and outputs are explicitly computed over all the prediction horizon at each sampling time by a local controller designed with MPC and they can be used as information to be transmitted to other local controllers to coordinate their actions. This data transmission can greatly simplify the design of a distributed control system and can allow one to obtain performances close to those of a centralized controller. In this thesis, we present a number of distributed control algorithms based on the Distributed Predictive Control (DPC) approach. They are designed for linear systems subject to constraints on inputs and states, described in state space form and decomposable in several nonoverlapping subsystems. The main idea of DPC is that, at every sampling time, each subsystem transmits to its neighbors the reference trajectories of its inputs and states over all the prediction horizon. Moreover, by adding proper additional constraints to the MPC formulation, all

the local controllers are able to guarantee that the real values of their inputs and states lie in a specified invariant neighborhood of the corresponding reference trajectories. In this way, each subsystem has to solve an MPC problem where the reference trajectories received from the other controllers represent a disturbance known over all the prediction horizon, while the differences between the reference and the real values of its neighbors' inputs and states can be seen as an unknown bounded disturbance to be rejected. To this end, a robust tube-based MPC formulation is adopted and implemented using the theory of polytopic invariant sets. This thesis also presents a completely decentralized version of DPC, named DePC (Decentralized Predictive Control) for strongly decoupled systems. In this case, no transmission of information is required, as all the interactions via inputs and states are seen as unknown disturbances to be rejected. The thesis is structured as follows. Chapter 1 provides an overview of the most common centralized MPC algorithms for regulation, tracking and disturbance rejection, as well as a short description of the main solutions for distributed and decentralized control, with particular attention to those

based on MPC.

In Chapter 2, the basic formulation of DPC for regulation of discrete-time systems is presented. Efficient techniques for the computation of polytopic robust invariant sets, for the initialization of reference trajectories and for the online solutions to possible large and unexpected disturbances are also shown, in order to provide both the theoretical results and some practical hints useful for real industrial applications. Since the discrete-time domain does not allow to consider the process inter-sampling behavior in the MPC optimization problem, in Chapter 3 a continuous-time version of DPC is illustrated. Chapters 4 and 5 present two extensions of DPC for the solution of the tracking problem: the first one can be used to track piecewise constant targets. The main improvement with respect to the standard DPC formulation consists in including among the optimization variables of the MPC problem the value of the reference point that each subsystem really tracks at each sampling time. An additional term in the cost function is also considered to penalize the distance of this extra optimization variable from the desired external set point. The second solution is based on inserting an integral action in the closed-loop by

rewriting the initial system in the so called "velocity-form". This formulation guarantees rejection of constant disturbances and can be used to track constant targets: in fact, recursive feasibility cannot be proved if the value of the reference signal is changed. In Chapter 6, to overcome the limitation of being able to manage only constant references for systems in velocity-form, a thorough study of their properties for centralized control is described for both the case of nominal and disturbed systems. The obtained results allow one to use the velocity-form also in presence of varying set points. Once again, time-varying external references are handled adding as optimization variable the set point really tracked at each time instant. The cost function term weighting its distance from the external one guarantees an asymptotic convergence of the first one to the latter. The outcomes of the previous developments for centralized control of systems in velocity-form have been exploited to develop a decentralized control method for tracking, called DePC, and presented in Chapter 7. It is able to manage piecewise constant set points and to reject external disturbances, asymptotically deleting their effects in case they are constant.

Eventually, some conclusions are drawn in Chapter 8.



## TEXTUAL AND CONTENT-BASED SEARCH IN SOFTWARE MODEL REPOSITORIES

**Bojana Bislimovska** - Supervisor: **Piero Fraternali**

The increased complexity and pervasiveness of software requires raising the level of abstraction, and automating labor-intensive and error-prone tasks to increase efficiency and effectiveness in software development. An approach that advocates software abstraction through the use of models is Model-Driven Engineering (MDE), widely used in academia and industrial organizations across different domains. MDE promotes the use of models in any engineering activity as abstractions that provide a simplified or partial representation of reality, useful to accomplish a task or to reach an agreement on a topic. Model-Driven Software Engineering specifically considers software models, i.e., abstractions of the static or dynamic properties of a software system. Studies demonstrate that the benefits of MDE in industry are perceived in terms of quickly responding to change of requirements, of streamlining communication among stakeholders thanks to more accessible organizational knowledge, and of improving the quality of code design and test case development. The adoption of MDE in academic and business organizations resulted in an increasing number of model collections, stored in model repositories.

Reuse and sharing of software requires the ability of effectively retrieving artifacts that meet the user's need, which is the goal of software search systems. Source code search is used not only within organizations, but also through online tools for share and retrieval of source code. Source code search allows for reduction of development time and costs, and may improve software quality, as novice programmers can learn from the code produced by more experienced ones. Model repositories are not yet as well developed and widespread as source code repositories. The same advantages of source code search could be achieved in MDE, if repositories allowed for efficient retrieval of models relevant to the user's needs. Model search approaches should exploit in more depth the main difference between source code and models, that is, the high level, structural, and often visual nature of a model representation. Ideally, an MDE developer should be able not only to search models via keywords, but also to sketch the idea he has in mind in his favorite language and retrieve all models that contain a similar design, properly ranked according to their relevance to the query. Therefore, similarity search techniques are essential to allow developers' needs

be formulated in the same language in which solutions are expressed.

The goal of this thesis is to study the implications of building search systems for software models expressed according to the notation of a modeling language, so to increase the reuse of modeling artifacts and promote the discovery of existing design patterns and the application of modeling best practices from previous projects. We study two different scenarios of model search: keyword-based search, which proceeds in continuity with classical Information Retrieval approaches and source code search techniques; and content-based search, which introduces the query-by-example paradigm into model search by representing software models as graphs and performing graph matching. Keyword-based and content-based search techniques are extended with the injection of metamodel knowledge in the search process, to test its effect on retrieval performance. Three approaches (one text- and two content-based) are implemented, configured and technically evaluated on a real world collection of 341 industrial models, with a panel of 10 queries. Models in the experimental repository are encoded in the Web Modeling

Language (WebML), a Domain Specific Language (DSL) for Web applications. Furthermore, the same text-based approach was adapted and configured for assessing its performance on a publicly available repository of 84 UML models (class diagrams) with a set of 20 queries.

In the keyword-based search approach, state-of-the-art methods were extended in order to incorporate metamodel-specific information. We show that augmenting the IR index with metamodel knowledge leads to a performance improvement with respect to conventional, metamodel-agnostic text-based IR techniques.

The first content-based search approach is implemented by means of graph matching by extending standard techniques for sub-graph isomorphism (the A-star algorithm) by considering a formulation of the matching score function that takes into account metamodel-specific information. The thesis also investigates how the locality of the match between the query and the project graph affects performance. In order to overcome the limitations of the A-star algorithm with respect to efficiency and scalability, we propose and implement another content-based search approach that employs graph matching. It is a novel approach that uses approximate representation of graph nodes as points in multidimensional space, obtained by multidimensional scaling. These points are used to build an index that allows efficient and scalable search. The method also utilizes neighborhood information for

finding similar subgraphs of a project graph with respect to a query graph. Metamodel information is incorporated in the index, and it is considered while performing search, as well as in the scoring function used to rank the subgraphs.

Performance of the proposed text- and content-based search approaches with respect to retrieval accuracy (precision and recall), ranking accuracy, and stability of the results across different queries is assessed by a technical evaluation based on a gold standard defined by experts. Moreover, the text- and one of the content-based approaches (graph matching technique based on the A-star algorithm) applied on the WebML repository, are also evaluated with two user studies, which engaged 25 MDE practitioners in the subjective evaluation of utility and quality of search results. Different variants of the technical configurations of the two systems have been evaluated and compared. The user studies examine the relationship between the performance of the systems and the user-perceived utility of retrieved results. The results of the experimental evaluation have shown that traditional text indexing techniques can deliver good performance for keyword-based queries, while adding metamodel knowledge to the index can improve accuracy. For A-star graph-based search, the conclusion is that matching the textual content of projects is still important, but the system benefits from an appropriate injection of metamodel knowledge regarding both the types of the elements and

the structure and topology of models. Furthermore, A-star graph-based search results exhibited greater variability and dependency on the queries than keyword-based results. Considering the multidimensional scaling graph-based search, introduction of additional metamodel information improves performance and generates alternative matching patterns. Another important observation for multidimensional scaling graph search is that it has an increased efficiency with respect to the A-star graph-based search at the price of performance loss. The results of the user studies show that the A-star graph-based search system provides, on average, better results than the keyword-based system, which suggests a correlation between the performance of a retrieval system and the user-perceived utility for reuse. Moreover, the user studies provide an insight about what expert modelers consider suitable queries and responses and the way in which two different classes of information retrieval systems can be configured to respond to the expectations of these searchers.

## INNOVATIVE SOLUTIONS FOR HIGH-BIT RATE BACKPLANE-BASED DATACOM INTERCONNECTION

**Anna Boletti** - Supervisor: **Prof. Pierpaolo Boffi**

In the next years telecom operators plan to expand the range of innovative services to guarantee the telecom market growth: the High Definition television, the teleworking, the digital health care system, the digital public administration, the digital learning, the Internet gaming are only some of the considered high quality telecom services. Hence, in the next future the telecom network will be characterized by an always-increasing demand of bandwidth and transmission rate, which increase the need to have transmission equipments with always increasing information transmission capacity.

The state of the art in the field of the transmission equipment consists of electro-optical boards with external optical interfaces, interconnected by means of high-speed electrical backplanes in the same subrack.

The solutions currently used are based on copper interconnections via backplane, driven by a suitable transceivers subsystem. Looking at high-end telecommunication equipment transceivers are directly integrated on the same silicon chip used for the switching function or Input/Output (I/O) function. The transmitting and receiving interfaces, the connectors, and the Printed Board (PB) technology evolved and allowed us to increase the

single interface bandwidth.

In parallel the amplitudes of the transmitted signals has been reduced, to limit power consumption and increase the integration.

In general we can observe the progressive capacity to enhance the power density in a single device and while the consumption measured in mW per Gb/s has had an exponential decrease, the consumption of the interface including Serdes, clock data recovery (CDR), pre-emphasis and equalization is growing. Furthermore, complexity and density increase is supported by evolving silicon technology able to manage more W/mm<sup>2</sup>. The realization of the transmission lines in PB even if based on the most sophisticated materials available on the market, shows very significant losses at very high operation frequency. Also connectors, even if now very optimized, contribute to increase loss and crosstalk. The electrical backplane is close to reach its limit and for that reason we need to explore alternative solutions.

Due to the fact that optical technologies are able to provide higher capacity over longer distances than electrical transmission systems, a natural answer to the limitations shown by the copper-based interconnection is to exploit

optical solutions.

For optical interfaces the bandwidth density scaling necessary to face the future Ethernet requirements may be achieved through the reduction of the pitch size (presently to about 50µm,) and applying 2D arrays of Vertical Cavity Surface Emitting Lasers (VCSELs) and PINs. Optical fibers constitute a good trade off in terms of bending losses, coupling efficiency with VCSEL sources, costs and attenuation. Multimode multicore fiber and/or 2D polymer waveguides are being prototyped to match the demand. The reduction of the distance at increased bit data rates for the existing multimode fibers is an important trend and continuous development is offered. Last but not least novel approaches to increase bandwidth density and transmission length are being investigated, for example based on multi-level coding like Pulse Amplitude Modulation (PAM) and compact wavelength-division multiplexing systems realized in silicon photonics or InP-based photonic integrated circuit approaches. Furthermore, many studies are carrying out to better achieve the integration using silica or polymeric waveguides, instead of copper striplines.

The aim of this doctoral

dissertation is to understand and quantify the real benefits of an optical backplane from a point of view of cost and complexity of implementation, transmission performance and power consumption. In particular, it has been tried to identify solutions that ensure simplicity of implementation, and which exploit nowadays off-the-shelf technology.

The objective is to understand the real applicability of the proposed optical solutions and their potential for evolution to next capacity upgrade in data center applications.

At the beginning an overview of different levels and types (electrical or/and optical) interconnections used in network applications is presented. Limitations of the electrical and optical approaches are discussed, considering their capabilities to support more than 20-Gb/s transmission and the possibility to evolve towards higher bit-rates. Some possible alternative optical solutions, to integrate the source into the transmitter module on the boards, interconnected via backplane, are presented. Based on optical interconnection the exploitation of low cost, power efficiency source as Vertical Cavity Surface Emitting Laser (VCSEL) is also described.

Then, a comparison between an electrical and optical backplane is given. Based on the nowadays technology and through simulations, it has been possible to design the best achievable electrical backplane. Hence, a comparison with respect a possible optical fiber-based backplane, using

VCSEL as sources, in terms of power consumption and capacity has been carried out. A detailed analysis on different specifications (as scalability, low cost, low complexity and transparency) that a real optical backplane has to satisfy has been discussed. Some of relevant aspects, like material, choice of optical fibers, type of connectors, design and layout have been taken into account to allow a physical realization of a real fiber-based optical backplane.

Focusing on one of the backplane requirements, the transparency, further studies and experimental analysis have been made. Optical data links should guarantee a complete transmission transparency in the interconnection also with the external network, i.e. a traveling signal must be allowed to pass through the backplane network on a determined internal link without any need of opto-electronic and electro-optic conversion and processing. A center-launching technique between single mode optical fibers and multimode fibers has been exploited to satisfy that requirement. Experimental and simulative analyses have been carried out in order to test the achievable performances and robustness of this technique, making it suitable for fiber-based backplane connections. To reproduce different operating conditions in backplane applications, two experimental configurations have been taken into account and bit error rate curve have been obtained.

Finally the work focuses on how to increase the optical backplane transmission capacity,

maintaining the same number of optical connections and, in the meantime, satisfying stringent requirements like low cost, low complexity and low power consumption. Multi-level modulation format have been considered, in particular close attention is paid on Pulse Amplitude Modulation (PAM) and On-Off-Keying (OOK) modulation format. Beyond a common VCSEL-based technology, a possible alternative technology as Silicon Photonics have been analyzed, comparing them in terms of power consumption and performances through simulations results. The work concludes providing some considerations related to the best technology suitable for short and medium reach applications in the near future, taking into account power efficiency, cost, complexity and integration aspects.

## SINGLE-PHOTON COUNTING INSTRUMENTATION FOR GATED-MODE DETECTION

Gianluca Boso - Supervisor: Alberto Tosi

### Description:

The measurement of fast-changing and very faint light signals in the visible and near-infrared wavelength range with picosecond timing resolution has proven to be an effective technique to study different physical and biological structures like biological tissues in functional brain imaging, optical mammography and molecular imaging, not to mention Fluorescence Lifetime Imaging, Fluorescence Correlation Spectroscopy, Quantum Information, LIDAR and many others. Single-Photon Avalanche Diodes (SPADs) are some of the most used detectors to measure fast light pulses at single-photon level.

SPAD detectors are microelectronic single-photon detectors that are becoming increasingly widespread thanks to their reliability, low cost, ruggedness and possibility to be integrated into a standard CMOS process in order to obtain large arrays or to integrate complex logic structures together with the detector. Performance-wise, they have good Photon Detection Efficiency (PDE) (up to 50% in the visible range) and low timing jitter (few tens of picoseconds). Moreover, a SPAD can be gated from OFF to ON in few hundreds of picoseconds by modulating its bias voltage. Detector gating

is an effective way to keep the detector blind during intense photon fluxes, while enabling its single photon sensitivity only during desired gate-ON time windows. My major research topic consists in the development, characterization and experimental exploitation of a fast-gated SPAD detector module (see Fig. 1) intended to operate different Silicon SPADs (built in a CMOS or custom process, with different active areas and structures) in the fast-gated regime with ON and OFF transitions below 200 ps. Fast-gating a SPAD detector involves different electronic challenges both in the driving circuitry (the pulse generator) and in the read-out circuitry. These challenges are addressed and many circuit solutions are presented. In particular two fast-gated instruments have been constructed: (i) a first prototype which monolithically included all the electronic circuits and optical components to fast-gate a single photon detector; (ii) a second implementation which featured a flexible detection head (including the SPAD detector and its ancillary circuits) – control unit topology (including power supply, PC interface and synchronization circuits). This second implementation is based on a commercial platform as starting point and is aimed at

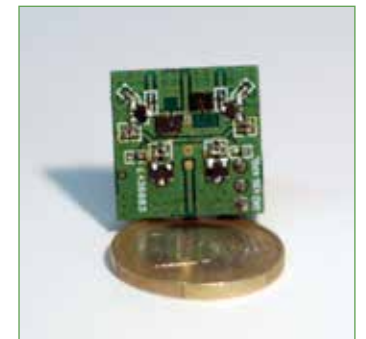
obtaining a more robust module with consistent performance. The various fast-gated modules have been exploited in different research collaborations. Different fast-gated prototypes were specifically tailored to obtain the best performance for each application. Such collaborations include Dipartimento di Fisica of Politecnico di Milano (Italy), CEA-LETI, Grenoble (France), PTB, Berlin (Germany), UC Irvine (USA), École Polytechnique de Montréal (Canada), Istituto Italiano di Tecnologia, Genova (Italy). During these collaborations, the developed fast-gated module helped enhancing the performance of different imaging techniques at single-photon level like functional NIR spectroscopy at null source-detector distance, non-contact scanning functional spectroscopy, tomographic reconstruction of solid inclusions in liquid phantoms, STED confocal microscopy and the study of Rabi oscillation in nitrogen isoelectronic vacancies. The various imaging scenarios are described and the improvement introduced by the fast-gating technique is highlighted. Different custom instruments have also been developed to enhance the performance at the detection level in fast-gated time correlated single-photon counting (TCSPC) setups

including a gated counter with two detection windows scalable down to 70 ps, a fast pulse generator with rise and fall times below 50 ps and different driver circuits to gate SPADs or to pulse laser diodes in the gain-switching regime with FWHM below 100 ps. In particular, different laser structures have been tested, including side-emitting and VCSEL diodes, with wavelength ranging from 600 nm up to 1550 nm. Every structure has been tested in pulsed mode, through a custom electronic circuit developed as part of this thesis work, and evaluated in terms of optical power range and stability, extinction ratio and pulse shape. The evolution of the fast-gated TCSPC setup led to the development of the first setup for contact time-resolved brain imaging where both a fast-gated CMOS detector and a picosecond pulsed VCSEL diode are placed in a compact probe (see Fig. 2) in contact with the patient head. This new approach is leading to different future prospective like dense arrays of sensors or cheap and disposable optical probes with improved photon collection efficiency with respect to actual bulky setups. As a proof-of-principle experiment, a first in-vivo measure has been performed in collaboration with Dipartimento di Fisica of Politecnico di Milano

where blood and hemoglobin variations in the motor cortex of a healthy subject have been monitored during a standard finger-tapping experiment. Employing the same electronic background used to fast-gate Silicon SPADs, a gigahertz sinusoidal gating setup for InGaAs/InP SPADs has been developed (see Fig. 3). Compared to Silicon detectors, InGaAs/InP SPADs have a good detection efficiency (up to 50%) at NIR wavelengths (up to 1700 nm) but suffer from high dark noise level and afterpulsing noise that usually limit their use for high throughput applications. During this PhD thesis a custom setup was developed. The setup, thanks to the gigahertz sinusoidal gating of these detectors, achieves good performances in terms of detection efficiency (up to 4%) and timing jitter (below 150 ps) and boosts the throughput of the system up to 100 Mcps (1000 times faster than standard InGaAs/InP detectors). The different gating solutions are described including the issues that arise in the avalanche signal read-out when a SPAD detector is enabled with a gigahertz sinusoidal bias.



1. Fast-gated SPAD module



1. Compact probe for contact time-resolved optical spectroscopy

## A METHODOLOGICAL FRAMEWORK TO UNDERSTAND AND LEVERAGE THE IMPACT OF CONTENT ON SOCIAL MEDIA INFLUENCE

Leonardo Bruni - Supervisor: Prof. Chiara Francalanci

Social media have a strong impact on the way users interact and share information. The process through which users create and share opinions on brands, products, and services, i.e., the electronic word-of-mouth (eWOM) is gaining increasing attention. In the online context, the eWOM has been transformed from a communication act that takes place in a private one-to-one context to a one-to-many complex interaction. The reach of information sharing through eWOM can be both broad and fast. Companies know that controlling the dynamics of information sharing is very difficult. This need for improving control is one of the reasons why there is a growing interest in understanding how social networks can affect the dynamics of user interaction and information sharing. Indeed, by identifying and convincing a small number of influential individuals, a viral campaign can reach a wide audience at a small cost.

Several previous studies have focused on the role of influencers, i.e., nodes with a central position in the network. In particular, microblogging platforms such as Twitter are the focus of a wide range of studies that aim at understanding how messages spread inside the social network and how

the role of the message author impacts on message reach. Microblogging networks are more and more used by companies as a communication medium for the promotion and engagement of customers. While on traditional media attention can focus not only on content, but also on the way a message is conveyed, on social media content plays a more central role. Content is even more central with microblogging, as the shortness of messages compels users to focus on the core of the information that they want to share. On Twitter, the standard size of a message limited to 140 characters is roughly the typical size of headlines and encourages users to produce content that are easy to consume. As a matter of fact, some of the influencers on social media are a "network phenomenon" and, clearly, is what they have said that has had an impact in the past and represents the reason for their current breadth of audience, not the other way round. Our claim is that despite the fact that the information shared by influencers has a broader reach, the content of messages plays a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. This thesis takes a first

step towards understanding the relationship between content and influence by addressing the following research questions: i) is content a driver of social media influence?; ii) what are the characteristics of content that help increase social media influence? These questions are rather unexplored by the current literature.

Even though the breadth of the audience was considered the first and foremost indicator of influence, the more recent literature has associated the complexity of the concept of influence with the diversity of content. Considering the successful case of Twitter, it has been observed that influencers are prominent social media users, but we cannot expect that the content that they share is bound to have high influence. These literature results provide the theoretical foundations to support the hypothesis that the content of messages plays a critical role and can be a determinant of the social influence of the message irrespective of the centrality of the message's author. The goal of the research presented in this thesis is to analyze the set of relationships existing between content of posts and influence on social media. Content is evaluated from both a per-post and per-author perspective.

Measuring influence is a complex task because there is no precise consensus as to what influence is and how to measure it. In the context of Twitter, some studies indicate that a person's influence is related to the propagation and repercussion of tweets. Previous literature also recognizes that fundamental goal of any social media user is to post content that is not only shared by many users but also shared frequently. Based on these considerations, the influence variable is defined in this thesis by considering the two most suggested attributes in literature, i.e., volumes and time dynamics.

The relationship among the content and the influence on social media is analyzed by considering two different yet complementary perspectives, such as those of per-post and per-author. One of the most commonly weaknesses in the literature is that analysis often relies on assumptions based on the literature on traditional media, such as television or radio. However, traditional media are based on broadcasting rather than communication, while social media are truly interactive. It is very common that influencers say something totally uninteresting and, as a consequence, they obtain little or no attention. On the contrary, if social media users

are interested in something, they typically show it by participating in the conversation with a variety of mechanisms and, most commonly, by sharing the content that they have liked. A content that has had an impact on a user's mind is shared. Thus, the analysis is performed on a per-post basis and by characterizing content in terms of sentiment, inclusion of multimedia and specificity. This focus at a post level is important to make a clear separation between influence and influencers and measure the impact of content without a bias from the content's author. Then, we make an effort to translate our post-level findings into guidelines to be adopted by authors to maximize their influence. This is performed in two steps. First, we describe an empirical analysis investigating the relationship between content and influence by approaching content measures at an author's level and, thus, searching for the rules that tie author-level behavioral variables and authors' influence. Second, we attempt the application of these rules in a practical case to create and promote a brand. To support both steps we have developed and continuously improved a software suite that supports assessment.



# ADVANCED DIGITAL TECHNIQUES AND ELECTRONIC ARCHITECTURES FOR HIGH PERFORMANCE SCIENTIFIC MEASUREMENTS

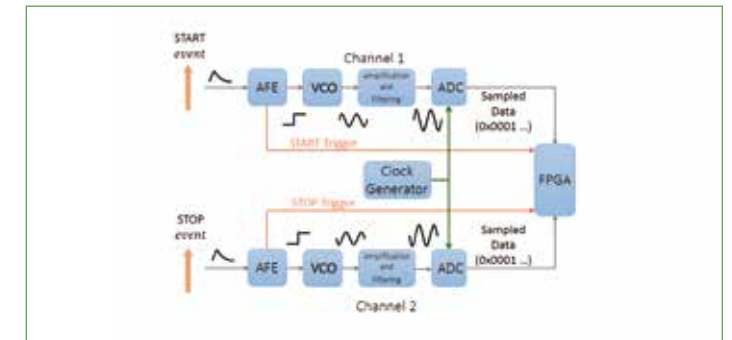
Francesco Caponio - Supervisor: Prof. Giancarlo Ripamonti

The precise measurement of time intervals is a primary goal in a growing number of applications and the challenge to achieve increasingly higher resolutions than ever is a main topic of research. In this sense, time-of-flight measurements and time-correlated photon counting are two milestones. Since the intrinsic resolution of the sensors used today is in the order of tens of picoseconds, the measurement systems must guarantee performance at least of this order. The choice of making digital a part or the totality of the measurement electronic systems exploits well-known advantages from the adaptivity, to the versatile calibration, to the easiness of implementation of powerful processing algorithms, but also new methods must be put in place to significantly improve the resolution of the measurements. The well-established method for high-accuracy time interval measurements is the time-to-amplitude conversion (TAC). Another diffused method is the time-to-digital conversion (TDC), which has the advantage of easy integration. Methods based on phase measurements have also been proposed in literature. During my PhD research, I developed a new strategy for the high resolution time measurement. This is able to

reach time resolution in the order of few picoseconds, based on the additional features of very limited analog hardware, easy integrability in contexts where computation power is already present, with no need of intensive calibration procedures, and minimizing the necessity of a-priori information. The proposed technique relies on a mixed analog – digital architecture, in order to exploit the full power of the two electronics worlds. The time measurement method is based of the waveform interpolation. The time interval measurement system is based on a Voltage Controlled Oscillator (VCO), whose oscillating frequency is changed according to the arrival of an external event, to be time-resolved. The VCO acts as a time reference for the measurements. The event changes the DC bias of the varactor, thus the oscillating frequency changes at the exact time of arrival of the event. From this point on, the identification of the instant of arrival of the event becomes a mathematical issue. The signal corresponding to the event is processed by an analog front-end circuit, which provides a stable DC bias to the VCO varactor. In particular, when the event signal arrives, the front-end circuit generates a step function on the output port which resets after the processing

ends. The sinusoid generated by the VCO is then sampled by an ADC. Data processing is carried out by the FPGA. The test board is composed by two identical channels. Each of them is able to measure the time distance between the arrival of the event and its sampling clock. Thus, the two measurement channels need to be driven by the same sampling clock. At the arrival of the event which triggers the measure, a steep transition on the tuning voltage is generated by an analog circuit. The VCO adjusts the frequency of the output sinusoid. Since the phase of the sinusoid is a continuous function during the frequency transition, there is a relationship between the phase of the sinusoid before and after the event arrival. In order to evaluate the measurement method performances, the entire system has been simulated in Matlab. The simulation validated the theoretical hypothesis on the working principle, and set to baseline of performance to be achieved with the developed test system. In order to test the method, an application specific test board has been developed. The board is composed by two separate channels, with the same sampling clock. Each channel can measure the time of arrival

of the event with respect to the sampling clock. The presented technique is an innovative method to achieve high-resolution measurement of time intervals, which is suitable to be exploited in various scientific applications, for instance as an instrument of test for high-resolution time measurement devices. The proposed technique offers the possibility to measure time interval between two or more events with high resolution and high linearity. The achieved results place this method in the highest rank of the high-resolution time measurement techniques, with a resolution of few picoseconds both in simulation and in experiment. The method relies on the purity of the sinusoid generated by the VCO, which is the time reference device. The optimization of the VCO is the key for the improvement of the achievable resolution.



1. Block diagram of the Time Measurement Instrument



2. Picture of the realized instrument

## C-BASED HIGH LEVEL SYNTHESIS OF PARALLEL APPLICATIONS TARGETING ADAPTIVE HARDWARE COMPONENTS

**Vito Giovanni Castellana** - Supervisor: **Fabrizio Ferrandi**

The ever increasing complexity of embedded systems is driving design methodologies towards the use of abstractions higher than the Register Transfer Level. In this scenario, High Level Synthesis (HLS) plays a significant role by enabling the automatic generation of custom hardware accelerators starting from high level descriptions (e.g., C code). Typical HLS tools exploit parallelism mostly at the Instruction Level (ILP). They statically schedule the input specifications, and build centralized Finite State Machine (FSM) controllers. However, the majority of applications have limited ILP and, usually, centralized approaches do not efficiently exploit coarser granularities, because FSMs are inherently serial. Novel HLS approaches are now looking at exploiting coarser parallelism, such as Task Level Parallelism (TLP). Early works in this direction adopted particular specification languages such as Petri nets or process networks, reducing their applicability and effectiveness in HLS. This work presents novel HLS methodologies for the efficient synthesis of C-based parallel specifications. In order to overcome the limitations of the FSM model, a parallel controller design is proposed, which allows multiple flows to run concurrently, and offers

natural support for variable latency operations, such as memory accesses. The adaptive controller is composed of a set of interacting modules that independently manage the execution of an operation or a task. These modules check dependencies and resource constraints at runtime, allowing as soon as possible execution without the need of a static scheduling. The absence of a statically determined execution order has required the definition of novel synthesis algorithms, since most of the common HLS techniques require the definition of an operation schedule. The proposed algorithms have allowed the design and actual implementation of a complete HLS framework. The flow features automatic parallelism identification and exploitation, at different granularities. An analysis step, interfacing with a software compiler, processes the input specification and identifies concurrent operations or tasks. Their parallel execution is then enabled by the parallel controller architecture. Experimental results confirm the potentiality of the approach, reporting encouraging performance improvements against typical techniques, on a set of common HLS benchmarks. Nevertheless, the interaction with software compilers, while profitable for the optimization of the input

code, may represent a limitation in parallelism exploitation: compilation techniques are often over-conservative, and in the presence of memory operations accessing shared resources, force serialization. To overcome this issue, this work considers the adoption of parallel programming paradigms, based on the insertion of pragma annotations in the source code. Annotations, such as OpenMP pragmas, directly expose TLP, and enable a more accurate dependences analysis. However, also in these settings, the concurrent access to shared memories among tasks, common in parallel applications, still represents a bottleneck for performance improvement. In addition, it requires concurrency and synchronization management to ensure correct execution. This work deals with such challenges through the definition of efficient memory controllers, which support distributed and multi-ported memories and allow concurrent access to the memory resources while managing concurrency and synchronization. Concurrency is managed avoiding, at runtime, multiple operations to target the same memory location. Synchronization is managed providing support for atomic memory operations, commonly adopted in parallel

programming. These techniques have been evaluated on several parallel applications, instrumented through OpenMP pragmas, demonstrating their effectiveness: experimental results show valuable speed-ups, often close to linearity with respect to the degree of available parallelism.

# TCAD MODELING OF PERFORMANCE AND RELIABILITY OF NANO-SCALE MEMORY DEVICES

Niccolò Castellani - Supervisor: Prof. Alessandro Spinelli

This work presents the most part of the results obtained during collaborative working projects between the nanolab group of the Politecnico di Milano and external organizations. Since each organization has a proper research field and specific requests, results are presented in three divided sections focused on different topics concerning the applied analysis of innovative memory devices and following a unique scientific approach. Each problem was faced by a theoretical point of view and using numerical tools based on finite elements methods in a *Technology Computer Aided Design* (TCAD) framework. The following list provides an introductory description.

I. Flash memory devices: *An insight study of threshold fluctuations, Politecnico di Milano and Numonyx Inc., Agrate Brianza, Italy.* A comprehensive study of Random Telegraph Noise (RTN), a very specific issue of MOS-based memory technologies, is presented. MOS-based technologies have been studied for long time so literature works provide a full description of the device working principle pointing out a detailed list of problems that might affect next technological nodes. What we did is to study one of these issues using our approach and

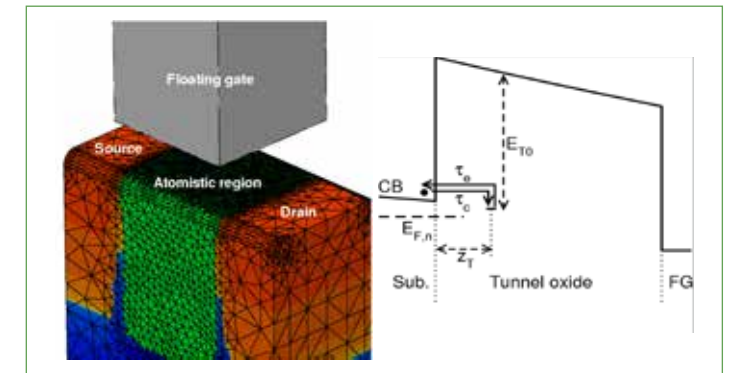
without having specific models or tools at our disposal. The Random Telegraph Noise is the fluctuation of threshold voltage of a Flash memory device between two (or more) well definite levels. It is caused by one (or more) trap sites located in the tunnel oxide and randomly over the channel area. The RTN is one of the most challenging reliability issues for nanoscale Flash memory devices and the statistical analysis of its features is a mandatory task to fully understand its impact on state-of-the-art technologies. In particular, the statistical distribution of the RTN amplitude is an extremely important piece of information for Flash technology because the large number of devices per chip requires the exploration of the distribution down to very low probability levels. We start analyzing on the numerical approaches used to reproduce the statistical RTN distribution for scientific and industrial purpose. We highlighted the simulation speed limit of the conventional Monte Carlo approach and studied the reliability of faster alternative methods. Then we focused the attention on the RTN physics and we developed a new model to study the three-dimensional electrostatic implications on the spread of RTN time constants.

Moreover, since spectroscopic analyses of RTN traps have always represented a useful way to gather information on the quality of a technology process and on the nature of the defects mining it, we also studied the implication of three-dimensional electrostatic effects on the most common experimental extrapolation methods, providing for the first time a quantification of their roughness.

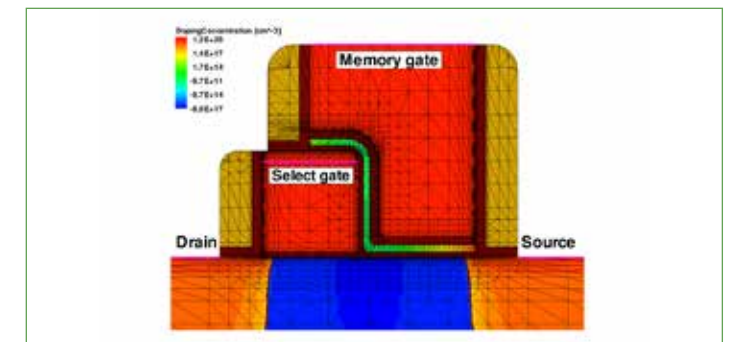
II. Charge-Trap Split-Gate memory devices: *A numeric study for technological optimization, Politecnico di Milano and CEA - LETI, Grenoble, France.* This is the most company-related research, strictly bonded to production process: here the request was to face a technological problem, understand it and, above all, provide an affordable solution. A Charge trap Split Gate memory device is composed by two overlapped MOS devices controlled by two different gates. The memory gate controls the storage unity, while select gate is introduced to modulate the current at the drain contact, to reduce the current dissipation during program operations. Charge trap Split-Gate memory devices are considered a promising technology because they can combine typical

advantages of a discrete storage layer, such as robustness and scalability, with a low power consumption design, small circuitry, common lithographic realization and high speed operations. This technology is a good candidate to be used for embedded applications as Floating-Split-Gate memory replacement, but phenomena of program disturb were addressed to limit the technology scaling. Here the program disturb was faced using a numerical approach: we suggest a possible solution and analyze the trend with scaling, keeping an angle of view as close as possible to real industrial concerns.

III. Thyristor-Based memory devices: *Physical comprehension of a brand new device working principle, Politecnico di Milano and Micron Inc., Agrate Brianza, Italy.* This work took place as scientific support to the comprehension of the working mechanisms of a brand new memory device. In this interesting challenge the most part of the of the efforts were spent to provide a reasonable description of device behavior and we decided to discern and optimize existing models rather than developing new tools. Here we studied the working principle of a new thyristor-Based memory device, designed



1. Left: schematics for the TCAD structure of the MOS-based template cell investigated in Part I. Right: schematics for the conduction band profile, highlighting a single RTN trap and its possibility to capture/emit a channel electron.



2. Schematics for the charge trap split gate cell used in Part II

as a possible candidate for SRAM or DRAM replacement. This promising technology was studied in the Micron Technology Inc. laboratories with the aim to determine the industrial feasibility of the project, to be introduced on the market. This work takes place as scientific support in this scenario,

with the aim to reach a good physically based comprehension of the phenomena involved in common device operations. Here we provide an accurate description of the involved physical phenomena that occurs during static and dynamic operations.

# COMPUTATIONAL PREDICTION OF GENE FUNCTIONS THROUGH MACHINE LEARNING METHODS AND MULTIPLE VALIDATION PROCEDURES

**Davide Chicco** - Supervisor: **Marco Masseroli**

Genes are the most important and essential molecular units of a living organism, and the knowledge of their functions is a crucial key in the understanding of physiological and pathological biological processes, and in the development of new drugs and therapies. This association between a gene and its function has been named as *biomolecular annotation*. Unfortunately, the discovery of new annotations is often time-consuming and expensive, because so are biological *in vitro* experiments carried out by physicians and biologists.

Rapid advances in high-throughput technology have been making many new gene functions available online in public databases and data banks. Despite their undeniable importance, these data sources cannot be considered neither complete nor totally accurate, because annotations are not always revised before their publication, and sometimes include erroneous information, beside being incomplete by definition. In this scenario, computational methods that are able to quicken the curation process of such data are very important.

This has motivated the development of computational algorithms and softwares that

utilize the available genomics information for gene function prediction, able to provide prioritized lists of biomolecular annotations to the biologists, in order to orientate their future research and experiments.

With this thesis, we first face the problem of predicting novel gene functions (or biomolecular annotations) through different computational machine learning methods, in which we take advantage of the properties of co-occurrences of annotations to suggest new likely gene functions. We propose some computational methods, implemented in an integrated framework, able to produce prioritized lists of predicted annotations, sorted on the basis of their likelihood. Particularly, we enhance an annotation prediction method already available in the literature, and then developed two variants of it, based on gene clustering and term-term similarity weight.

In addition, we also deal with the issue of the validation of the predicted annotations. Scientists keep adding new data and information to the annotation data banks as long as they discover new gene functions, and sometimes these data are erroneous or inaccurate. In addition, new discoveries are made every day, and the

available information cannot be considered definitive. For these reasons, such databases are always incomplete. This leads to a significant problem of validation, because we do not have a true *gold standard* to refer. So, we designed and developed different validation procedures able to check the quality of our predictions. We introduce a validation phase consisting of a Receiver Operating Characteristic (ROC) analysis, a search for the predicted annotations into a new updated database version, and possibly an analysis of the available knowledge in the literature and through some available web tools.

To better understand the variation of the output predicted lists of annotations, we design and develop new measures, based on the Spearman coefficient and the Kendall distance. Such measures are able to state the level between two lists by analyzing the difference between positions of the same element in two lists, and by evaluating the number of element couples having contrary order in the two lists. These measures demonstrated to be able to show important patterns otherwise difficult to notice.

Finally, we provide a visualization and statistical tool able to state the novelty of the

predicted gene annotations, denoted as *novelty indicator*. For each gene, this tool is able to depict the tree graph of the predicted ontological annotation terms, producing images easily understandable also by non-experts, and also a statistical value that states the level of *novelty* of the prediction.

Our tests and experiments confirmed the efficiency and the effectiveness of our algorithms, by retrieving manifold predicted annotations as confirmed in the updated database or in the literature. The similarity measures resulted very useful to understand the similarity of our predicted lists, making us able to see specific similarity patterns while key parameters vary. The *novelty indicator* possibly, resulted very useful in producing tree graphs able to make our lists of predicted biomolecular annotations clearly usable by biologists and scientists.

Currently, the discovery of new biomolecular annotations is a time-consuming process, in biology. Biological experiments are long-lasting and expensive, and so are the *in vitro* validation procedures of biomolecular annotations that finally state if an annotation is biologically correct and reliable. In this thesis, we described some computational methods that we explored to predict new gene functions, to verify their consistence with multiple validation procedures, to compare different annotation lists, and to comprehend the value of our predictions.

We started from an existing algorithm by Khatri et al.

(“A semantic analysis of the annotations of the human genome”, 2005) and first enhanced it with automated procedures able to select the best key parameters for the SVD truncation level and for the likelihood threshold. Departing from a previous work developed in our research group, we moved to new algorithms, based on gene clustering and on term-term similarity weights. Our tests showed that these methods are able to improve the reconstruction of the output annotation matrix. This led our prediction to increase the percentage of annotations later found confirmed by the annotations become subsequently available within updated versions of the public data bases, while decreasing the number of total predictions.

Together with these new techniques, we also provided some useful validation procedures. Since we do not have any facility to run *in vitro* experiments, we designed and implemented different validation methods based on ROC analysis, comparison with available versions of annotation databases, and literature and analysis. Our validation procedures resulted very efficient in checking the correctness of our predicted annotations.

To better compare annotation lists, when varying the predictive methods or the input parameters, we introduced new coefficients, based on Spearman rank and on Kendall measure, which are able to detect *how much* a list differs from another one.

These new measures demonstrated very effective and able to reach the objectives which they were designed for, showing similarity values that vary proportionally to the difference between the lists.

Finally, we designed and developed software functionalities to analyze the *importance* level of our predictions. First, with a powerful DAG viewer visualization tool, able to depict the DAG tree of the (former and predicted) annotation terms of each gene. Afterwards, by implementing a *novelty indicator* measure based on the semantic similarity between the original gene and the new annotated gene, fit to classify each prediction from *little novel* to *very novel*.

The DagViewer with its DAG graphs that demonstrated to be very useful for understanding which predictions more relevant and interesting than others, can be easily interpreted also by non-expert readers. The introduced novelty indicator values resulted consistent with the visualization tool results, and “translated” the concept of *prediction novelty* from qualitative visualizations to quantitative values. After designing these modules, we integrated them all into a unique software, that may be extended in the future with new functionalities, too.

We believe that the tools presented within this thesis may be very useful to the bioinformatics and scientific community to address future research experiments about gene functions.



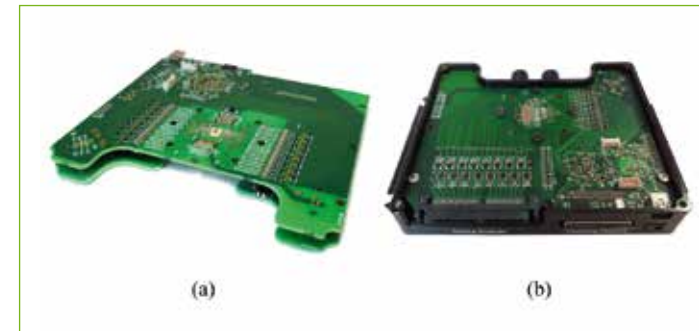
## DEVELOPMENT OF ELECTRONIC SYSTEMS FOR SINGLE-PHOTON AVALANCHE DIODE ARRAYS

Andrea Cuccato - Supervisor: Ivan Rech

In recent years, the biomedical and chemical research has focused on optical analysis to better understand biological processes, indeed these non-invasive measurements are the best solution for in vivo experiments and medical diagnostic tests. This trend has pushed the research in the electronic field towards the development of high-performance photodetectors, in order to meet the strict requirements imposed by applications. As an example, in the last decades single-molecule spectroscopy (SMS) has rapidly grown in the biomedical and biochemical field. The technique involves examining very low concentration samples, therefore it requires photodetectors with extremely high sensitivity and low noise, and nowadays the best answer to these requirements is represented by single-photon detectors. In particular, great achievements in single-photon avalanche diode (SPAD) arrays have been recently made, pushed by a growing demand for parallel experimental setups. Indeed a multispot approach significantly reduces the measurement time, allowing the study of fast dynamic processes in analysis such as SMS, single-molecule Förster resonance energy transfer (smFRET) and fluorescence lifetime imaging microscopy

(FLIM). Moreover, the temporal response of the single-photon detector is of utmost importance when the time-correlated single-photon counting (TCSPC) technique is employed to obtain fluorescence decay curves with subnanosecond time resolution. Nowadays, state-of-art imaging sensors integrate thousands of single-photon detectors on the same chip. As an example, devices designed in CMOS technology make possible to integrate the detector and circuits for time-to-digital conversion within the pixel, thus allowing the fabrication of large 2-D SPAD arrays for high frame-rate time-resolved imaging applications, like laser ranging (LIDAR) or FLIM. However, from the detector point of view, these systems suffer from poorer performance with respect to the best in literature, which is represented by SPADs (and SPAD arrays) developed in a customized technology process. As a drawback, the latter limits the number of pixels that can be integrated in a single chip, due to on-chip signal routing issues. Moreover, to exploit the best detector performance, a hybrid CMOS-custom technology pixel structure has to be implemented, which leads to off-chip signal routing issues and increased area occupation. To break the strong tradeoff between performance

improvement and parallelization, in our research group a complete 32-channel system for TCSPC measurements, based on a custom-technology 32x1 linear SPAD array, has been designed and fabricated with the aim to maximize the detector performance. Part of this complete system constitutes the main subject of this thesis work, which involved the design and fabrication of a compact 32-channel time-resolved single-photon detection instrument, capable of performing single-photon counting analysis as a stand-alone module, and TCSPC analysis when connected to a multichannel TCSPC instrument. The basis of this project are represented by an 8-channel detection head, developed in a previous thesis work, and the experience acquired from the development of a complete single-channel TCSPC instrument. The 32-channel time-resolved detection head exploits the extremely high performance of the 32x1 custom-technology SPAD array, which works in conjunction with a 32x1 CMOS AQC array. However, this hybrid architecture has led to many design issues to deal with, like signal routing, electrical crosstalk between parallel channels, connectivity, power management, heat dissipation and the development



1. (a) Picture of the two boards that make up the 64-channel single-photon detection head; (b) picture of the 32-channel time-resolved single-photon detection head (the upper aluminum cover has been removed to show the integrated arrays).

of suitable packaging solutions. In particular, since the SPAD dark counts can be reduced by cooling the detectors, a customized sealing system has been designed to allow a temperature control while preventing the formation of dew. The main advantage with respect to a standard sealed package (i.e. TO-like packages) is the larger number of connections that can be implemented between integrated circuits and PCBs for signal processing and power management purposes. A complete and parametric characterization of the detection head has been performed, and results are in good agreement with SPADs developed in a custom technology process. During typical operating conditions, i.e. with 6V of

SPAD excess bias voltage and  $-10^{\circ}\text{C}$  of SPAD temperature, the module features a mean dark count rate lower than 400cps, a photon detection efficiency peak close to 45%, an afterpulsing probability lower than 2% and  $10^{-2}$  of optical crosstalk probability for adjacent pixels. Moreover, the system response function to a photon-detection event features a time jitter as low as 60ps FWHM at a mean count rate up to 1Mcps per channel, hence suitable for multidimensional TCSPC measurements. Besides the time-resolved module, during this thesis work a 64-channel single-photon detection head for single-photon counting measurements has been designed and fabricated. Indeed, 2-D SPAD arrays and SPADs with enhanced efficiency

in the red region of the visible spectrum have been recently developed in our research group. The new detection head has been developed with the aim to make these new detectors employable in an experimental setup, thus meeting a growing demand for high performance and high measurement throughput in biochemical applications. The main modifications with respect to the 32-channel module concern the improvement of the power management system and the digital processing capability. In the new detection head, the channel number increase has been obtained at the cost of a reduced time resolution performance, due to signal routing issues at both chip-level and PCB-level. However, the two systems developed in this thesis work are complementary, each one targeting a specific range of applications. Moreover, they actually represent an attempt to investigate the limit of the tradeoff between performance and parallelization for custom-technology-based single-photon detection systems.

# ACOUSTIC PIPELINE MONITORING: THEORY AND TECHNOLOGY

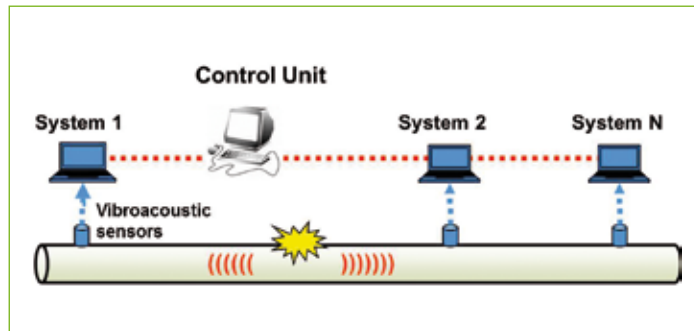
**Silvio Del Giudice** - Supervisor: **Giancarlo Bernasconi**

Nowadays pipelines are a widespread, efficient and cheap mean of transportation for many fluids, in particular hydrocarbons. Possible failures or damages to such structures might cause serious economic, environmental and health consequences, that's why it's of paramount importance to monitor the working condition of pipelines.

In particular as regards to oil and gas pipelines, in many country the law compels fairly to install systems that detect leaks of a certain amount within a certain delay.

Since conduits are often buried or laid on the sea-bottom it's difficult to perform direct monitoring and there arose the need to think out techniques of indirect monitoring, often called non-invasive or non-destructive. The technique dealt with in this thesis is acoustic monitoring which exploits the fact that many events to be monitored in the pipeline produce or influence acoustic transients and these propagate as waves within the transported fluid at distances of many kilometers, carrying information on the originating event and on the propagation channel.

During my PhD I contributed to the development of a vibro-acoustic monitoring system (e-vpms® registered technology) which employs sensors to



1. e-vpms® vibroacoustic monitoring system

measure acoustic and other signals to monitor pipelines. The design and development of this technology have required the study of the theory of wave propagation in pipelines.

## Theory

I analyze the theory of acoustic wave propagation as far as it is useful to describe propagation of waves that can be exploited for pipeline monitoring, considering both liquid and gas as filling fluids. In the case of gas-filled pipelines, the pipe can usually be considered a rigid medium and the simple wide-tube approximation which describes the fluid-borne wave propagation is sufficient for most practical application; in the case of liquids, the compressibility of the fluid is comparable to the compressibility of the

pipe material and therefore the properties of the shell are important and possibly even the properties of the external medium.

I report a selection of propagation models according to the phenomena they take into account, most of the models were selected for their relative simplicity, being analytic or semi-analytic. For each model the computation method is presented.

Then I choose a common scenario of particular interest to hydrocarbon transportation, that is a buried oil-filled pipelines, in order to compute and compare the results of the models presented.

The comparison shows that none of the methods presented can accurately compute the propagation parameters in the whole range of frequency of the

practical interest, at least from unit Hz up to hundred Hz, but a more general matrix method is needed.

The method described allows to consider a waveguide made of 3 co-axial cylindrical layers each of which can be either solid or fluid. In case of solid it has an elastic behavior, described by elastic parameters such as Lamé's. In case of fluid medium, both thermal and viscous effects are considered, and for each of them, both bulk and boundary layer types.

The switch from solid to fluid is easily performed by replacing Lamé parameters with proper functions of fluid's viscous and thermal parameters.

In this way a unique model can be used for both gas- and liquid-filled pipelines, which can support technologies such as e-vpms®, for monitoring of pipelines in different ways like for example real-time remote detection of third-party interference or long-term monitoring of transportation efficiency or detection of malfunctioning, failures and anomalies along pipeline. This general matrix method is called AXSYM-3L.

Then I show some of the results of the model, limited to fluid-filled pipes, buried or submerged and to the fundamental fluid-borne propagation mode. In particular some tests are

performed to verify the effect of the surrounding media, and others of the thermal boundary layer.

Results of AXSYM-3L are shown together with the other models, and AXSYM-3L looks suited to provide accurate results in a wider frequency range than other models.

Finally AXSYM-3L as a matrix method is used to compute and display also the profiles of field variables and these are shown in physical units for the usual scenario at the typical frequency of 50 Hz.

## Technology

e-vpms® monitoring system relies on a discrete network of pressure and vibration sensors installed on the pipeline, at relative distances of tens of kilometers. The acoustic and elastic waves produced for example by third party interference and by flow variations (leaks, spills, valve regulations, pig operations, etc.), propagate along the pipeline, and they are recorded at the monitoring stations.

The propagation parameters thus have a significant effect on measurements and here lies the importance of wave propagation theory.

Sensors and GPS signals are acquired in each local unit and local units are connected to a control unit that executes

the multichannel processing, enabling the detection, localization and classification of the triggering event, but also to gain useful information on the pipeline system through advanced processing. The e-vpms® system has been installed on both oil and gas transportation lines during several field campaigns.

I show some interesting experimental results obtained on different oil and gas pipelines: The field test campaigns have permitted to:

- derive and validate mathematical models of sound propagation within pipes;
- tune and realize a detection system for leaks and TPI (third party interference);
- track PIGs (pipeline inspection gauges) both in low pressure and high pressure scenarios;
- characterize the pipeline as equivalent acoustic transmission channel;
- apply long term monitoring of pipe for analysis of the transport efficiency, detection of anomalies such as deformations/obstructions and identification of fluid properties.

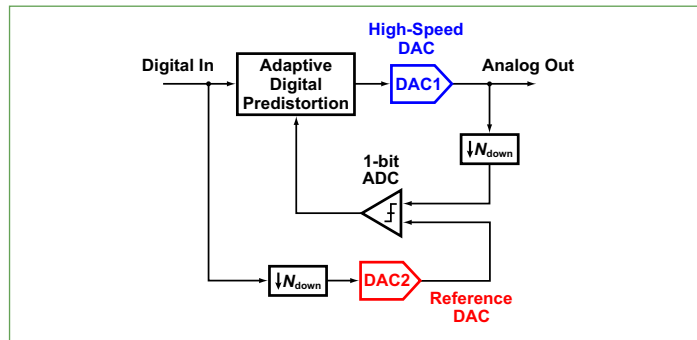
# TECHNIQUES FOR HIGH SPEED AND LOW POWER DIGITAL-TO-ANALOG CONVERTERS

Andrea Fenaroli - Supervisor: Salvatore Levantino

The extraordinary evolution of today communication systems towards higher levels of integration is increasingly leading to the implementation of more and more complex functions in the digital domain. This process drives the need for the realization of high performance data converters in low-cost ultra-scaled nanometer CMOS processes, connecting the systems digital core to the real analog world.

In particular, academic and industrial research in the field of high-speed Nyquist-rate Digital-to-Analog Converters (DACs) is pushed forward by the growing interest in multi-carrier multi-band transmitters, for both wireless and wireline systems. Several communications standards have been recently developed requiring a DAC in the transmitter path with a sampling frequency in the GS/s range, while posing, at the same time, extremely stringent constraints on resolution, linearity and power.

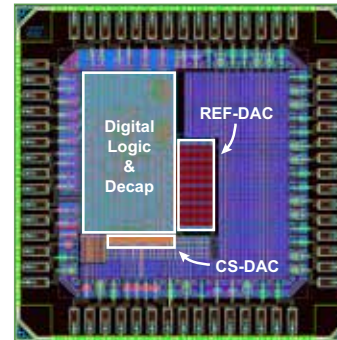
Unfortunately, the feature device scaling and the supply voltage reduction, in addition to the typically noise environment of large Systems-on-Chips (SoCs), are introducing critical issues on the design of such converters. More precisely, the analysis of state-of-the-art CMOS DACs reveals two fundamental trade-offs limiting their performances:



1. Architecture of the proposed linearization scheme

the first one between low-frequency and high-frequency linearity while the second one between linearity and power efficiency. The essential objective of this thesis is the definition and the development of new design methodologies and techniques for the realization of high-speed high performance DACs suitable for the integration in ultra-scaled CMOS technologies, which allow overcoming the fundamental trade-offs limiting performances. In particular, the thesis core is the introduction of a novel digital technique based on the use of a new multipath LMS adaptive filter that continuously measures and cancels static non-linearity errors in the background of normal operation. Since analog circuits impairments are canceled out in the digital domain, a design full-

oriented at optimizing dynamic performances is allowed, overcoming in this way the typical trade-off between low-frequency and high-frequency linearity. The proposed scheme includes an additional voltage comparator and an ancillary slow rate, yet linear, DAC, resulting in an extremely simple and low-power overall DAC architecture. Behavioral simulations confirm the effectiveness of the algorithm in increasing the linearity across the full bandwidth, with an improvement ranging from 26dB at DC to 15dB at the Nyquist frequency. A 10-bit 2.5GS/s current-steering DAC including the digital calibration scheme has been designed and implemented in 28nm CMOS. Using a single supply voltage of 1V, it provides a 10mA full-scale current onto a



2. Chip prototype in 28nm CMOS

50Ω differential load, resulting in a -2dBm output power. While the current-steering circuit consumes 36mW power at the maximum operating frequency, the ancillary 10-bit 25MS/s resistor-string DAC draws only about 20μA. Simulation results show a DAC SFDR greater than 65dB across the entire Nyquist bandwidth, revealing the potential effectiveness of the proposed linearization technique to advance the state-of-the-art.

## INTEGRATING FORMAL METHODS WITH INDUSTRIAL STANDARDS IN THE DEVELOPMENT OF FLEXIBLE MANUFACTURING SYSTEMS

Luca Ferrucci - Supervisor: Dino Mandrioli

The thesis presents a complete, innovative, formal verification approach to model certain class of manufacturing systems, Flexible Manufacturing Systems (FMS). It is the final product of a collaboration project between a group of automatic control engineers of the "Istituto di Tecnologie Industriali ed Automazione" (ITIA) of the CNR of Milano and a group of informatics engineers of the "Dipartimento di Elettronica, Informazione e Bioingegneria" (DEIB) of the Politecnico di Milano. A complete, integrated Model-Driven environment is the final target of the project; this environment should support the designer from the modeling phase to the formal verification phase, to obtain an agile, flexible development process. To overcome the difficulties using formal languages, the approach is based on widely-used, graphical but semi-formal high level modeling languages. We have chosen Stateflow, which is a graphical language derived from Statechart, and encoded it in terms of formulae of the TRIO logic, which is a first-order metric MTL-like temporal logic; by this way we provided a rigorous, compositional, run-to-completion semantics for Stateflow, based on concepts of micro and macro-steps. The formal model of the system and the semantics,

glued together in a set of logic axioms, permit us the formal verification of a set of real-time qualitative or quantitative user-defined properties, using a fully automatic bounded model checker tool, Zot, which has been developed internally at the DEIB. The approach is general enough to allow using different modeling languages, logical formalisms or model-checking tools. The approach is illustrated and validated through the development of a realistic case study of a Robotic Cell, a simple FMS. Each component of the cell has been modeled in the standard IEC61499 language, a widely-used graphical language based on the concept of Function Block, analyzed during my third year of the doctoral studies and standardized recently. Components are then translated in Stateflow in a straightforward way, while the component's network has been implemented using Simulink. In the last part of the work, we have defined a new metric linear temporal logic called X-TRIO, derived from TRIO, which permit to predicate over micro and macro-steps. The particularity of this new logic is the time domain: given a temporal domain  $T$  (for example, Real or Integer numbers), it has been augmented with the set of infinitesimal numbers, which are numbers that, in absolute

value, are smaller than any real number. The augmented domain, called  $\overline{T}$ , is a total ordered metric domain. The idea is to substitute each zero time transition with an infinitesimal time transition, that is not a "real, tangible" time transition (since it is less than any real number), such that it is possible to give a temporal order to micro-steps. By this way, we note that is not possible to the system to be in the same "state" at the same time, so the state is a function of time and we can avoid logical contradiction (e.g. a propositional variable that is true in a micro-step and false in the next one can lead to a contradiction in a "normal" time domain with zero-time transitions). In X-TRIO has been defined a new Next time operator  $X$  to refer to the next system state which occurs in a non-standard time instant in case of a micro-step, or in a standard one in case of a macro-step. We have analyzed only a fragment of X-TRIO, called  $X\text{-TRIO}_N^P$ , where the domain is restricted to the number of the form  $x=v+k\varepsilon$ , where  $v,k \in \mathbb{N}$  (the set of natural numbers),  $\varepsilon$  is a fixed infinitesimal constant and formulae are quantifier-free. It has been demonstrated that the  $X\text{-TRIO}_N^P$  logic is undecidable, but has been identified a suitable syntactic restriction to the

formulae such that the new fragment of the  $X\text{-TRIO}_N^P$  obtained by this restriction is decidable. Decidability has been demonstrated by reducing the satisfiability problem of the fragment  $X\text{-TRIO}_N^{P-R}$  to that of PLTLB (Propositional LTL with both past and future operators) through a translation function that preserves satisfiability. The decision procedure works also in the opposite way, demonstrating that a PLTLB formula and a corresponding restricted  $X\text{-TRIO}_N^{P-R}$  formula obtained by the translation function are equisatisfiable. The translation function of the decision procedure has been used to encode the decidable fragment of  $X\text{-TRIO}_N^{P-R}$  into LTL. A plug-in for the model checker Zot has been developed, which perform the translation of formulae written in the decidable fragment of  $X\text{-TRIO}_N^{P-R}$  into LTL, to permit the verification of models and properties written in the new logic. Another plugin for Zot, called SF2TRIO, has been developed during the last year, to allow end-users to specify the various parts of a Stateflow graph using a set of commands which are used to declare states and transitions of the graph, hiding the logical encoding. The initial case study, defined during first year, has been rewritten in the new logic and a set of properties have been verified

using Zot, to test experimentally the correctness of the translation function.



# MODELLING AND CONTROL OF AUTOMATIC TRANSMISSION SYSTEMS IN TWO-WHEELED VEHICLES

Paolo Giani - Supervisor: Prof. Sergio M. Savaresi

Transmission systems represent the main component of all vehicles powertrains, as they provide the link between the engine and the driving wheels. Due to their crucial function, the automatic management of these systems is a challenging and difficult task, which requires a thorough knowledge of both the powertrain and the vehicles dynamics.

In years, several automatic systems have been developed to assist the driver in the transmission management. However, these are mainly tailored to cars, trucks and other heavy vehicles, while applications in the two-wheeled vehicles field are still largely unexplored. In this context, the present research aims to expand the current state-of-the-art considering the latter class of vehicles. In particular, two vehicle categories have been considered, namely high-performance motorcycles and conventional bicycles for city usage.

In details, two motorcycle prototypes have been employed, see Figure 1. All of them were based on the standard aprilia® RSV4 model, and each of them was properly equipped with the required sensors and actuators needed to automatically control of the transmission. It is worth pointing out that this motorcycle



1. Motorcycle prototypes considered in the work.

model particularly suits the application, as it is one of the most powerful motorcycles actually available on the market and its weight and engine power characteristics make the automatic transmission management a complex task. The high-performance motorcycle powertrain is, indeed, a complex system, made of several components, such as the engine, the clutch, the gearbox and all the additional equipment needed to make the system electronically controllable.

On the other hand, the innovative bicycle analysed in the work is based on a standard bicycle, see Figure 2, which was equipped with an automatic transmission system and all the required additional devices that make it

electronically controllable. In this application, the model of the powertrain system requires a consistent effort to be derived, as, in bicycles, also the rider (i.e., the vehicle source of power) is part of the powertrain system. This process, however, is indispensable to achieve an effective control of the vehicle powertrain. As a matter of fact, the bicycle source of power is a (very) low power engine and the efficient use of its energetic resources is the crucial point to considerably improve the whole system efficiency.

The methodological pathway adopted in the investigations can be summarised into three main steps:

- System modelling. To understand the underlying dynamics of the systems, consistent models of both transmissions and powertrains



2. Experimental bicycle prototype.

of the two vehicle classes are needed. To this purpose, an accurate analysis of each powertrain component was first addressed, and the models parameters were identified through specific experimental tests on the systems. The derived models were finally validated using experimental data.

- Transmission-level control design. Based on the derived models, suitable controllers for the transmission components were designed, so as to ensure prescribed performance levels. In motorcycles, such controllers allow the precise modulation of the clutch position and the selection of the inserted gear. On the other hand, the transmission controller on bicycles allows to finely regulate the transmission ratio between the pedal and

the wheel velocities. These are the enabling steps for the higher-level controllers implementation.

- Vehicle-level control design. The proper and coordinate use of the transmission-level controllers allows pursuing higher-level objectives, as, e.g., manage comfort and safety related aspects or regulate the energy flows in the vehicle. In motorcycles, the controlled transmission system was used as main actuator to perform automatic gear shifts and launch manoeuvres. On the other side, in bicycles, the automatic transmission system was employed to actively regulate the heart rate of the rider, in order to optimize the riding effort.

The achieved results, in both vehicle and for each milestone

of the work, were carefully analysed and validated through suitable experiments, witnessing the effectiveness of the proposed approaches for the control of the transmission systems in the two vehicle classes.

## AGENT-BASED WATER RESOURCES MANAGEMENT IN COMPLEX DECISION-MAKING CONTEXTS

**Matteo Giuliani** - Supervisor: **Prof. Rodolfo Soncini-Sessa**

The presence of multiple, institutionally independent but physically interconnected actors is a distinctive feature of the complexity characterizing most of the decision-making problems in environmental contexts. As dealing with many conflicting stakeholders requires to replace the concept of optimality by that of Pareto efficiency, the presence of many decision makers requires some kind of coordination and a cooperative attitude of the involved parties, who agree on adopting a fully coordinated strategy to maximize the system-level performance. These assumptions are often unpracticable in real world contexts, where the decision makers generally belong to different institutions or countries. In these situations, they independently pursue local interests and produce negative externalities leading to a low system-wide efficiency. Game theory and simulation-based approaches are generally used to analyze these issues from a descriptive standpoint, while their prescriptive use in decision support systems to design coordination mechanisms between the originally self-interested decision-making actors remains a challenge.

This thesis contributes a novel decision analytic framework based on multi-agent system

(MAS) to study water resources planning and management problems in complex decision-making contexts. The aim of the proposed decision analytic framework is to combine descriptive and prescriptive methods, which provide informative tools to represent the actual decision-making context as well as decision support procedures to recommend proper coordination mechanisms. The adoption of an agent-based framework naturally allows the representation of a set of decision makers or stakeholders (agents), which act in the same environment and need to coordinate to maximize the system-wide efficiency in the use of the available water. This agent-based perspective aims to move beyond the traditional centralized approach to water resources management and to explore different levels of cooperation, from fully coordinated strategies and full information exchange to completely uncoordinated practices. Moreover, coupling the agent-based modeling scheme with state-of-the-art Control Theory techniques allows a better understanding of the feedbacks between agents objectives, agents decisions, and the environment they share, as well as the description of their co-evolution and co-adaptation

under change.

The proposed framework is demonstrated in four different problems characterized by distinctive decision-making contexts, which require the adoption of different tools and methodologies. The first application introduces the problem of balancing system-wide efficiency and solutions practicability working on a hypothetical water allocation planning problem. Two extreme situations, corresponding to a fully centralized and a totally uncoordinated strategies, are first computed. An extensive analysis of the intrinsic conflict between efficiency and practicability is then developed, by highlighting the limitations (i.e., low practicability) of centralized and fully cooperative solutions in systems characterized by the presence of distributed decision making institutions. Different levels of coordination are explored by means of MAS methods based on distributed constraint optimization problems to develop effective constraint-based mechanism design strategies.

The agent-based decision analytic framework is then applied in two real world case studies to show the advantages of coordination

mechanisms based on the exchange of information. The first application assesses the value of cooperation and information exchange in transboundary systems. The Zambezi River basin is used as a case study to estimate the benefits potentially achievable by the downstream country (i.e., Mozambique), under different levels of cooperation with the two upstream countries (i.e., Zambia and Zimbabwe). The differences in the system-level benefits attained under different scenarios of cooperation provide an estimate of both the economic value of full cooperation, measured as the benefits obtained by full cooperation with respect to the one with coordination only, and the economic value of information exchange, measured as the benefits obtained with coordination with respect to the ones with no cooperation. The second application shows the potential for a co-adaptation strategy between farmers and water managers to enhance the effectiveness of agricultural water management practices in the Lake Como system. The proposed co-adaptation aims to match the needs of the farmers with the design of water supply management policies, under current and projected climate. The framework proposes an integrated procedure to model water supply and demand as coupled human (farmers and water managers) and natural (crops) systems, where people and nature interact reciprocally, form complex feedback loops, and co-evolve under changing conditions. The analysis is performed under current and future climate conditions, in

order to assess the potential for the adaptation loop to enhance the efficiency of agricultural water management practices and foster crop production as well as to mitigate climate change adverse impacts. Finally, in the last application, the framework combines reservoir policy identification, many-objective optimization under uncertainty, and visual analytics to characterize current water reservoir operations and discover key tradeoffs between alternative policies for balancing evolving demands and system uncertainties. Moreover, this chapter contributes a novel method based on input variable selection techniques to support the identification of effective mechanism design strategies for environmental protection. These tools are demonstrated on the Conowingo Dam, located within the Lower Susquehanna River, USA. The Lower Susquehanna River is an interstate water body that has been subject to intensive water management efforts due to the presence of many stakeholders (agents) affected by the Conowingo Dam operation. To provide effective support to the Susquehanna River Basin Commission and avoid policy inertia and myopia, the current regulation of Conowingo Dam is identified and refined to balance the conflicting objectives as well as the uncertainties related to the hydroclimatic variability. Then, alternative policy mechanisms are designed by directly constraining the decision space in order to dynamically condition the reservoir operating policy and better balance the primary operating objectives (i.e., guaranteeing the public

water supply and maximizing the hydropower revenue) with environmental protection and recreational interests.

Part of this thesis' contributions has appeared (or has to appear) in the following main publications: *i)* M. Giuliani and A. Castelletti (2013), Assessing the value of cooperation and information exchange in large water resources systems by agent-based optimization, *Water Resources Research*, 49, 3912-3916; *ii)* M. Giuliani, A. Castelletti, F. Amigoni, and X. Cai (2013), Agent-based distributed optimization as a tool to balance efficiency and practicability in watershed management, *Journal of Water Resources Planning and Management*, (under review); *iii)* M. Giuliani, J.D. Herman, A. Castelletti, and P.M. Reed (2013), Many-Objective Reservoir Policy Identification and Refinement to Reduce Institutional Myopia in Water Management, *Water Resources Research*, (under review); *iv)* M. Giuliani, Y. Li, A. Castelletti, and C. Gandolfi (2013), Co-adapting water demand and supply to changing climate in agricultural water systems, *Global Environmental Change*, (in preparation).

## HARNESSING ADAPTIVITY ANALYSIS FOR THE AUTOMATIC DESIGN OF EFFICIENT EMBEDDED AND HPC SYSTEMS

**Silvia Lovergine** - Supervisor: **Fabrizio Ferrandi**

Embedded Systems (ESs) and High-Performance Computing (HPC) systems belong to two distinct areas of the Information Technology (IT). On one side, embedded systems are usually designed as hardware implementation of a specific functionality, with the aim of either reducing power consumption, or accelerating the most frequently executed portion of an application. On the other side, high-performance computing systems are designed as massively parallel supercomputers with tens of thousands of processors, usually employed to solve complex, highly parallel scientific problems. Due to the distinct nature of the two domains, these systems have historically evolved following different trends.

However, many issues and design challenges are lately affecting both the domains. Among the examples, increasing computational demands and large parallelism degrees, together with the growing capabilities of silicon technology, led to the design of increasingly complex architectures and applications. As a consequence, modern embedded systems exploit the potentialities of hundreds or thousands of processing units, often heterogeneous and physically

distributed, which run in parallel on the many-core platform. These same factors led to massively parallel systems in HPC domain. As another example, power constraints, which have always been a critical requirement for embedded systems design, are becoming more and more relevant also in the high-performance domain. As a consequence, communication minimization techniques are required also for modern HPC systems design. This trend suggests that HPC systems design techniques can be exploited, at a different abstraction level, by embedded systems designers to address shared issues, and vice versa. In other words, designing modern supercomputers, as well as modern embedded systems, requires a holistic approach that relies on tightly coupled hardware-software co-design methodologies.

Among the numerous issues which are shared between the ES and the HPC domains, this work focuses on unknown, uncertain and unpredictable behaviors. Modern computing systems have to deal with this kind of behaviors, which are often related to incomplete information at design/compiler time. Unknown, uncertain and unpredictable information can be originated by different

sources. For example, the interaction with external modules (e.g., sensors, IPs, memories) generates variable and possibly unpredictable communication latency. Moreover, the nature of high-level applications is inherent uncertain (e.g., unknown number of loop iterations, unknown values of the incoming arguments of a function, unknown outcome of conditional instructions evaluation, or unpredictable memory accesses). Incomplete information at design or compile time often leads to suboptimal designs: to prevent the system from a failure, the designer must either consider all the possible scenarios, with consequent increase in area and complexity of the circuit, or take a conservative approach, with consequent decrease in performance. Moreover, this approach is error prone, since the correctness cannot be guaranteed in all the situations (such as run-time events which the designer cannot predict, or technology parameters which change unpredictably during the life cycle of the device, for example due to component degradation). In this scenario, systems able to dynamically adjust their behavior at run-time appear to be good candidates for the next computing generation in both the domains.

This work aims at defining efficient design methodologies for modern embedded systems and high-performance computing systems, able to deal with unknown, uncertain and unpredictable behaviors. For such purpose, this work introduces the concept of Adaptivity Analysis, which provides a formal approach to study the adaptivity properties of the applications. Given the different scale of the problem in the ESs and HPCs domain, Adaptivity Analysis is defined at two distinct abstraction levels.

In the embedded systems domain, Adaptivity Analysis addresses the problem of designing efficient adaptive hardware cores. More in detail, this analysis identifies, at design-time, the conditions that each instruction (or group of instructions) must satisfy to execute, according to their dependences. Such conditions, called Activating Conditions (ACs), are logic formulas. Since unknown information may occur at design-time, the ACs provide parametric results, which depend on conditions that will be resolved at run-time (e.g., outcome of the evaluation of a conditional instruction). At run-time, as soon as the unknown information is resolved, the ACs are evaluated. Once an AC is satisfied, the corresponding instruction can safely execute, thus providing an explicit activation mechanism for the instruction, which allows their dynamic auto-scheduling. Such a scheduling technique, called dynamic AC-scheduling, provides support for the High-Level Synthesis (HLS) of adaptive hardware cores. Moreover, this

work defines a prototype for this kind of system. Furthermore, this work defines a proper Intermediate Representation (IR), called Extended Program Dependence Graph (EPDG), able to provide a parallel execution model for adaptive hardware cores. The EPDG is defined starting from the Program Dependence Graph (PDG), and adding minimum control-flow dependences to it. Indeed, since the PDG only contains minimum data and control dependences, this kind of IR is not suitable to represent a parallel execution model for the specification. A formal definition of minimum control-flow dependences is provided. Finally, the dynamic AC-scheduling (together with EPDG construction and ACs computation) has been implemented as part of the PandA framework, thus providing a fully automated design methodology.

In the HPC domain, Adaptivity Analysis is used as automatic compiler-based generation of parallel irregular applications. In HPC design, run-time systems are entitled for managing unknown, uncertain and unpredictable behaviors. Hence, adaptivity analysis is used in this context as a compiler technique for the support of run-time systems. More in detail, as previously mentioned, this work focuses on the automatic parallelization of a particular class of applications, known as Irregular Applications. Due to their irregular nature (i.e., irregular data structures, irregular control flow, irregular accesses to the communication network), the efficient parallelization of irregular

applications is significantly challenging. Indeed, such a class of applications suffers of poor locality, it is characterized by high-intensive communication and synchronization, and its data structures are usually difficult to partition in a balanced way. This work introduces the Yet Another Parallel Programming Approach (YAPPA) compilation framework. YAPPA aims at efficiently parallelizing irregular applications running on commodity clusters. The novel parallel programming approach supported by YAPPA is defined by the underlying run-time library, called Global Threading and Memory (GMT). GMT cross combines different parallel programming approaches, which are usually exploited separately in the literature. It integrates: i. Global Address Space (GAS) across cluster nodes; ii. lightweight multithreading for tolerating memory and network latencies, iii. support for a fork/join programming model. YAPPA extends the LLVM compiler with a set of transformations and optimizations, which at first instrument the sequential code to run GMT primitives (parallelization phase), and then apply a set of transformations to improve the efficiency of the generated parallel code (optimization phase). A novel set of transformations is defined in this work targeting irregular applications needs. YAPPA is provided as a plugin for the LLVM compiler.

## JOINT USE OF RECONFIGURABLE ANTENNA PATTERN AND ADAPTIVE CODING AND MODULATION IN SATELLITE COMMUNICATIONS AT HIGH FREQUENCY BANDS

Marco Luccini - Supervisor: Prof. C. Capsoni, Prof. S. Primak

This Ph.D. thesis covers a large number of topics related to radio propagation for satellite communications at high frequency bands. Though it is not meant to be a comprehensive discussion of all the issues involved in new multimedia satellite communication systems, our goal was to further the research in the area of the joint use of Fade Mitigation Techniques (FMTs) to cope with atmospheric attenuation at high frequencies.

Satellite and terrestrial Telecommunication systems greatly benefit from the use of high frequencies (Super High Frequency (SHF) band and above in this thesis). The main advantage is the availability of larger bandwidths, required for the increasing demand of multimedia services. In addition, higher frequency bands require smaller antenna size for a fixed gain, which certainly represents a concrete benefit. Moreover, the possibility of using on-board antennas with reconfigurable directivity is attractive for satellite systems whose coverage area is sub-divided in spot beams for frequency reuse or regional services.

The drawback of radiowaves at frequencies above the Ka band (Q/V and W bands) is the consistent increase of the impairments caused by the

lower part of the atmosphere (troposphere). Due to the strong propagation impairments at those frequencies, the design of Telecommunication (TLC) satellite systems must rely on advanced techniques to counteract atmospheric fades. The extensive use of Fade Mitigation Techniques (FMTs), such as site diversity and on-board adaptive power allocation, from the propagation point of view, Adaptive Coding and Modulation (ACM) and Data Rate Adaptation (DRA), from the information-theoretical point of view, is mandatory.

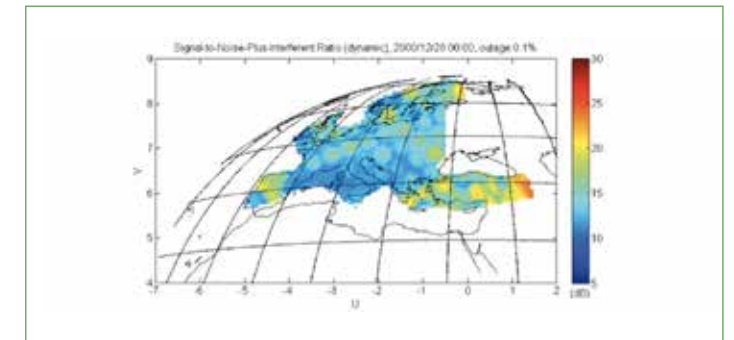
The choice of the most appropriate technique and its implementation is based on the intensive use of propagation models, able to mimic the quality of the radio channel with sufficient accuracy. In this respect, the models able to represent the spatial distribution of atmospheric attenuation over large (continental) areas and derived from the simple knowledge of basic meteorological information are of specific interest for both Broadcasting Satellite Service (BSS) and Interactive Multi Media Broadband Service (IMMBS) scenarios. For these application scenarios, important models are those able to provide the attenuation time series experienced by different

end-users spread over the service area. Useful models should simulate high resolution data in space and time based on available low resolution meteorological data. In addition, they should be able to reproduce with good accuracy first order and second order statistics (such as fade slope and fade duration) for multiple locations, including the correct spatial correlation of atmospheric attenuation.

This thesis addresses the problem of the joint use of two FMTs for fixed satellite communications: the on board reconfigurable antenna pattern coupled with ACM. The problem is analyzed in its entirety: we consider the use of reconfigurable antenna and ACM in conjunction with a developed model to simulate the propagation environment based on meteorological data. To the best of our knowledge, this is the first attempt to take into account all these techniques together.

A complex satellite system, characterized by a reconfigurable antenna front-end, able to spatially arrange power distribution according to the actual time-variant propagation needs will be referred to On-Board Dynamic Power Allocation (OBDPA) system. In such system, the

power is allocated only in the amount that is needed to support required QoS, thereby allowing management of the limited on-board available resources (Fig. 1). The antenna adaptation procedure relies on updated meteorological information, provided to the reconfiguration algorithm on a regular basis. On the other side, ACM is a sophisticated form of information rate control in which single user information rate is adapted by changing modulation type and coding rate, while maintaining a constant symbol rate. The aim of the ACM technique is to dynamically select, based on the current channel quality, the transmission Modulation and Coding (ModCod) pair which maximizes spectral efficiency, satisfying the required Frame Error Rate (FER) performance. In order to test the joint use of the above techniques, the attenuation information at single user level with high resolution in time and space is required. A new model has been developed to reproduce correct statistics of rain attenuation on the short and long-term basis. On the short basis, the simulated attenuation evolution should reproduce with the highest accuracy single rain events evolution (within 6-hours time interval), which is comparable to a communication session



2. Example of dynamic power allocation (OBDPA system) over the european service region. Signal-to-Noise+Interference Ratio (SNIR) with the beam excitation coefficients optimized to the simulated rain attenuation conditions.

duration; on the long term basis, resulting statistics should represent the sites specific meteorological conditions over a time period of a few years. Moreover, the need of considering multiple users not co-located and subject to spatially correlated attenuation conditions appeared evident. As input information, this model uses a large database of rain attenuation measurements available. Furthermore, the model is trimmed and refined to reproduce the spatial rain attenuation information which are provided to the OBDPA system. Finally, the model is used in a case study evaluation of the performance of the OBDPA system with and without ACM. The outcome of this study could potentially be applied to the following scenarios

(among others): 1) the design and analysis of the Satellite Communication System before its deployment, which requires the use of channel models and input data whose computational complexity must be traded off with the accuracy they provide; 2) the system operational stage, during which the model may be used to activate specific mitigation techniques. In this case, requirements are in terms of real time or about real time application, of ability to identify the actual state of the radio channel and of the confidence interval of the performances.



## INSTRUMENTATIONS FOR DIMENSIONAL MEASUREMENTS BASED ON SELF-MIXING INTERFEROMETRY

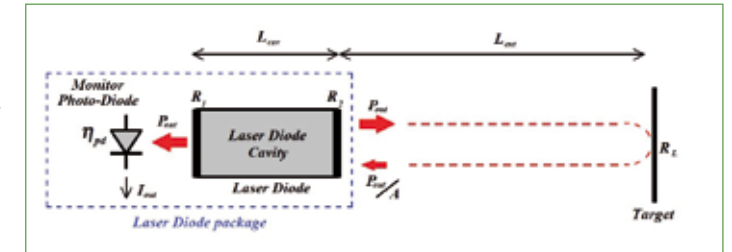
Alessandro Magnani - Supervisor: Michele Norgia

During the last years, self-mixing interferometry has been considered even more as a valid alternative to the classical optical instrumentations. Its characteristics such as contactlessly, low-cost and high-resolution, allow to achieve novel kinds of measurement systems dedicated to dimensional measures. The advantage in using self-mixing technology consists, in fact, in the good trade-off between low production cost and obtainable accuracy; these interferometers permit to reach sub-micron displacement measures with a very simple optical setup, constituted by the laser diode only, as shown in Fig. 1. Moreover there is no requirement for any external component and the classical optical configuration is not invasive, allowing to operate on sensitive and remote surfaces. Basing on self-mixing interferometry, the aim of this research work was to develop different instrumentations for dimensional measurements; in particular, four electronic prototypes functioning in real-time was studied. The first one is a vibrometer operating in open-loop mode, which allows spatial resolution down to about 100 nanometres over an undefined spatial range. As shown in Fig. 2, this instrument acquire the self-mixing signals and reconstructs

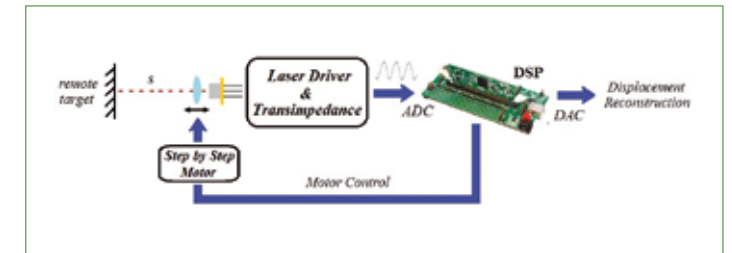
the remote displacements by means of two different algorithms; in way to increase the quality of the interferometric signal, an autofocus system based on step-by-step motor was implemented. The second instrument is another kind of vibrometer, based on close-loop technology, which guarantees spatial resolution down to few nanometres, high linearity, but operating on a dynamic range of about 100 microns. Substantially, the interferometric signal is injected into the laser current, causing the wavelength modulation and then (if the loop is kept stable) compensating the remote displacements. A third developed device is oriented to the measure of the speed of a target, based on a dedicated frequency-domain algorithm which analyses the phase-spectrum in real-time. This prototype demonstrated good performances for speed until 3 meters per second with a resolution better than one millimetre per second. The last developed device is a rangefinder (or a telemeter); the device takes the absolute distance measurement from the remote target, in the spatial range between 10 centimetres and 2 metres, with a resolution of about 100 microns. All the developed instrumentations have been studied, designed and finally implemented in a

real-time electronic device, providing the results on both an analog and digital output. Currently, these instruments are only prototypes and not yet distributed commercially: future works should be done in way to increase the robustness and the reliability of the measurements, but they represent now a starting point of a novel series of industrial devices. Typical applications range from the mechanical vibrations measurements to the monitoring of the positioning of objects in industrial processes; other applications are the measurement of the liquid level during the filling of bottles, or also the analysis of the micrometric eccentricity of pipes diameter. An important part of the PhD research was devoted to a novel application for the self-mixing interferometry: the measurement of the holes depth in ablating systems based on high-power lasers. In fact, the self-mixing phenomenon could represent a valid solution for many ablation processes, in particular texturing, where the measure of the depth of the laser excavation is necessary to monitoring (in real-time) the correctness of the process. At the present, many experimental measurements have been done with different target materials, power lasers and optical configurations. The main optical-

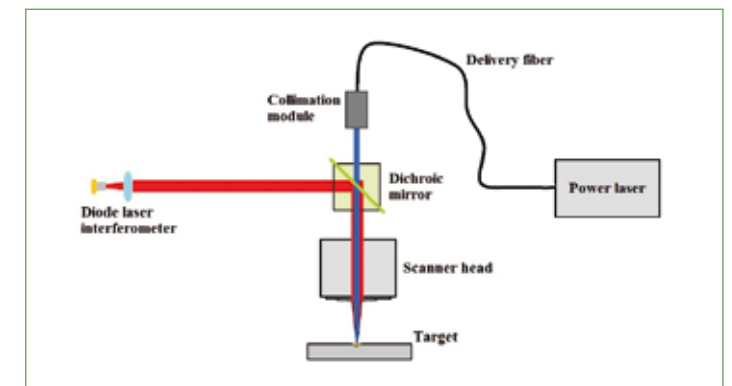
mechanical setup, shown in Fig. 3, consists in the overlapping of the high-power laser beam with the interferometric beam; ideally, it is possible to measure the progress of the hole depletion. The results indicate that self-mixing technology may be used to these purposes: the classical fringes signal is visible, indicating the depth of the ablation. In the future, however, it is necessary to better understand many spurious phenomena due to the evaporation of the ablated material so that the self-injection interferometry may be used for this kind of processes; further experimental tests are required and theoretical analysis before having a reliable measurement system. In conclusion, this Thesis work showed the great capabilities of the self-mixing interferometry in dimensional measurements contexts. Even if much other work should be done, the author believes that these instrumentations will have a future in the choice of a measurement system, if not will overcome other similar devices.



1. Optical setup of a self-mixing interferometer.



2. Simplified scheme of an open-loop vibrometer based on self-mixing technology.



3. Optical-mechanical setup for the ablation depth measurement of high-power laser systems.

# CHARACTERIZATION OF GROUND BACKSCATTERING IN PRESENCE OF VEGETATION THROUGH POLARIMETRIC SAR TOMOGRAPHY

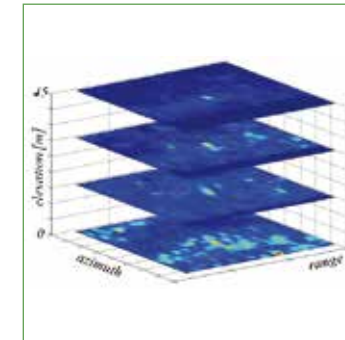
Mauro Mariotti D'Alessandro - Supervisor: Prof. Stefano Tebaldini

The impact of human activity on the global environment has dramatically increased in the last two centuries. The necessity of understanding the effect of the human development on the natural equilibrium led to many efforts in the recent years. On the one hand measures of Earth parameters multiplied; on the other hand, modeling attempts to describe the Earth dynamics relying on such measurements spread. Of particular interest in this scenario is the study of the above the ground biomass together its the relationship with carbon cycle and global warming. Synthetic Aperture Radars (SARs) working at microwaves offer a perfect tool for global mapping due to its capability of imaging day and night, with any atmospheric condition and due to the fast coverage of very large areas. In particular longer wavelength SARs enable to penetrate the vegetation layer up to the ground level. It follows that each depth inside of the vegetation and the underlying ground contribute to the radar signal. Thus the whole forest layer can be imaged, still refined processing chains are needed to interpret this data properly. The main processing issue is to separate the contributions coming from different depths inside of the vegetation layer as they reach the sensor mixed

up together. This goal can be attained by means of SAR tomography. SAR tomography enables to associate the proper contribution to the elevation where it belongs. An innovative way to perform tomography not relying on any a priori assumption is introduced in this thesis: it is referred to as *model free SAR tomography*. A novel tomographic calibration technique is presented also: it is called *phase center double localization (PCDL)*. These operational tools, together with many others known by literature, have been exploited to achieve a good characterization of the ground contribution when imaging a vegetated area. In particular the ground signal has been studied varying ground topography, forest structure and underlying lithology. Also an algorithm to estimate the digital terrain model beneath the forest has been presented: it is named *ALGAE*. Furthermore experiments have been carried out on areas covered by a much thinner layer of vegetation, a grass layer. In this case no dedicated processing tools have been developed, the feasibility of a lithological classification relying upon SAR features has been investigated. SAR tomography relies on the joint exploitation of several radar images of the same area. The availability of the phase delay

information enable to combine together the radar echoes in order to achieve resolution along the three spatial directions. For this process to be successful a very precise calibration of the image stack is needed. Usually man made targets can provide enough constraint to achieve a satisfactory calibration; such targets might be totally missing in case of a survey on a forest. The proposed calibration algorithm enables to jointly recover the uncompensated phase delay associated with each image (in terms of displacement of the sensor from its nominal position) and the elevation of the phase centers associated with each resolution cell. *PCDL* has resulted very effective in calibrating the image stack; it turned out to be the only way to perform a proper 3D reconstruction of the forest. Major advantages are the fast convergence time, the possibility to apply the algorithm to areas where man made targets are totally missing and the possibility to calibrate the data stack also when the ground cannot be reached. Whenever the penetration capability of the electromagnetic wave is good, the contribution coming from the ground is clearly visible. In this case a different calibration technique can be applied: the *sum of Kronecker product*

*decomposition (SKPD)*. This calibration technique has been published in the past years and applied in this work; a further processing step is presented here: it enables to estimate the digital terrain model under the forest by processing the calibration phases SKPD returns, it is called *ALGAE*. *ALGAE* algorithm explicitly takes into account the null space of the forward operator associated with an airborne SAR survey. It algebraically inverts the forward problem up to the null spaces and returns a full solution by weighting the null space basing on some a priori information. The effectiveness of this approach has been proven by comparing with the LIDAR estimation of the topography of the area. Standard calibration algorithm together with innovative ones has been used to achieve a good understanding of the interaction of the electromagnetic wave with the target vegetation plus ground. A wide experimental section is devoted to the exploration of the physical phenomena occurring when imaging a vegetated surface at microwaves. The goal of this section is to provide an insight of which quantities contribute to the radar signal and to what extent, which are the significant parameters that can be reliably inverted starting with the SAR



1. The Guyaflux tower as imaged through model free SAR tomography.

data. SAR signal has been found to be strongly dependent on ground topography; mainly because a flat ground allows the double bounce scattering mechanism to take place. The  $90^\circ$  angle made by flat terrain and tree trunk produces a strong echo masking the contribution of the upper canopies. SAR tomography enabled to quantify the tilt angle needed for the double bounce mechanism to fade. The issue of relating this threshold angle to a physical quantity of the forest arose. A simple model has been fitted and it revealed that this angle can be possibly related to a free path length in front of the trunk rather than to the tree height. Also model free SAR tomography has been exploited to neglect the strong contribution coming from the ground and to focus on the

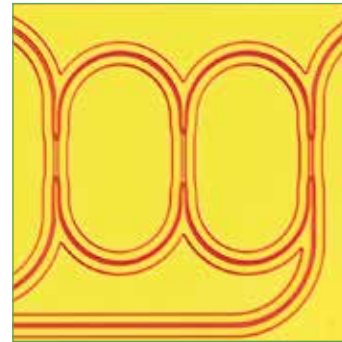
canopy to estimate biomass. The relationship between SAR polarimetric features and under the forest lithology has been explored. A ground truth map has been compared to several maps associated with many SAR derived quantities. A very poor agreement has been found in this case. The presence of understorey and the thickness of the vegetation layer likely masked the effect of underlying lithology. Also the connection between ground and canopy signal has been explored. The SKPD made it possible this kind of analysis as it relates together the three dimensional structure of the canopy to the polarimetry of the ground and vice versa. A strong experimental proof against many published models of ground backscattering in a forest has been found. A polarimetric analysis of the ground contribution in presence of a thin grass layer has been carried out. In particular SAR features have been compared to underlying lithology. This work is strongly innovative and it might be considered as the first attempt in this direction. The agreement with the ground truth map prevent to draw definitive conclusions.

## A DESIGN KIT PERSPECTIVE ON INP-BASED PHOTONIC INTEGRATED CIRCUITS

**Daniele Melati** - Supervisor: **Prof. Andrea Melloni**

Over the last decade, integrated photonics has experienced a large widening of its application fields. Photonics community has exploited the advantages of the integration of optical circuits for telecommunications, sensing, medical/biological applications, microwave filtering, long-wavelength devices. Among other, Indium Phosphide-based photonics emerged as one of the most promising and versatile technology for the integration of numerous different functionalities. The opportunity to grow active materials alongside passive layer stacks in a feasible and reproducible manner, enabled monolithic integration of complex circuits comprising sources, amplifiers and detectors together with passive circuitry as power splitters, waveguides or filters. As innovative components led to new applications and business opportunities, became clear to the photonics operators how a real market breakthrough for integrated photonics could be only related to the possibility to access a very large scale of integration, increasing the complexity of the optical circuits. On the other hand, it also emerged how a very small number of promising photonics devices resulted at the end in commercial products available on the market, highlighting an intrinsic limit in the methodology

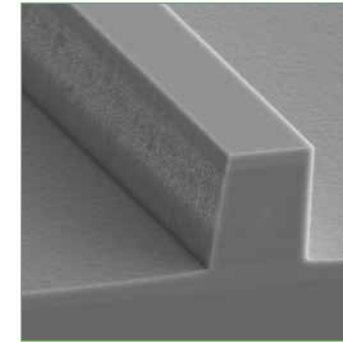
adopted for the development and fabrication of the circuits. Nowadays many groups work in the field, each one designing, producing and testing chips internally, grounding the whole chain on technologies developed expressly for application-oriented projects, generating a huge fragmentation. The potential market of each technology is too small to justify investments for development and industrial-oriented reconversion, suitable for high volume production and complex integration scale. Obviously, this phenomenon severely hampers the emergence of a real market for these products, limited to few business niches. What is needed is the availability of stable and cost-effective technological processes that can be exploited for the realization of photonic circuits. New methodologies are required in order to establish an organization of the photonic field which can be attractive for market players and encourage the implementation of photonic-based solutions for the development of commercial products. This goal can be reached exploring the approach followed by microelectronics long time ago and resulted in a wide diffusion of inexpensive electronic chips. The idea is to develop few standard integration technologies not



**1. Through generic foundries, the advantages of integrated optical circuits could be exploited in many application fields with reliable, inexpensive and high-performance photonic chips.**

oriented to a single high performance application but rather able to offer the most used components, optimizing their functionalities and guaranteeing their characteristics. If each single device exhibits a repeatable behaviour, their combination can likely result in circuits with higher complexity and better performances. A generic foundry (as a foundry adopting this manufacturing approach is commonly named) can open its process to a large group of external small users which share the same production runs and hence share the whole fabrication costs, resulting in a large reduction of the cost per chip.

This approach has been proposed for photonics in the last years and large efforts have been spent to set up this new organization of the production processes. However, key instruments are still missing in order to allow the users to easily access this new technologies. In particular the lack of Process Design Kits which help abstracting from the technological levels and focusing on circuit problems prevents nowadays the full exploitation of generic integration processes. The present work addresses this subject, currently of great interest, successfully developing the theoretical and experimental instruments which permit the realization of complete Process Design Kits, providing a clear demonstration of the effectiveness of this approach for the design and fabrication of integrated-optic circuits. The precise description of the optical characteristics of a component is one of the most important information included in a Process Design Kit because it permits to properly design a circuit relying only on fast circuit simulations. Advanced characterization methods are hence needed to collect foundry-specific data. A novel experimental technique based on coherent frequency-domain reflectometry is presented and theoretical developed. This technique allows to identify both localized and distributed reflective sections present inside a component with a resolution up to ten microns and to exploit them to extract the device parameters and to perform a complete tomography of the structures. Alongside with experimental



**2. Any Process Design Kit requires advanced theoretical models and foundry-specific data to describe a particular technological process. In this work, waveguide sidewall roughness (visible in the figure) is extensively analyzed, a difficult but fundamental aspect to consider for a realistic description of an integrated optical device.**

instruments, theoretical models are fundamental instruments for the realization of a Process Design Kit. Here, an innovative model for the description of the interaction of the light with sidewall roughness (which affects any kind of fabricated integrated component) is developed and demonstrated to correctly predict both backscattering and radiative losses. The proposed model is applied to both 2-D and 3-D laterally confined waveguides and compared to experimental data, finding a good match and providing a clear understanding of the physical phenomenon. A circuitual model is also developed to allow simulation of the roughness effects directly during the circuit design, including also the statistical properties of backscattering. Since part of the light radiated as consequence of sidewall roughness can generate optical cross-talk in nearby waveguides, an extensive

characterization of this aspect is carried out. Characteristics of the radiated power in terms of absolute value and phase behaviour are described. The last part of the work is devoted to the analysis and design of innovative circuitual solutions able to expand the capabilities of a generic foundry. A novel test-on-wafer technique is developed for a fast mapping of the optical properties of the waveguides immediately after the conclusion of the production process. Relying on the proposed reflectometric technique, spectral acquisitions of few Point Reflector Optical Waveguides allow to sample the properties of the waveguides on the entire wafer, providing precious information to the foundry. Lastly, a Two Mode Interference coupler is designed and characterized, demonstrating properties complementary to widely exploited MMI couplers. An actively controlled version is proposed for the design of an innovative wide-band continuously tunable optical true time delay line.

## THE R2P FRAMEWORK FOR ROBOT PROTOTYPING: METHODOLOGICAL APPROACH, HARDWARE MODULES, AND SOFTWARE COMPONENTS

**Martino Migliavacca** - Supervisor: **Prof. Andrea Bonarini, Prof. Matteo Matteucci**

Starting from the observation that robotics still fails to enter the mass market, we identify in the problem of robot prototyping one of the main limiting factors in today's robotic research.

Robotic applications need to be prototyped at an early stage of development, as their success depends on the interplay of a variety of components, involving several, different, competencies.

Today, the prototyping process can be even more demanding than developing the application itself: no standardized components for the development of generic robots exists, and systems are often implemented from scratch, even if they share most of the requirements.

Many unexpected issues often show up at early stages of development, slowing down, or even preventing, the translation of interesting ideas into working systems.

In this thesis, we present a novel approach to robot prototyping based on modular development techniques which have been recognized effective in many other fields.

The goal of the thesis is to extend the modular, component-based, approach to hardware level, developing hardware devices and programming tools

focused on robot requirements, to enable massive hardware/software reuse and fast prototyping.

By analyzing common robotic systems, we identified a set of functional requirements that we implemented at hardware level as distributed modules with on-board computation, each focused on a specific task.

Modules share a standard physical interface, with a simple connection schema, so that robot prototypes can be developed by selecting the modules that satisfy the identified requirements and by assembling them in a plug-and-play fashion.

Real-time interaction between the different devices is provided by RTCAN, a novel protocol for the CAN bus we developed starting from the requirements of distributed architectures for robots.

It combines the advantages of different approaches to communication scheduling: time-triggered communication, e.g., from control loops, is handled with a pure TDMA approach to guarantee high temporal determinism, while event-triggered communication requests, e.g., from sensors, are handled with EDF scheduling, to be delivered with low latency.

A lightweight publish/subscribe middleware extends modular development also to low-level control software, which is implemented by means of distributed, loosely coupled, nodes executed on the hardware modules.

In this way software components running on resource-constrained devices can be reused and shared through different projects, with a programming model most robot developers might be familiar with.

To easily integrate robotic platforms with high-level control software, we developed  $\mu$ ROSnode, a lightweight ROS client that can run on low-cost microcontrollers, publishing native ROS messages over TCP/IP.

$\mu$ ROSnode enables directly interfacing with the low-level control network, without the need for specific adapters and device drivers; in this way, the same interfaces are shared among different robot platforms, allowing to adopt the same ROS software in different projects.

The contributions of this thesis led to the Rapid Robot Prototyping (R2P) project, an open source hardware and software framework aiming at reducing time and effort needed to develop robotic systems.

With R2P, complex systems can now be implemented by assembling off-the-shelf components and easily programming their interaction, without the need for domain-specific knowledge in electronics and low-level control.

Individually, the components of the framework may also be used independently from R2P, and easily integrated in other projects, thanks to the open source development approach. The overall approach has been validated with some use cases to demonstrate the effectiveness of the proposed architecture to develop real applications.



## MODELS OF THE SPATIO-TEMPORAL DYNAMICS OF THE HIGH ALTITUDE ALPINE FAUNA IN THE CLIMATE CHANGE CONTEXT

**Andrea Mignatti** - Supervisor: **Prof. Marino Gatto, Prof. Renato Casagrandi**

High altitude Alpine regions are hotspots of biodiversity and are very sensitive to the occurring climate change, displaying a warming rate higher than the global average. The most evident response of Alpine species is an uphill movement towards higher elevations. Summit species are the most vulnerable because they cannot shift over the ridges or the perennial snow. This thesis's aim is to develop innovative models for the occurred and expected responses of high altitude Alpine fauna to the climate change. We developed both species distribution models and demographic models. Species distribution models describe the relationship between environmental variables and the suitability of a territory to host a given species. Temporal dynamics are instead taken into account in demographic models, which show how the abundance of individuals changes in time. In both cases, there is a need for appropriate methods that identify, from data, which environmental and/or climatic variables have an important influence on the spatial distribution and on the demographic parameters of the target species. The identification of the best predictive models has been carried out by using standard selection criteria. When the model selection

was uncertain, we relied on multimodel techniques to produce predictions. The three high altitude species under study, chosen according to their already known sensitivities to the environmental conditions, are: the Alpine marmot (*Marmota marmota*), a small mammal that hibernates during winter; the black grouse (*Tetrao Tetrix*), a tetraonid bird that lives just above the tree line; the Alpine ibex (*ibex ibex*), a long-living large mammal that uses several different habitat during the year. In the Alpine marmot case study, we investigated the fine scale species distribution of the species in an undisturbed high altitude valley near the borders of the Stelvio National Park (Italy). The positional data of the burrows were personally collected. The availability of a high resolution vegetation map permitted to study the fine scale influence of vegetation type on the suitability of the habitat for the marmots. We found, in fact, that the vegetation cover is one of the most important factors that determine the suitability of the habitat for the species. The most important topographical variables are related to (i) sun exposure, which can regulate both winter conditions and snowmelt in spring, and (ii) a burrow position favourable to defend against predators

and to avoid extreme weather conditions. As for the black grouse, we studied the influence of spatial position, population density and meteorological conditions on four demographic rates (growth rate and 3 components of fertility) that characterize the populations of 17 Alpine districts in the Piedmont region (Italy). Currently, few studies have investigated the influence of climate on the Alpine populations of black grouse, while most of the studies refer to lowland populations. Our results are mostly consistent with those published for lowland populations. In fact, high temperatures in the breeding season reduce the growth rate while the fertility decreases with the increase of rainfalls in the hatching period, when newborns are more sensitive to harsh conditions. The winter precipitations are likely to negatively affect the population growth, probably through a reduction of the food availability in the critical winter period. We also found that the direct (negative) density dependence is the main driver for growth rate. Moreover, our approach permitted to separate two groups of districts that are characterized by different carrying capacities. Future projections show that the variations in climate have the

potentiality to strongly affect dynamics, even if no clear future trends are detected.

Alpine ibex populations are characterized by a strong age and sex structure that has never been considered in a population dynamics model. Past studies show that the main driver of population growth rate are the population density and the accumulation of snow in winter, while survival and fertility are not constant with the age, but are typically smaller and more variable for the youngest and the oldest individuals. Using the Gran Paradiso National Park (Italy) population data, we developed models that, alongside with density and snow depth, take into account the age and sex-structure of the population at different levels of complexity. First, we considered four population groups, as reported in the population counts. We relate the growth rate and the demographic rates (adult male survival, adult female survival, weaning success and survival of kids) to the linear and nonlinear effects of snow depth and population density. Ibex population is divided into spatially segregated groups composed by individuals of the same sex; we found that the demographic rates related to a specific population group are more likely to be affected by the density of that group only. The values of all the demographic rates are clearly determined by the interaction between population density and snow. Moreover, we found that the nonlinear effects of snow depth are important for the juvenile compartments, which are favoured by intermediate levels of snow. However, considering

the population groups instead of the total population does not satisfactorily explain the occurred dynamics, especially the recent population decline. We therefore developed more sophisticated models that permit to reconstruct the complete age structure of the population. Using those models, we studied the effects of maturation time and senescence to determine the population dynamics. Our results underline that taking into account senescence is particularly important to explain the survival of adult females and their ability of breeding kids, while is less important for adult males and weaning success. Moreover, harsh environmental conditions are likely to enhance the senescence rate, and, thus, their effects are stronger in old individuals. The simulation of the population structure shows that the number of old individuals increased in the last 25 years of the study, at the expense of the fraction of young adults. The stability in the fraction of middle age adults, which are characterized by both a high survival and a high weaning success, can be interpreted as a strategy to damp the short term effects of harsh conditions. The methodologies adopted in this thesis are a sound solution to tackle the problem of studying the spatial distribution and the demography of the species in the climate change context. Our models permit in fact to investigate the role of climate-sensitive variables, such as the status of the vegetation and the meteorological variables, in affecting the spatial and temporal dynamics of the case-study populations, taking also into account the peculiar

characteristics of each species. This kind of analysis is crucial for conservation purposes for several reasons. Without precise modelling studies that include the peculiarities of the species it is in fact difficult to predict and understand the expected impacts of climate change on the populations. Indeed, using the species specific results to make generalizing conclusions can be misleading because each species has different characteristics and different vulnerabilities. Our results highlight which are the most critical areas (for marmot and grouse), periods of the year (for grouse and ibex) or periods of life cycle (for grouse and ibex) on which is worth to put more effort in future studies. On the other hand, a lot of work is still needed in this field, in particular to include the interactions between species, which is certainly of paramount importance to explain their spatial and temporal distribution.

## REQUIREMENTS VERIFICATION OF VARIABILITY-INTENSIVE SYSTEMS

**Amir Molzam Sharifloo** - Supervisor: **Carlo Ghezzi**

### Problem Formulation:

Modern software is increasingly complex. That is why it is often designed and implemented through an iterative and incremental process. The development process is conducted in a series of iterations, through which a high-level design is refined into detailed specification and finally in machine-readable code. As the result, a large-scale problem is gradually tamed while the evolution is carried out by elaborating and implementing functionalities of interest. Similarly to iterative-incremental development, the research on adaptive systems deals with some kind of evolution triggered by the need to adapt and modify functionalities at run time. Techniques to develop adaptive systems continuously monitor the changes in the operational environment of a software system in order to adapt their behavior and satisfy the requirements. Adaptation is often realized through introducing some degree of dynamism into the system such that it switches among different configurations depending on the changing environment. There exist different techniques to design and develop adaptive systems. In most cases, the ability to adapt is embedded via adding variation points which can be realized by different

alternative implementations. For example, a concrete service realizing an interface can be rebound at run time following a service-oriented architecture. This implies that system specification may dynamically evolve over time. Research on Software Product Lines (briefly SPLs) is another research direction where variability and evolution play the central roles. Relying on planned modularity and reusability, and techniques to introduce variation points and alternatives, engineers derive a large number of different products from a shared SPL specification. In this case, the specification contains all the functionalities while each product takes only a subset. Although the research on iterative-incremental development, adaptive systems, and SPLs are conducted by different communities, they are linked via the common concern of variability. That is why we take the name of Variability-Intensive Systems (briefly VIS) to refer to the system specifications produced in these disciplines, from which a large number of systems can be derived. In particular, we are interested to guarantee that a VIS specification satisfies its requirements. This is drastically crucial for safety-critical systems, where any violation may lead to expensive

penalties. This urges to employ verification techniques able to check specifications before any real implementation and execution. Formal verification, in particular model checking, has interestingly progressed through the last two decades and has been successfully applied in industry. However, there are two main obstacles to apply existing model-checking tools. First, they are not able to deal with incomplete specifications, that may be produced through early stages of development when there are not enough information to construct the final specifications. Moreover, model checking is an expensive task in terms of time and computation since it exhaustively explores the state space of specification. Running a model checker frequently for any change may not be economically reasonable and timely feasible. Hence, there is a need for cheap model-checking techniques dedicated to check evolving specifications. Regarding model checking to SPLs, requirements verification of individual products can be quite expensive. This is due to the large number of products that may be derived from a single SPL.

### Contributions:

The contributions of this thesis target the verification of variability-intensive specifications. However, we study this problem in two different areas, and this is the reason that the thesis is organized in two main parts. In both cases, we aim to develop model-checking techniques capable of handling incompleteness and variability in specifications. To reach efficiency demands, our techniques enjoy the reuse of verification results for invariant parts of VISs. The thesis initially focuses on the techniques to verify an incomplete specification against temporal properties, that are appropriate to express safety and liveness requirements. We choose Computational Tree Logic (CTL) as our formal property language due to its power in expressing temporal properties. However, the general idea of our approach is extensible to other temporal languages e.g. LTL. We motivate the need for verification of temporal properties for incomplete specifications through a running example Secure Information Retrieval. We introduce a new formalism Incompletely Labeled Transition Systems (ILTS), as a new variation of Labeled Transition Systems (LTS), to capture incomplete specifications. A new verification

technique IMC is presented to verify ILTS against CTL, which is able to deal with unspecified components. Unlike the standard CTL model checking that is unable to handle unspecified components, IMC is able to produce constraints for such components to guarantee the satisfaction of a property. We study the scalability of IMC by running our prototype implementation on the enlarged versions of SIR system. Moreover, we show how IMC can be applied to efficiently check adaptive systems. Since ILTS formalism is rather low-level, we extend our approach to Statecharts that are well-known to model a system behavior within a user-friendly and compact notation. The new technique is able to verify incomplete Statecharts, and by increasing the verification reusability it alleviates the time to re-verify slightly modified specifications. We discuss the applicability of the approach through the classic Railway Cross case study, and report on its performance. We later focus on modeling stochastic behaviors of SPLs, that we use to check a set of non-functional properties, namely reliability and energy consumption. We initially focus on scenario-based specifications, in which augmented sequence diagrams linked to feature

diagrams and enriched with stochastic information are used to represent a stochastic SPL behavior in a compact form. To support the stochastic verification of SPLs, we introduce a new variation of Markov models - FDTMCs - that proposes the notion of features as the first-class concept. We describe three different techniques to verify FDTMCs against two stochastic temporal logics: PCTL and Reward logic, that are popular to express reliability and energy consumption properties. We study efficiency of the approaches through two case studies.

## DEVELOPMENT AND APPLICATION OF A TEST SUITE FOR RADIATION DETECTORS OVER A WIDE RANGE OF CHARGE INJECTION LEVELS

**Giuseppe Vito Montemurro** - Supervisor: **Prof. Andrea Castoldi**

New techniques aimed at the direct characterization of the response of semiconductor detectors and of the associated front-end electronics are an essential factor for the successful development of a new generation of detection systems compatible, for example, with the present and future high brightness Free Electron Laser sources of X-rays because of their unprecedented beam properties which demand to the detector side the ability to handle high charge levels (up to  $10^7$  electron-hole pairs per pixel), a dynamic range up to 1:10000 and ultra-fast readout speed. Even higher levels of charge generation are expected in the planned detection arrays for the upcoming nuclear physics experiments which require suitable techniques and instrumentation for calibration and diagnostics in such non-standard operating conditions. To this aim we investigated two techniques, based (i) on the use of mono-energetic proton bunches, available at the DEFEL beam-line at LABEC (Laboratorio di tecniche nucleari per i Beni Culturali), Italy, for high levels of charge injection and (ii) on a table-top pulsed IR laser system, suitable for low and medium injections levels.

Laser sources are an efficient tool for the characterization of detector response and in

particular for the possibility of mapping the relevant properties of radiation detectors with micrometer resolution. The main advantages are intrinsic capabilities of accuracy in beam positioning and in time synchronization, finely tunable intensity. The laser source has also some disadvantages; it can be difficult to calibrate the laser intensity for all the available intensity range. In fact, optical reflection losses and refraction at the considered wavelength can affect the spatial distribution of the generated charge within the detector volume. However, at low levels of charge injection, absolute calibration of the intensity can be gained with direct comparison with the response to calibration radioactive sources. As an attractive tool to probe the response of the detector-front end system at higher levels of charge injection with high precision in space, time and for different ionization profiles across the detector depth, the possibility of using a pulsed mono-energetic proton beam has been exploited. The protons projected range in silicon is perfectly matched with the typical thickness of silicon detectors in the energy range 1-6 MeV. The ionization profile of protons can be tuned from shallow ionization ( $16 \mu\text{m}$  at 1 MeV) to an extended ionization

( $295 \mu\text{m}$  at 6 MeV) along the wafer thickness. Finally the radial range, which sets the ultimate achievable position resolution, is limited to  $9.5 \mu\text{m}$  at 5 MeV. By increasing the number of protons in a bunch, it is possible to probe the detector at increasing levels of charge injection. These levels correspond exactly to multiples of the generated charge by a single proton; therefore the total generated charge is intrinsically calibrated. Being the time duration of a proton bunch (i.e. the spread of the times of incidence of individual protons in the same bunch on the detector surface) much smaller than the detector charge collection time, the detector output amplitude is proportional to the bunch multiplicity. Protons are hence suitable for probing the detector and front-end electronics at high, and precisely calibrated, levels of charge generation with high spatial resolution and tunable ionization profile. In addition to the optimization and expansion, with the introduction of new instrumentation, of the pulsed IR laser facility, the activity has been focused on the improvement and qualification of the performance of the DEFEL proton beam line. We upgraded the beam line spatial resolution of the proton beam with novel remotely controlled in-vacuum hi-resolution profiling slits and a

CMOS imager has been installed as 2D beam monitor. Up to now a beam spot down to about  $60 \mu\text{m} \times 40 \mu\text{m}$  has been achieved and potentially can be further improved. The time jitter of the proton bunches is better than 0.5 ns opening the way to probe also the transport dynamics of the charge carriers with accurate time resolution.

A great effort has been also devoted to the implementation of a novel general purpose multi-channel data acquisition (DAQ) system, 500 Msample/s 12 bit, capable of full shape digitization of the output signals coming out from the devices under test. A general purpose multichannel system is versatile and suitable for the different foreseen experimental situations, especially the case of charge sharing among adjacent detector channels, which may be more noticeable at high levels of charge generation. Moreover the acquisition of the full shape of the output signals allows gaining deeper knowledge of the detector behavior, for example extracting the real current pulse at the detector anode (directly correlated with the charge cloud) by means of a dedicated deconvolution algorithm. The system is based on SIS3350 sampling boards, produced by Struck Innovative Systeme GmbH, featuring 4 channels, 12 bit, 500 MSample/s. The boards are connected via VME Bus, while a SIS3100 bridge acts as VME-PCI interface for the communication with the host pc via optical fiber. The dedicated acquisition software, written in C language, has been developed in order to be as user friendly as possible and, at the same time, to give

to the user the possibility of fully exploits the features offered by the sampling boards, in order to adapt the data acquisition system to different experimental conditions. Essentially, the software can be described as consisting of three sections: the configuration of the system, the acquisition management and the data transfer-storing. We present the development of the DAQ system, including the arrangement of the hardware, the development of the dedicated software and the functionality tests carried out with the system developed so far. Up to now a maximum data transfer speed of 30 MByte/s is provided. In the future the system, limited to 8 channels by the available hardware, can be easily expanded with additional sampling modules. This can be achieved thanks to the care taken in the development of the software, which is able to manage a system with an arbitrary number of boards (with reasonable limitations, for example the slot capacity of the VME crate). We probed the test suite through qualification campaigns on two relevant case studies. The first one is the characterization, with monochromatic protons, of the Double Sided Silicon Strip Detectors (DSSSDs) used as first and second detection stages of the telescopes that constitute the FARCOS (Femtoscope Array for COrelations and Spectroscopy) array. The tested detection system required an exhaustive study of the amplitude response at high levels of charge injection and, because pulse shape analysis must be performed,

of the induced signals shape as a function of the injection position. The second case study consists of a detailed qualification, carried out at low and medium levels of charge generation with the laser test suite, of a pixel matrix with DePFET (Depleted P-Channel Field Effect Transistor) readout, that is a first prototype of the DePFET-based pixel detector developed by the DSSC consortium (DePFET Sensor with Signal Compression), a collaboration aimed at the development of a large area 2D X-Ray imaging pixel detector for the new European XFEL. This characterization provides a deeper understanding in the properties of DePFET-based detectors and offers a detailed data-set useful to predict the detector response as a function of operating conditions. A test of the DePFET matrix has been carried out also with mono-energetic 1 MeV protons at the DEFEL beam line (not yet upgraded) to probe its amplitude response. Due to the limited spatial resolution of the proton beam at that moment it was not possible to perform amplitude calibration and in-pixel mapping, which would be now possible in the upgraded beam line.

## ALGORITHMS FOR THE VERIFICATION, COMPUTATION AND LEARNING OF EQUILIBRIA IN EXTENSIVE-FORM GAMES

**Fabio Panozzo** - Supervisor: **Prof. Carlo E. Fiorini**

The study of strategic-interaction situations has recently received increasing attention in artificial intelligence with the aim of designing autonomous software agents able to act optimally. These situations are customarily modeled as games in which the mechanism describes the rules and strategies describe the behavior of the agents. Particular attention is focused on the study of formal methods to theoretically guarantee the optimality of the agents' behavior. Game theory and microeconomics provide mathematical tools to model strategic-interaction situations and characterize the appropriate solution concepts. However, they do not provide computational tools to find solutions. This problem, commonly called equilibrium computation, is instead central in computer science, whose aim includes assessing the complexity of finding an exact or approximate solution, designing exact or approximate algorithms, and evaluating the application of the algorithms in practical settings. In this thesis, we focus on non-cooperative general-sum extensive-form games. Extensive-form games are more suitable to represent real-world situations providing a richer representation than strategic-form games, where the sequential structure of decision-

making is described explicitly and each agent is allowed to be free to change her mind as events unfold. While in zero-sum extensive-form games the main problem is to solve games of large dimension (e.g., poker), in general-sum extensive-form games there are other important issues. The most important is the choice of the most suitable solution concept: it depends on the scenario modelled by the game, e.g. the agents' common knowledge, and it depends on its computability, i.e. the amount of time and memory necessary to solve the game. We can identify two main scenarios, characterized through the knowledge available to agents: the case where the agents have common information and the case where they have incomplete information. Interestingly, with common information, the appropriate solution concepts for extensive-form games refine the concept of Nash equilibrium, while, when agents learn, the appropriate solution concepts relax the concept of Nash equilibrium. More precisely, in the first scenario, it is well known that the Nash equilibrium is not suitable, allowing strategies to be non-sequentially rational. The game theoretic literature provides a number of refinements of Nash equilibrium for extensive-form

games that are the appropriate solution concepts. For these solution concepts the verification and computation problems can be hard. In this thesis we show that for some solution concepts the verification problem is easy and we design algorithms to find some refinements of Nash equilibrium. When the information is not common every agent needs to have a set of beliefs over the opponents' behaviour. These beliefs are updated during the learning process, i.e. while the agents repeat the game. In this thesis we discuss the suitable solution concepts for this scenario and we develop efficient evolutionary game theory techniques and algorithms that work with the sequence form representation of extensive-form games allowing an exponential reduction of time and space w.r.t. the algorithms presented in literature. Finally, we experimentally evaluate each presented algorithm.



# SAFE AND EFFICIENT DYNAMIC UPDATES OF DISTRIBUTED SOFTWARE SYSTEMS

Valerio Panzica La Manna - Supervisor: Prof. Carlo Ghezzi

Modern software systems are subject to continuous, often unanticipated changes. Changes may occur in the environment in which the system is operating, or in the requirements, when new functionality is added or the existing functionality is modified. To deal with these changes software systems need to continuously evolve.

Software evolution yields specific updates, upgrades, or software patches that are typically managed and installed *off-line*. That is, the system is shutdown, updated, and restarted. However many advanced technical systems in transportation or production, but also information systems in commerce, financial transaction processing and healthcare, often have to operate continuously while their environments change or their requirements evolve. It is often expensive and impractical to shut down these systems in order to perform software updates, so the software needs to be *dynamically updated*, at runtime, while the system is operating.

Compared to off-line maintenance, *Dynamic Software Updates (DSU)* are much more difficult since in addition to the correctness of the new version, they must also preserve on-going activities. Especially in critical applications, updates must therefore be *safe*, i.e. they must not lead the system to

unexpected erroneous behavior. At the same time, dynamic updates must also be *efficient*. They should minimize the interruption of system's service (usually called *disruption*) and improve the *timeliness*, i.e. the delay with which the system is updated.

## The problem

Research in dynamic software update and dynamically updating systems mainly focuses on two different levels of abstraction. A dynamic update can be performed to support the runtime evolution of a system at *architectural level*. Software architecture defines the components of the system and how they are interconnected and its runtime evolution is typically supported by *dynamic reconfigurations*. A dynamic reconfiguration consists in unlinking at runtime the current version of the component, replacing it with its new version, and linking the new version to the rest of the system. Dynamic updates have been intensively studied also at the program level. Approaches at the *program level* allow for a running process to continue the execution with the new version while it is active. In these approaches the update is performed immediately without waiting that a consistent state is reached.

An architectural-based approach has the advantage of a clear separation of concern, and rely on fully automated approaches. However, it is based on restricting safety criteria that also affect the efficiency of the approach.

Approaches at the program-level, are based on weaker criteria for safe updates and are more efficient. However, the update process is embedded in the source code and must deal with the specific implementation details of the systems. Moreover, they require a non-negligible and error prone human intervention. The property of abstraction, and safety of a dynamic update are in trade-off with the efficiency and the human intervention required by the approach. Existing approaches have not yet found a proper balance between these contradicting properties and this lack calls for further investigation.

This thesis provides a contribution to fill this gap by addressing the following research questions: (1) Under which condition software can be safely and efficiently updated at runtime? (2) How can this condition be *automatically* derived?

## Contribution

In order to address the above problems, the goal of this thesis is to propose novel

criteria for the dynamic update of distributed systems. New approaches are presented to relax the restrictions of previous research, and *automatically* identify the states of the system in which these criteria are satisfied. Both the criteria and the approaches are independent of the specific programming language and abstract away from the implementation details. A prototypical implementation of these approaches is provided by means of runtime environments supporting the dynamic update, and of tools supporting the design of dynamically updating systems.

The contribution of this thesis is twofold. First, a new approach is proposed for the dynamic reconfiguration of component-based systems. Second, an approach is presented for the dynamic update of software-intensive systems based on the changes in the system's specification.

## Version-consistent dynamic updates of component-based distributed systems

This contribution considers the problem of supporting the runtime replacement of components executing *transactions*. A transaction is a sequence of actions executed by a component that completes in bounded time. Actions include local computations and message exchanges. A transaction can be initiated by an outside client or by another transaction executed by another component. In this setting, the global functionality of the system is represented by a distributed transaction that contains all the transactions locally executed by its components. In order to ensure

a safe update of components, the *distributed transactions* that the system is currently executing need to be preserved.

Version consistency of distributed transactions is proposed as a safe criterion for dynamic updating. Version consistency ensures that distributed transactions be served as if there were operating on a single coherent version of the system despite possible updates that may happen meanwhile. A distributed algorithm is introduced to maintain dynamic dependencies between components at architectural level and enable low-disruptive version-consistent dynamic updates. Dynamic dependencies model the past and possible future interactions between components. The approach relies also on the software architecture of the system to identify the static dependencies. The combination of these two models defines a more fine-grained condition for an efficient and safe dynamic update.

An implementation of the proposed approach is provided for the dynamic reconfiguration of SCA-based (Service Component Architecture) distributed systems. The approach has also been extended in the context of Service Oriented Architectures (SOA) to manage the version-consistent runtime evolution of service-based system in which each partner service is implemented as a BPEL process. These implementations, in conjunction with experiments, show the benefits of version consistency with respect to previous approaches both in terms of timeliness and low degree of disruption.

## Specification-driven dynamic updates of software-intensive systems

This contribution focuses on dynamic updates of software-intensive systems, which consists of many distributed components fulfilling complex functionality by reacting to events coming from the environment. A specification describes what are the assumptions about the surrounding environment and how the system has to react. Despite the existing techniques for developing dynamically updating systems, thus far almost no attention has been paid to defining when updates are correct with respect to a changing specification, i.e., when a system can safely disregard its current obligations and change its behavior to satisfy the new specification.

This contribution elaborates a formal definition for correct updates of a current implementation with respect to specification changes. Based on the specification, different criteria are defined to identify states in which the system can be safely updated. Moreover, the contribution introduces an approach for synthesizing a dynamically updating controller from the current implementation and changes in a scenario-based specification that updates to the new behavior as soon as possible.

The approach has been implemented as part of a tool-suite, called *ScenarioTools*, which support the engineer in the design of dynamically updating systems that can be safely and efficiently updated at runtime.

# AUTOMATIC MODEL SIMPLIFICATION FOR CONTINUOUS AND DISCONTINUOUS SYSTEMS

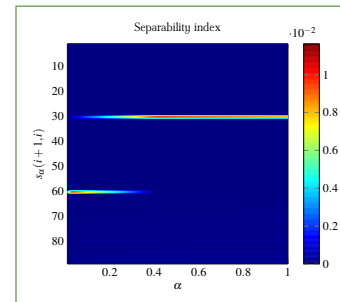
**Alessandro Vittorio Papadopoulos** - Supervisor: **Prof. Alberto Leva**

In the last years equation-based languages straightened and simplified the way sophisticated models of complex physical systems are built. In particular, the object-oriented modelling paradigm allows to obtain such models in an affordable way by focusing on the development of single “building blocks,” i.e., the objects, and connecting them so as to obtain the overall description of the physical system. This often results in large and complicated models, hard to simulate and in general to manage. The aim of the research path to which this work belongs, is to provide methodologies and automatic techniques to cope with said complexity, in order to simplify and streamline the model analysis and simulation.

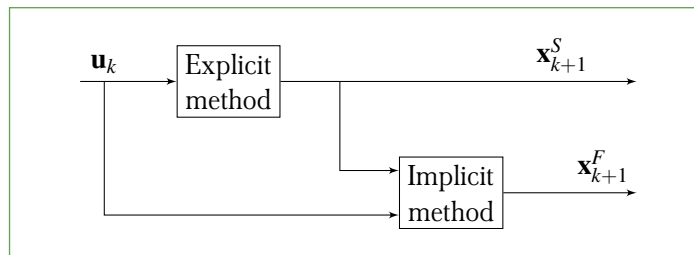
In the context just sketched, the thesis proposed several contributions. The first one is the development of a simplification framework which includes most of the model manipulation techniques available in the literature and novel ones. Functional to this contribution is the proposal of a technique called “Cycle Analysis,” able to perform a structural analysis of a DAE system, and return a dependency graph representing the way the dynamic variables are interacting, associating with each dynamic variable a time scale. Such information can be

used to ease and automate the partition of a complex model into decoupled subsystems, also with a parametric analysis (see Fig. 1).

Novel indices are introduced to characterise some structural properties of the system, like its stiffness, and to quantify “how much” a system is suited to be partitioned in the sense above. Also, the results of



1. Parametric cycle analysis performed on a thermofluid system.



2. Mixed-mode integration scheme for simulation efficiency.

the Cycle Analysis are used for improving the simulation efficiency by means of mixed-mode integration methods, represented in Fig. 2, and of a co-simulation (multi-rate) architecture, showing the effectiveness of the approach on some applications of interest.

In addition, an extension to classical model order reduction techniques conceived for continuous systems is here proposed for hybrid ones,

focusing on the switched affine case.

As a result of the above contributions, and moving from the methodological to the technological side of the addressed matter, the integration of the proposed simplification framework in state-of-the-art modelling and simulation tools is here considered, providing a viable and complete solution.

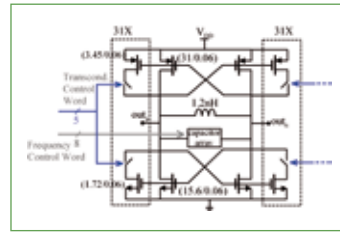
# ANALYSIS AND MINIMIZATION OF FLICKER NOISE UP-CONVERSION IN RADIO-FREQUENCY LC-TUNED OSCILLATORS

Federico Pepe - Supervisor: Prof. Andrea L. Lacaita

The world cell phones market experienced a steady growth in latest years, driven by the increasing number of applications available on smart-phones and tablet computers. Today's portable devices integrate multiple communication capabilities with ever-increasing data-transfer rates. In this scenario, the implementation of single-chip radio transceivers capable of operating over multiple standards is of great interest. The need for a higher integration level to reduce board size and cost led complementary metal-oxide-semiconductor (CMOS) processes to fast become the technology of choice in radio-frequency (RF) integrated circuits (IC) development. However, to take full advantage of the switching characteristics of MOS transistors, a digital approach started to be adopted in the RFIC design. On the contrary, the analog section of transceivers must cope with the limitations imposed by the adoption of scaled CMOS processes, for example the large flicker noise corner frequencies. In this context, the design of local oscillators for signal (de) modulation is becoming a highly-demanding task. This is due to the need of oscillators

with an increasingly-broad tuning range to comply with different radio-communication standards. In the traditional design approach, however, this generally leads to a non-optimum sizing of the active devices of the oscillators. This further exacerbates the impact of flicker noise, eventually resulting into unacceptable phase noise performances. The main purpose of this doctoral thesis is to provide a detailed quantitative analysis of the flicker noise up-conversion mechanisms, filling the gap towards a complete understanding of this phenomenon. A summary of the most important original contributions of this thesis is given below. Regarding flicker noise up-conversion in voltage-biased oscillator topology, it is demonstrated that modulation of harmonic content is the main cause of  $1/f^3$  phase noise generation. This effect has been quantitatively assessed and linked to the oscillator non-linearity. In particular, it has been shown that two mechanisms play a fundamental role in the conversion of flicker noise into phase noise:

- the direct injection into the tank of PM tones since the current through active devices lags with respect to the voltage



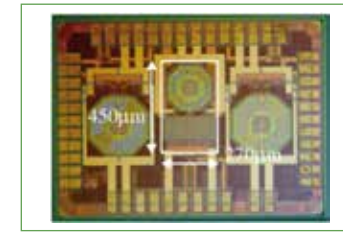
1. Improved voltage-biased oscillator.

- an AM-to-PM conversion effect due to the dependence of the oscillation frequency on the harmonic content of the output voltage. Design equations have been derived which highlight that flicker noise up-conversion is reduced in case of low excess gain, translating into low distortion of the voltage output. This suggested the implementation of an improved oscillator topology shown in Fig. 1. The up-conversion can be mitigated by adopting resistors in series to the drain nodes of the transconductor transistors. The resistors, together with the parasitic drain capacitances, introduce a delay in the loop gain shifting both the ISF and the current waveform of the transistors. It follows that  $1/f$  noise up-conversion can be reduced by judiciously tailoring the component values without degrading the start-up margin or adopting resonant filters.

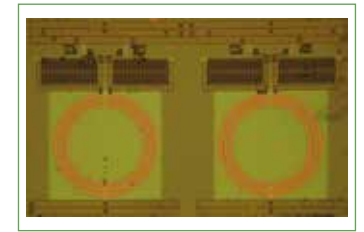
A theoretical explanation and a quantitative analysis have been carried out, addressing in details the different effects. The analysis has been verified by means of experimental measurements on a 65-nm CMOS oscillator.

Regarding the simulation of the impulse sensitivity function (ISF), a fast and accurate method based on periodic transfer function analysis for the impulse phase response in oscillators has been presented. The proposed technique overcomes the traditional simulation method based on transient analysis in terms of both computation time and accuracy. In particular, the method is helpful every time the magnitude and phase of each ISF component is of interest.

An alternative solution to reduce the flicker noise up-conversion has been presented, shown in Fig. 1, where a segmented transconductor is adopted in order to keep the excess gain almost constant and close to 2 as the frequency spans over the broad 1.6-to-2.6-GHz interval, thus breaking the conflicting link between tuning range and  $1/f^3$  phase-noise performance. The circuit has been fabricated in a 65-nm CMOS technology and the microphotograph of the die is shown in Fig. 2. A



2. Die microphotograph of the fabricated oscillator.



3. Die microphotograph of fabricated current-biased oscillators with two different widths of the switching transistors.

reduction of 10 dB of the  $1/f^3$  phase noise over a 47% tuning range was demonstrated, with no impairment the  $1/f^2$  phase noise performance. Moreover, measurement results are in good agreement with simulations, thus proving the validity of the proposed phase noise model. Regarding flicker noise up-conversion in current-biased oscillators, it has been shown that the  $1/f^3$  phase noise is mostly due to an AM-to-PM conversion arising from modulation of non-linear parasitic capacitances of the transconductor stage. A quantitative insight into this phenomenon has been carried out, highlighting that the AM-to-FM sensitivity term associated to the tail capacitance is the dominant contribution to the overall sensitivity. Moreover, it has been clarified why the AM-to-FM sensitivity

is null for a particular value of bias current, which can be adjusted by the circuit designer by adopting a sufficiently small transistors width. Measurements results carried out on the test chip in Fig. 3 confirm a reduction of flicker noise up-conversion of about 7 dB in case of smaller width with no impairment of the  $1/f^2$  phase noise.

Regarding the validity of phase noise model in current-biased oscillators, a further investigation has also been carried out in order to explain the discrepancy between measured and simulated values of  $1/f^3$  phase noise. A revised model of flicker noise up-conversion was proposed, which captures the effect of partial correlation between noise side-bands by taking into account the non-instantaneous response of noise to variation of bias conditions.

## END-USER DEVELOPMENT OF MASHUPS: METHODS, COMPOSITION PARADIGMS AND TOOLS

**Matteo Picozzi** - Supervisor: **Prof. Maristella Matera**

With the Web 2.0 revolution, new technologies, new standards and new application models have been introduced in the Web scenario. The Web has become more mature and full of potentialities as a platform for the development of interactive rich applications. The use of client-side scripting languages, the diffusion of Web Services and public APIs, and the always increasing basic skills of laypeople in the development of Web applications shaped up a scenario in which a new class of web applications, the mashups, was born.

Mashups integrate, at different levels of the application stack, data, functionality and user interfaces from different resources such as Web Services, public APIs or enterprise databases. Mashups emerged in response to the need of users, not necessarily experts of technology, to quickly assemble Web resources to create new Web applications solving their situational needs. One characterizing feature of these applications is that they are very often developed by the end users, i.e., people who actually need the final application. To accommodate this practice, which can be fruitful in several situations where the possibility of constructing applications satisfying specific

needs is required, in the last years different tools, conceived to offer intuitive composition languages, have been proposed. Unfortunately, most of such tools, after a couple of years of activity, were dismissed.

My PhD thesis aims at investigating and defining a framework that includes models, composition paradigms and tools for the End-User Development (EUD) of mashups. Two main reasons for the failure of the mashup tools so far proposed are indeed their incompleteness with respect to the users needs and their difficulty of use. The framework defined in this thesis aims therefore at covering the most salient activities in a mashup life cycle, and proposes a new composition paradigm based on abstractions that try as much as possible to hide technical details. The end users are enabled to integrate data of diverse resources, to create components that could be used in a mashup composition, to generate mashups that can be deployed on different kinds of devices (e.g., mobile devices or multitouch screens). We have also investigated collaboration mechanisms to allow groups of users to share resources and co-create applications. This last feature is particularly fruitful to promote the potential

of mashup composition as a paradigm for knowledge sharing and creation.

More specifically, this thesis introduces the following contributions:

(i) **User-driven development process.** We argue that the user-driven creation of mashups is more challenging than the provider-driven development of services. Firstly, for non-expert users it is desirable to have a mashup composition paradigm that hides the back-end complexity and simplifies the data, service and UI aggregation process through mechanisms as close as possible to the front-end organization, i.e., the “appearance” of the final application that the user would really experience. Therefore, our approach is strongly characterized by a visual composition paradigm allowing users to integrate contents into unified visualizations and synchronize the behavior of the resulting UI components based on an event-driven logic. The users manipulate exactly the data and the visual elements they will see and interact with in their final applications. A WYSIWIG style indeed allows the users to immediately get realistic examples of the data that will populate the visual elements and the way contents will be fused into unified visualization

and synchronized with other resources. In other words, the composition method is based on an “integration-by-example” paradigm, where the users can immediately observe the effect of their composition actions.

(ii) **UI-centric models for resource integration.** We propose a new visual-oriented model for mashup composition, based on associating data and function renderers, exposed by online resources, to visual renderers available in some pre-defined UI templates. Associating a UI to an integrated data set is one of the most challenging tasks in the construction of mashups. Our UI templates and the consequent visual integration model try to alleviate this task, also providing the users with an environment where they program data integration and synchronization of different services by expressing at the interface level examples of what they would like to experience during the execution of the final application. We therefore propose Domain Specific Languages (DSLs) that encapsulates the fundamental constructs of the visual mapping paradigm, yet abstracting from specific visualization styles and execution devices.

(iii) **Multi-device deployment of composite resources.** Our modeling abstractions enable a model-driven process that transforms examples of data integration and UI component synchronization, visually defined by users, into applications that can be run on multiple devices. A design environment allows users to visually express

data integration queries and automatically generates a composition description, based on our DSLs, that is then exploited by client-side execution engines running on different devices to dynamically integrate data and visually render the final mashup. Independently of our proof of concept implementation, we believe that the modeling abstractions under these mechanisms captures salient, generally valid aspects, that can be reused throughout different UI templates and execution platforms.

(iv) **Lightweight integration of resources.** Our visual integration model supports union and intersection joins to create integrated views of data coming from multiple services and a publish-subscribe, event-driven synchronization of data sources and services spanning multiple domains, both in form of remote APIs (e.g., the most common map-based services or other APIs commonly adopted in Web 2.0 applications), or of context- and content-based services local to mobile devices (e.g., the access to the GPS component or to the info stored in the personal address book). The main challenge is not the definition of a new data or service integration technique; rather our method introduces some abstractions that sustain a paradigm where the local and global integration schemas and publish-subscribe coupling of UI components are dynamically derived from the association operated by users through a visual interactive paradigm. One of our goal is indeed to promote a lightweight

integration of data and services that does not require any dedicated integration platform and is based on queries and integration mechanisms at the UI level that the users can intuitively define and easily execute on different client devices, even the mobile ones that are characterized by limited computing capabilities.

(v) **Collaboration.** We introduce asynchronous and synchronous collaboration techniques in mashup development and execution, enabling users to share resources with other people (e.g., friends, teammates, clients, co-workers) and to co-create artifacts, in line with the Web 2.0 style. In particular, we enable the production of annotations on the available resources and introduce techniques for synchronous co-editing that are based on the propagation of users’ actions to different active instances of a shared information space.

(vi) **Proof of concepts.** We illustrate a reference architecture for the execution of Web and mobile mashups that has guided the development of our mashup platform, PEUDOM (Platform for End User Development of Mashups). Through a performance test we show that a lightweight integration of resources is possible also under the limitations posed by the computing environments of client-side applications and mobile devices. Through user studies we also validate the adequateness of the composition paradigm with respect to the abilities of the average end users.



## A FAMILY OF READOUT ASICs FOR X AND $\gamma$ -RAYS SPECTROSCOPY.

**Riccardo Quaglia** - Supervisor: **Prof. Carlo Fiorini**

Aim of my Doctoral activity has been the study, the design, the characterization of different integrated circuits, designed for different applications in the field of nuclear electronics.

Nowadays, advanced microelectronics processes have made possible advanced signal processing with dedicated ASICs (Application Specific Integrated Circuit) for the readout of an increasing number of channels, ensuring at the same time compactness, low noise and low power in the systems.

My work in these years has been especially focused on the design of integrated circuits for Silicon Drift Detectors used both in X-ray and  $\gamma$ -ray applications. An interesting research field is to combine ASICs design for the acquisition of signals from SDDs in order to reach the best performances of energy resolution with compactness, low power and high efficiency. Silicon Drift Detectors (SDDs) are relatively recent devices, invented by E. Gatti and P. Rehak in 1983, that are now the de facto standard in low noise, high rate X-ray detection (for instance EDX, EDS, XRF, etc..) in the typical range 0.2 - 30 keV, but they are also a competitive alternative to photomultiplier tubes (PMTs) for the readout of light from scintillators, thanks to their high quantum efficiency and low electronic noise.

The simultaneous presence of this two important peculiarities makes SDDs in advantage in scintillator readout respect to other photodetectors, for instance PIN diodes (higher noise) or Silicon PhotoMultipliers (huge improvements in the last few years but still a lower photo detection efficiency) despite of the need to moderate cool them in real applications. The advantages of SDDs respect to PMTs are in robustness and compactness and they allow a simpler use in extreme conditions such as astronomical observatories on satellites or interplanetary missions; furthermore the compatibility of silicon detectors to high magnetic fields allow also the use of SDDs arrays in nuclear imaging combined to Magnetic Resonance Imaging (MRI). The activity was carried out in several projects: an 8-channels ASICs for high resolution and high throughput X-rays Spectroscopy with SDDs for the High Time Resolution Spectrometer (HTRS), a  $\gamma$ -rays spectrometer in the range of 150 keV-15 MeV based on large LaBr<sub>3</sub>:Ce scintillators with a 27-channels ASICs, a highly reconfigurable 8-channels multi detectors (SDD, PMT, Ge, SiLi, ecc..) readout chip (VERDI - VEratile Readout for Detector Integration) and the application of the designed ASICs for

nuclear physics experiments (INFN-GAMMA and INFN-SIDDHARTA). In the HTRS project in particular a high efficiency pile-up rejector circuit and logic has been discussed and designed in a 350 nm technology. The pile-up rejector takes advantages of the symmetrical shape of the output response of the filters (9th order Semi-Gaussian) but can be, in general, applied to any shaping filters because the working philosophy is based on programmable time windows that control different phases of the peak detection mechanism. The results obtained have been very interesting, absolutely at the state of art for X-ray spectroscopy in terms of both energy resolutions and electronic processing efficiency. The system designed combined to a CMOS preamplifier (developed in parallel to the activity, named CUBE) allows to reach a best energy resolution of 126 eV on the Mn-K $\alpha$  line with an input count rate higher than 100 kcps or 144 eV at 800 kcps in a single channel. The second project, in collaboration with the European Space Agency (ESA) and the Bruno Kessler Foundation (FBK), is in the framework of the use of SDDs as detectors of scintillation light. The project has started for the investigation of an alternative to PMTs for the

readout of large (1", 2" and 3") LaBr<sub>3</sub>:Ce scintillator for future interplanetary applications. The system has been designed with particular attention to all the components in order to obtain modularity and compactness. The detection head is not unique for all the scintillators but it is based on the assembly of single fundamental detector modules composed by an array of nine SDDs with active area of 24 mm x 24 mm (suitable alone for the 1" scintillator). Each SDD array is packaged in a single structure composed by the detector itself, the ceramic carrier and a copper block that has the same width of the sensor chip and ceramic and it is screwed to a common aluminum base cooled by Peltier stages. The desired active area can be realized placing side by side the appropriate number of modules. In case of 2" scintillator four modules are used while for the 3" configuration nine modules are required. The module is composed by squared 8 mm x 8 mm detectors that, in the single unit, have been deeply tested for noise and ballistic deficit estimation. The ASIC designed is based on 27 analog channels with a single multiplexed output and it is characterized by many functionalities (for instance multiple peaking times 2  $\mu$ s, 3  $\mu$ s, 4  $\mu$ s and 6  $\mu$ s, or multiple

gain settings) programmable with a custom designed SPI interface. The ASICs can be used alone for the readout of the 1" or in combination with another one or other two for the readout respectively of 2" and 3". Each ASIC is directly bonded on a dedicated PCB and the required number of these board are connected to another board that interfaces the ASICs with the acquisition system, that is composed by two parts, an analog to digital conversion boards (one per ASIC) and an FPGA boards (one per system) that manages all the ASICs and provides an interface with the PC software. The calibration of the system is carried out with a <sup>55</sup>Fe X-rays source, absorbed in direct conversion by each SDD. A linear transformation in gain and offset equalizes the channels mismatch. Test with the 1" scintillator detector have conducted obtaining an energy resolution of 3 % at 662 keV (and 2.1 % at 1.3 MeV) lower than the reference one obtained with a PMT (3.2 % at 622 keV) and confirming a further advantages of the readout based on SDDs arrays respect to a photomultiplier, a superior linearity at high energies. Detectors and electronics developed for the ESA-project

can be suitable for two physics experiments: SIDDHARTHA and GAMMA. Preliminary experimental results are very promising for the upgrade of the existing instruments with the new electronics.

# ENERGY-AWARE VISUAL ANALYSIS FOR WIRELESS MULTIMEDIA SENSOR NETWORKS

Alessandro Enrico Cesare Redondi - Supervisor: Matteo Cesana

The integration of low-power wireless networking technologies such as IEEE 802.15.4-enabled transceivers with inexpensive camera hardware has enabled the development of wireless multimedia sensor networks (WMSNs), also known as visual sensor networks (VSNs). VSNs can be thought of as networks of wireless devices capable of sensing multimedia content, such as still images and video, audio, depth maps, etc. Due to their flexibility and low-cost, VSNs have attracted the interest of researchers worldwide in the last few years, and are expected to play a major role in the evolution of the Internet-of-Things (IoT) paradigm. In this scenario, VSNs will be employed for carrying out a broad range of complex analysis task, such as object recognition, event detection, localization and tracking, etc. thus paving the way for several new applications ranging from enhanced surveillance to advanced services for health care.

The traditional approach to visual analysis is based on a two steps compress-then-analyze (CTA) paradigm: images are acquired in digital format by sampling and quantizing the light-field on a discrete lattice of pixels. Images, or image sequences, are then compressed

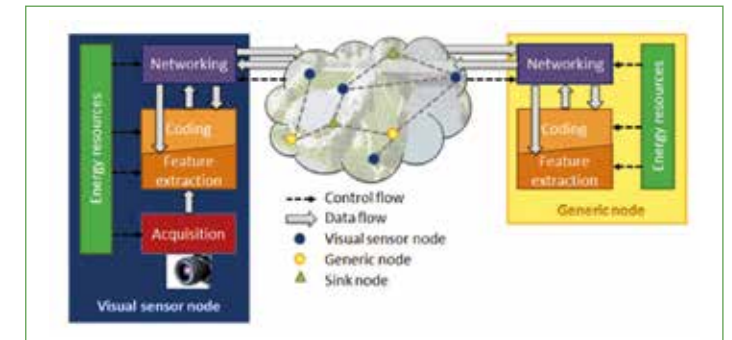
in order to be stored and/or transmitted for further analysis. A large body of research has focused on the analysis of visual data to accomplish high level tasks, e.g. recognizing letters, faces, objects, detecting events, etc. In some cases, as in the data processing pipeline of smart cameras, visual analysis and compression might be tightly coupled. Part of the analysis of raw data might be performed on-board and used to drive the configuration of the encoder, e.g. by varying the degree of lossy compression. However these systems are costly and often characterized by significant power consumption. In many scenarios, though, image analysis comes at last and it is decoupled from the acquisition and compression phases. Therefore, image analysis is often based on a compressed and hence lossy representation of the original image, which might significantly impair its efficiency. Moreover, when only the semantics of the visual content matter, transmitting image data retaining a pixel-level representation is inefficient in terms of the computational and network resources used, especially when the analysis is based on data sensed by more than one camera. In addition, energy-efficiency is often regarded as a secondary aspect, or is neglected altogether, since

most of the processing burden associated to common image analysis tasks is to be carried out at a centralized, power-eager, node.

This traditional approach is partially at odds with one of the most efficient visual analysis system we are aware of: the human visual system. There, visual stimuli are analyzed and ultimately converted into high level semantic concepts, which form the basis for efficient communication between individuals. Indeed, when interacting with others, people refer to such concepts, rather than resorting to the low level representation of the visual stimuli, since it entails a more efficient coding of visual information.

This thesis is focused on departing from traditional solutions and pursuing a paradigm shift that affects the way visual data is sensed, processed and transmitted, somehow mimicking the human visual system. The key tenet is that most visual analysis tasks can be carried out based on a succinct representation of the image, which entails both global and local features, while it disregards the underlying pixel-level representation. In order to satisfy the severe energy constraints dictated by visual sensor networks, it is imperative

to optimize the computation, coding and communication of the visual features (see Figure 1). On the coding side, this thesis tackles the problem by proposing a novel joint analyze-then-compress (ATC) paradigm. That is, image features are collected by sensing nodes, are processed, and are then delivered to the final destination in order to enable higher level visual analysis tasks. To this end, this thesis analyzes first the case of non-cooperative visual processing, that is considering one camera node as a single entity, directly connected to a remote server through a wireless link. Under this scenario, we propose novel solutions to encode different types of local visual descriptors, in order to minimize the amount of data transmitted and the corresponding energy consumption. The proposed encoding schemes are extensively evaluated in terms of their compression power and their impact on the distinctiveness of the descriptors themselves. We also propose a framework to achieve optimal internal allocation, i.e., to select the optimal number of visual features and the corresponding coding parameters to use in order to maximize the overall analysis accuracy. Such rate-accuracy optimization framework is used to compare



1. Overview of the system architecture

the proposed ATC paradigm to the traditional CTA paradigm, through a real VSN testbed.

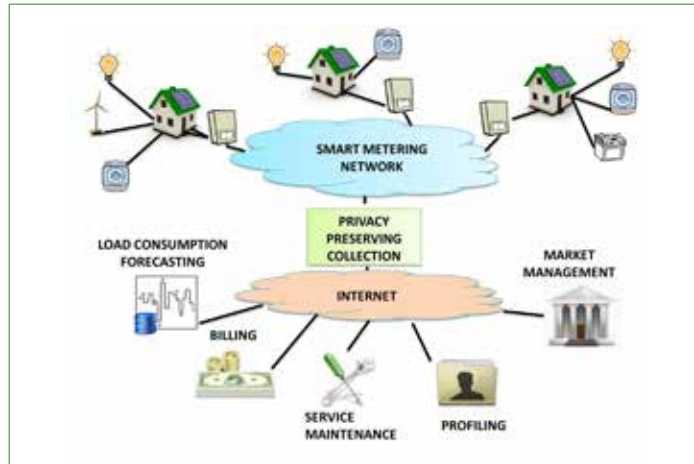
When camera nodes are considered as part of a network of devices instead of single entities, cooperative visual processing may be achieved. Different visual contents may be acquired from multiple camera nodes and transmitted to a central controller for further analysis with the aid of several generic nodes with relay capabilities connected to the network. Such generic nodes may also be employed to help camera nodes in processing the acquired content, e.g., extracting visual features. Also, the transmission of visual features may be subject to tight application-dependent requirements, and may be affected by network conditions. Under this scenario, this thesis addresses the design of several

networking tools for improving cooperative processing in VSNs. First, a mathematical formulation is proposed, formulated and solved to optimally allocate the available energy and bandwidth resources to single camera nodes in a sensor network, in order to maximize the overall accuracy of the analysis task. Since feature extraction is to be demanded to energy and processing-constrained sensors, cooperation among multiple sensors may be required to share the overall processing load. This thesis also addresses this problem by developing offloading schemes that can be implemented in the network to delegate part of the processing loads to neighboring nodes, on one hand to obtain significant speed-ups in terms of processing time, and on the other hand to relieve the burden of intensive visual processing on a single camera node.

# PRIVACY PRESERVING DATA COLLECTION IN THE AUTOMATIC METERING INFRASTRUCTURE OF SMART GRIDS

Cristina Rottondi - Supervisor: Giacomo Verticale

The energy industry is rapidly changing. Smart Grids are developed by massively integrating Information and Communication Technology (ICT) into energy grids to ensure security of supply. The new energy grid will be equipped with innovative sensing and control systems, capable of performing real time monitoring of power generation, transmission and usage, of analyzing consumption data and providing information about optimization and forecasting of power usage. Moreover, it will allow a consistent reduction of carbon emissions by integrating Distributed Energy Resources (DER) and increasing the efficiency of energy utilization. A pivotal role in Smart Grids is played by Smart Meters and communication Gateways, which are installed at the customer's premises. A Smart Meter performs measurements of the energy consumption, of the availability of energy storage capacity, or of local energy generation and sends these data via the Gateway to External Entities, e.g., to a metering operator or a meter service provider, which in turn provide these data to the energy supplier to enable accounting and billing. Also other entities such as Distribution System Operators (DSOs) or Regional Transmission Operators (RTOs)



1. The Smart Grid scenario with multiple External Entities collecting metering data

might be interested in such data to optimize the distribution network. The customer, i.e., the Gateway, does not only send data but could also receive data, e.g., pricing information when using variable tariffs to which it responds accordingly. Thus, the data of the smart metering system has a certain economic value, may enable several value added services and can be accessed by multiple entities. However, security and privacy are of paramount importance to ensure correct operation and protection of customers' personal data: it has been shown that customers' electrical usage readings can be used to profile their behavior and even to determine which

household appliances are being used. Therefore, through the analysis of the customers' electrical load profile, detailed information about personal habits and lifestyles can be inferred. The protection of customers' privacy can be realized by implementing a secure architecture enabling aggregation of the collected data. If data such as measurements or responses to pricing information is aggregated over a certain area (e.g., a network segment) the likelihood of revealing personal information (e.g., usage behavior, presence at home) is greatly reduced. This work proposes a privacy-friendly

infrastructure for allowing utilities and third parties (the so-called External Entities) to collect measurement data with different levels of spatial and temporal aggregation from Smart Meters without revealing the individual measurements to any single node of the architecture (see Figure 1).

The proposed architecture can be either centralized or distributed: the former introduces a set of functional nodes in the Smart Grid, namely the Privacy Preserving Nodes (PPNs), which are supposed to be controlled by independent parties and collect customers' data encrypted by means of Shamir Secret Sharing Scheme, performing different spatial and temporal aggregations for each External Entity according to its needs and access rights. By exploiting the homomorphic properties of the sharing scheme, the aggregation can be performed directly on the encrypted measurements. Conversely, the distributed security architecture relies on Gateways placed at the customers' premises, which collect the data generated by local Meters and provide communication and cryptographic capabilities. The Gateways communicate with one another and with the External Entities by means

of a public data network. For both architectures, the deployment of the information flows among the nodes of the network is also discussed, assuming that the routing of communication flows can be centralized or can be performed in a distributed fashion using a protocol similar to Chord. The problem is modeled by means of Integer Linear Programming formulations and heuristic algorithms are provided to tackle large instances. The performance and the security guarantees provided by the proposed architecture are evaluated assuming various adversary models, and the scalability of the infrastructure is first analyzed under the assumption that the communication network is reliable and timely, then in presence of communication errors and node failures. Another possible approach to privacy protection relies on data pseudonymization, which consists in replacing the identity of the subject generating the data with a pseudonym, which still allows to relate data generated by the same source, but makes it impossible to attribute them to a specific user. Therefore, for the centralized architecture, a data pseudonymization protocol is introduced to allow the collection of individual pseudonymized data, which

maintain their temporal sequentiality along a time span of finite duration, but cannot be related to the identities of the users that generated them or to the data generated by the same user in the preceding or following time windows. Moreover, both aggregation and pseudonymization techniques can be combined with data perturbation methods, which are aimed at decreasing the level of precision of the provided information. Therefore, we also discuss how the proposed privacy-friendly infrastructure can be integrated with data obfuscation by means of noise injection, as inspired by the framework of differential privacy. Finally, we discuss how the centralized infrastructure can be adapted to perform the distributed optimization of energy consumption without compromising the privacy of the users, with the aim of shaping the aggregated load profile of a neighborhood according to the local energy production by renewable sources. Such goal is achieved by scheduling the starting time of domestic deferrable appliances (e.g. dishwasher, washing machine, recharge of electric vehicles), without disclosing to the schedulers the energy consumption pattern of the single appliances nor the identity of their owner.

## MODELING, ANALYZING, AND MONITORING INTERACTING SOFTWARE COMPONENTS

**Mario Sangiorgio** - Supervisor: **Prof. Carlo Ghezzi**

Software systems are becoming more and more pervasive and their complexity is increasing as they need to solve novel problems. They tackle this complexity by building systems combining different software components, each one in charge of providing a particular feature. In this context, it is crucial that developers fully understand the behavior of all the software components they rely on: software engineers should be aware of all the possible abstract states of the components they use, the effects of their operations, and the protocols that clients have to follow to interact with them. Ideally, for each software component, developers should have access to specifications that describe how they work and how they interact with other objects. Only with this knowledge in mind developers can effectively build software that works properly and manages to meet its requirements. The majority of the software projects do not have any kind of formal specification available. Instead, the only way developers can know what a component is expected to do is to read either its natural language documentation or through its source code (when it is available). Natural language documentation is often useful but it usually describes the

behavior of the component too vaguely and in a way that is not always precise. In fact, there is no guarantee that the documentation covers all the special cases that it is up to date or even that it matches the actual behavior of the component. Conversely, source code lacks abstraction. It is the ultimate source when it comes to know what a software component does, but it contains too many details that could confuse developers instead of helping them to quickly figure out how a piece of software works. The software engineering research community tackled this issue with the definition of models to describe software behavior and with techniques designed to automatically infer them. Our contribution includes the definition of different kind of software behavior models, the design of suitable inference algorithms, and the study of two practical applications that can take advantage of the information encoded in the models. Our main research effort led to the introduction of different ways to model software behavior. The models we designed can effectively describe the behavior of a software component at different levels of abstraction. We designed models describing very precisely the behavior of

a software component (BEM), and models that only focus on the more abstract protocol of interaction that clients using it have to follow (PBM). BEM's very detailed description of the behavior of a software component is limited to a small scope. On the opposite, the more abstract PBM give up some details to describe the general behavior of a software component. For each model we defined a variant that describes the behavior of a single component in isolation and another that also takes into account interaction among the different objects that compose a software system. This extension is particularly important because it allows us to capture and formally describe all the interesting and significant behaviors that only arise when two or more software components interact. Interaction behaviors are often hidden, they are not obvious and they may depend on the specific instances of the software components working together. We formally defined what a composition is and how the operations on each of its components affect it. Developers seldom have the skills to properly write them and in the majority of the cases the only documentation comes in natural languages. On the other hand, formal specifications can help program understanding

and could be the basis for other analysis using them. We also designed inference algorithms to automatically build models starting from existing software artifacts. We address the lack of formal specifications for the commonly available software components by defining and implementing an inference algorithm that automatically builds behavior models for existing software artifacts. In this work we infer models through dynamic analysis with a black-box exhaustive exploration of a small, but significant, scope. The algorithms we designed work for both stand-alone and interacting software components. The problem setting we described so far concerns only the design phase of a software application lifecycle. Nevertheless, software behavior models usefulness is not limited to that phase. At design time, the knowledge encoded in the models help developers understanding what a component does and how it could be integrated in the system under development. The information they contain can be leveraged to build techniques that use behavior models during other phases of the software lifecycle. In this thesis we explored two ways to support developers when they implement the application and in the phases that follow its deployment. The first application targets the software construction phase with a static analysis tool. The purpose of the analysis is to check a piece of code against the behaviors encoded in our models. It then reports possible issues that emerge when there

is a mismatch with respect to the proper usage of a software component. The tool aims at finding both functional defects and implementation issues that lead to fragile code even if it is working properly. We developed a static analysis tool that checks the byte-code of existing programs against the specifications encoded in the model. The tool detects component misuses and reports the related errors to the developer. It also looks for fragile code, i.e. code that has a correct behavior only because it relies on some implicit assumptions, which happen to hold in the considered codebase. Those assumptions make it easy for the code to break: a change to a loosely related part of the application might turn an assumption to be false with the effect of introducing an error in the program. The second application of software behavior models extend their usage from their traditional scope to the run time phase. We can in fact integrate the information used in behavior models in a run time monitoring system. The monitoring tool can check that the system uses safely software components. This is particularly useful in the context of SOA where systems rely on external components providing part of the functionalities they need. These components give developers access to external services, but they usually do not have any control over them: services are usually designed, maintained and operated by a third party. The introduction of behavior models based monitoring system allows checking that the external components in charge of

providing functionalities behave as expected and, if needed, to report violations as soon as they happen. We faced these problems arising from Service Oriented Architectures by keeping models at run time. Doing that, we can leverage the knowledge encoded in the models to reason about the state of the external components. Keeping alive the models at runtime required the definition of techniques to monitor the execution of the system and to detect and report violations when the collected data contradicts the behaviors encoded in the model. Additionally we addressed the issue of keeping the models updated with respect to the information collected at run time. The models inferred at design time may be incomplete. Hence, we need a way to enrich them as soon as we have more information about the behavior of the external service. In conclusion, we designed models that can be useful during all the phases of software engineering: they help understanding the behavior of undocumented classes and they leverage on this information to build other tools that addresses program construction (helping finding errors), maintenance (through the elimination of fragile code) and even program execution with runtime monitoring.



## PERFORMANCE AND POWER ANALYSIS OF LARGE-SCALE VIRTUALIZED INFRASTRUCTURE

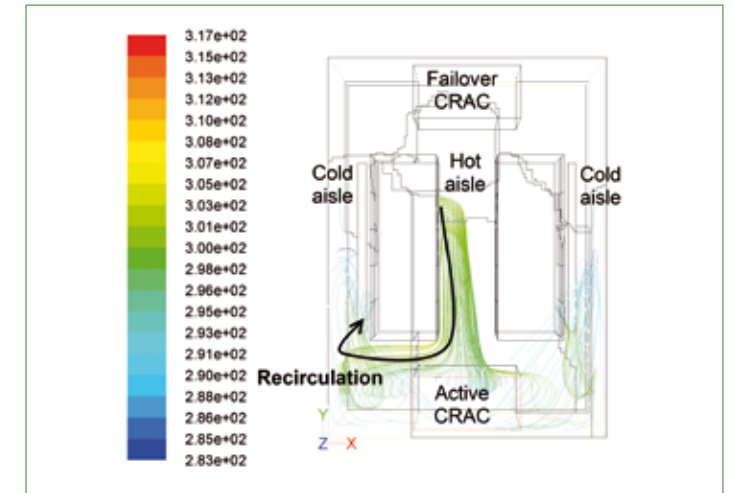
Andrea Sansottera - Supervisor: Paolo Cremonesi

Datacenters host hundreds or thousands of different applications. Traditionally, most applications run on dedicated servers, due to compatibility, security and performance reasons. This often leads to underutilization of processing and memory capacity, a problem often referred to as over-provisioning. With the rapid growth in datacenter power consumption, *over-provisioning* is no longer affordable. In fact, running servers at low utilization leads to inefficient energy usage, due to leakage currents and other inefficiencies. A strategy to fight the over-provisioning of datacenters is the adoption of *full virtualization*, which allows the multiplexing of multiple virtual machines, intended as efficient and isolated duplicates of the underlying hardware, on a single computer. The recent introduction of virtualization technology on commodity server hardware allows the process of *server consolidation*, i.e., the reassignment of a set of applications, typically running on separate servers, on a smaller set of target servers, using a virtual machine monitor to isolate them. The availability of virtualization also gave birth to the idea that storage and computation could be sold as a utility, similarly to electricity. Datacenters previously built

to run large Internet services, such as the Amazon online-store datacenters, began to provide hosting services to other companies, allowing customers to quickly startup and shutdown hundreds of virtual machines. This model of dynamic provisioning of computational and storage resources is known as *Infrastructure-as-a-Service* (IaaS). While virtualization addresses the compatibility and security concerns related to running multiple virtual machines on the same server, performance modeling is more fundamental than ever for sizing virtual machines and satisfying the response time requirements established in service level agreements. At the same time, virtualization allows one to flexibly schedule and reassign workloads within the datacenter. This provides the opportunity to optimize the utilization of resources, reducing the overall cost and power consumption. The two problems are tightly related. In fact, consolidation must be tightly controlled through appropriate performance modeling, because solving the over-provisioning problem should not lead to costly service level agreement violations. The models aimed at finding the optimal allocation of workloads within a datacenter,

subject to several constraints, can be classified in two broad categories: *static* and *dynamic* workload placement problems. In static workload placement problems, such as server consolidation, the set of workloads is known in advance. In dynamic workload placement problems, faced in IaaS clouds, virtual machines are provisioned and terminated on-demand. The first part of our research work consists in the formulation of a static workload placement problem that aims at minimizing the power consumption of the datacenter. Contrary to previous work we consider heterogeneous servers and performance constraints. The optimization of server and cooling power is tackled as a joint problem, since prioritizing the most efficient servers might lead to an increase in heat-recirculation (see Figure 1). In turn, this causes an increase in the power consumption of the air conditioner, which needs to lower the temperature of the cold air to guarantee that the inlet temperatures of the servers are within given thresholds. Our model establishes a relationship between workload allocation, server utilization, server power consumption and cooling power consumption. To ensure that the increase in resource utilization does not lead to unacceptable response times, performance

constraints are incorporated in the model. A case study consisting of other one thousand real-world applications and two hundred servers shows that taking into account power consumption while allocating workloads leads to more than 8% power savings, without increasing the average response time of each workload. To solve the problem, we present both a Mixed Integer Non-Linear Programming formulation as well as an effective Variable Neighborhood Search heuristic. In the context of dynamic workload placement problems, we consider a model with virtual machines belonging to several classes, each characterized by different resource requirements and revenue, and heterogeneous servers, with different resource capacities and costs. The aim is to compute the scheduling policy that maximizes the infinite horizon average profit. This problem, relevant to the providers of Infrastructure-as-a-Service clouds, is formulated as a continuous-time Markov Decision Process. We show that the optimal policies often take counterintuitive choices. This information is valuable for the design of scheduling heuristics. Our tuned and memory-optimized implementation of the solution algorithm can efficiently solve problems with several millions of states, using either multi-core CPUs or GPUs. The policy is computed offline using our algorithms and then can be easily applied by the virtual machine scheduler of the cloud provider. Performance models, such as the response time model used for cooling-aware static workload placement, require



1. A bird-eye view of a datacenter room. Cold air exists from the floor in the cold aisles and hot air is extracted from the top of the computer room air conditioner (CRAC). However, some of the hot air recirculates from the exhausts to the inlets of the servers.

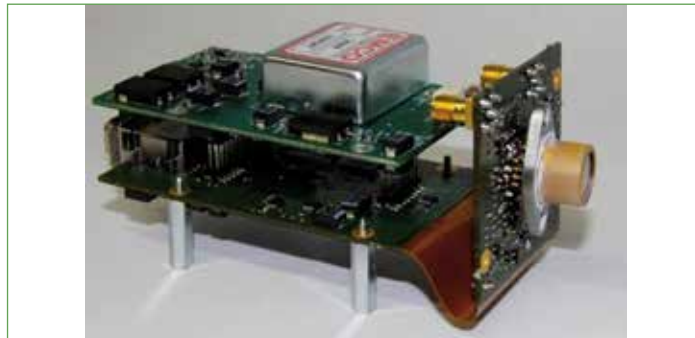
several parameters. In particular, for a queuing network model, the service demand as well as the characteristics of the arrival processes have to be measured or estimated. Simplistic models or bad estimates of the parameters can lead to dangerously underestimating the required computing resources or wasteful over-provisioning, neglecting the benefits of performance-driven workload placement. Hence, we presented new results in the context of performance parameter estimation. We designed a novel algorithm for estimating the service demand from throughput and utilization measurements when the monitored system is subject to changes. In virtualized datacenters, this is more common than ever, since virtual machine settings can be changed in a few seconds, severely affecting the service demand. Improving upon the state of the art, the proposed

algorithm can provide accurate estimates even in the presence of both outliers and short-lived, recurring changes in the behavior of the system. The accuracy of our algorithm is assessed challenging synthetic data sets, with up to 30% of contaminated observations, and on real-world datasets. To improve the modeling of the arrivals at a queuing system, we developed new techniques for the parameterization of Marked Markovian Arrival Processes. This class of models allows to represent workloads with correlation between arrivals as well as interdependencies between different workload classes. To the best of our knowledge, we are the first to provide closed, analytical formulas to match the model parameters to the empirical characteristics of arrival traces. Convex optimization problems for approximate matching are also presented.

# NEAR-INFRARED PHOTON COUNTING INSTRUMENTATION AND APPLICATIONS

**Carmelo Scarcella** - Supervisor: **Alberto Tosi**

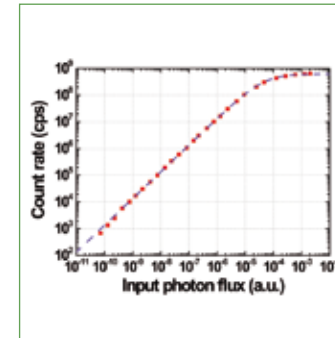
Nowadays many scientific applications need to detect very faint and fast light signals, down to the single-photon level. Advanced applications can be found in many fields, such as measurement of fluorescence decays, single molecule detection, characterization of new materials, non-invasive testing of VLSI circuits, single-photon source characterization, fiber optics testing, laser ranging in space and telemetry, quantum cryptography, quantum computing and studies in quantum physics. Different kinds of single-photon detectors have been developed in the past. Solid-state detectors have the advantage of high reliability and robustness, and they can be very compact and portable. Single photon avalanche diodes (SPADs) proved to have good performance in terms of photon detection efficiency and timing resolution. Silicon SPADs are sensitive up to 1100 nm, while InGaAs/InP SPADs can be employed to detect photons from 900 nm to 1700 nm. In this work, innovative electronics performing photon counting and photon timing measurements have been developed. The extensive characterization of the developed systems proved outstanding performance, thus becoming an enabling technology for several scientific applications.



1. Picture of the InGaAs/InP module detection-head electronics.

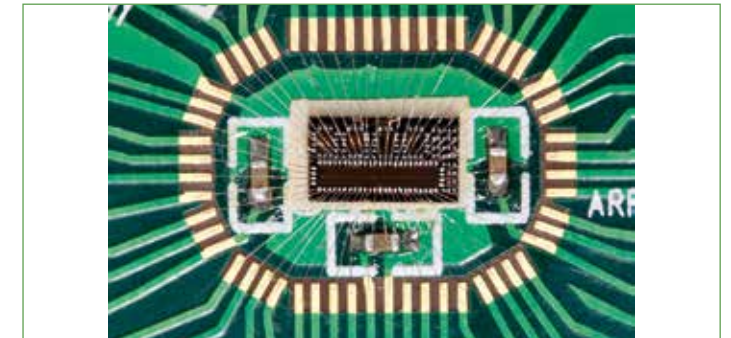
A high-performances InGaAs/InP single-photon detection module has been developed. Through all steps of the design and development the goal was to offer high flexibility of the overall instrument, in order to be adapted to the different requirements of the near-infrared applications mentioned above. All main system parameters, in fact, are configurable in a wide range. The SPAD temperature can be set down to 225 K (in order to reduce its dark count rate) by means of a TEC controller. The module can either work with an internally generated trigger or can be synchronized with an external trigger. The wide-band front-end electronics allows for the fast gating of the detector and for the low time jitter avalanche sensing. The module consists of a control-unit and a detection-head connected through a wide-band cable. Figure 1 shows a picture of

the detection-head electronics. An extensive experimental characterization has been carried out to test the whole system, proving outstanding performance, superior with respect to other state of the art single photon detection modules. In particular an effective afterpulsing reduction was measured thanks to prompt quenching of the avalanche by acting both at the anode and cathode side of the SPAD. The sensitivity of the detector within the enabling window is highly uniform. This is a fundamental requirement for photon-counting applications, in order to avoid distortion in the reconstructed optical waveform. Moreover, the developed instrument is able to work at high repetition rates (up to 133 MHz). The response of the system to a sharp laser pulse has been characterized, obtaining about 70 ps (FWHM) timing resolution.



2. Count rate dependence on impinging photon flux in the gigahertz gating system. The count rate is linear over five orders of magnitude, and saturates at about 650 MHz.

Two novel gating schemes for the sinusoidal enabling at gigahertz of InGaAs/InP SPADs have been conceived and developed. A free-running equivalent mode is obtained by keeping the SPAD gate signal unlocked from the synchronization reference of the optical waveform to be reconstructed. We called this working mode "gate-free". An extensive experimental characterization shown how the SPAD performance improves compared with a classical square-wave gating scheme, especially for the sensing of non-periodic high-throughput optical signals. When employing a 25  $\mu\text{m}$  diameter SPAD, the average photon detection efficiency at 1550 nm is 4 %, the afterpulsing probability with 1.5 ns count-off time is of about



3. Picture of the monolithic 32 pixels linear SPAD array.

1 %, the timing resolution is 90 ps (FWHM) and the maximum count rate is 650 Mcount/s, see Figure 2. This approach is therefore a new state-of-the-art for high throughput single-photon counting applications in the near-infrared range up to 1700 nm. The first scheme exploits an approach based on notch filters and self-differencing for the avalanche signal sensing. The second scheme developed, allows an easy tuning of the gating frequency, maintaining outstanding performance. Finally a 32 channels single-photon counting module for time-resolved measurements from the near ultraviolet to the near infrared range has been developed, Figure 3. It is based on a linear array of silicon CMOS SPADs, monolithically integrated together with the quenching circuits. The detection system provides a NIM output for each channel, suitable to feed external

time-correlated single-photon counting instrumentation, hardware correlators, or multichannel counters. The experimental characterization showed timing resolution uniform for all the channels ( $\sim 100$  ps of FWHM). The photon detection efficiency is good over a wide wavelength range, and achieves about 50 % at  $\lambda = 450$  nm. The extremely low dark count rate and afterpulsing probability represent the strong point of this CMOS SPAD array: 150 dark counts per second at room temperature and 1% of afterpulsing probability. This linear CMOS SPAD array can pave the way to new applications where multi-channel acquisitions of fast and faint optical waveforms have to be acquired, such as time-resolved spectrometers, through a very compact detection head with single-photon sensitivity.

## SYSTEM SUPPORT FOR ADAPTIVE PERFORMANCE AND THERMAL MANAGEMENT OF CHIP-MULTIPROCESSORS

Filippo Sironi - Supervisor: Prof. Marco D. Santambrogio

Computer architecture crossed a critical juncture at the beginning of the last decade. Single-thread performance stopped scaling due to technology limitations and complexity constraints. Therefore, chip manufacturers started relying on multi-threading and multicore processors to scale-up performance efficiently while keeping other figures of merit like energy and power consumption under control. In fact, whenever parallel software is available, a multicore processor harnessing Thread-Level Parallelism (TLP) can outperform a massive superscalar processor exploiting Instruction-Level Parallelism (ILP) within the same power budget. As a consequence, on-chip parallel architectures, which once were rare, are now commodity across all domains, from embedded and mobile computing systems to large-scale installations. Nevertheless, achieving efficient performance accounting for energy and power consumption progressively became increasingly complex requiring significant innovation across the hardware/software execution stack, even for commodity solutions.

At a high level, two challenges arise that hinder multicore processors efficiency. First, it

must be possible to effectively partition hardware resources among co-runner applications within multi-program workloads and avoid the negative effects of sharing when hardware resource cannot be partitioned. Hardware resource partitioning is necessary because most multi-threaded applications do not fully exploit the parallelism available in commodity multicore processors due to the major difficulties of fine-grain parallelism. Among the hardware resources worth partitioning, there are: compute bandwidth and cores, and possibly others depending on the workload. Ideally, the system software layer of the hardware/software execution stack should act on hardware resource partitioning to attain fair application performance and provide Quality of Service (QoS) guarantees while respecting system constraints. Second, the system software layer should operate in a transparent fashion without burdening application programmers with all the complexities of the hardware/software execution stack.

The focus of this dissertation is twofold. First, support efficient hardware resource partitioning for commodity multicore processors through a system software layer, which operate transparently for applications.

To this end, I present solutions to attain fair application performance and provide QoS guarantees for co-runner applications within a multi-program workload accounting for application-specific performance measurements and performance goals. Second, support efficient Dynamic Thermal Management (DTM) for commodity multicore processors through a low-level system software layer. For this purpose, I present a solution to constrain temperature when a multi-program workload of single-threaded applications runs on a Chip-Multiprocessor (CMP). The resulting artifact is a set of changes, runtimes, and libraries for the GNU/Linux operating system.

On the performance side, I present the Heart Rate Monitor (HRM), Metronome, and Metronome++. First, HRM is a split-design subsystem consisting of an extension of the Linux kernel and a user-space library to attach applications to the subsystem. HRM addresses the impedance-mismatch problem by providing application-specific performance measurements that are meaningful to both programmers and users and, at the same time, useful to the system software layer of the hardware/software execution stack. libhrm

provides programmers a simple Application Programming Interface (API) to instrument applications to define performance measurements and allow users to specify performance goals. HRM and libhrm make the operations of the system software layer I developed transparent to application programmers, which just exploit their knowledge of the application domain to define meaningful performance measurements. Second, Metronome is a kernel-space runtime introducing the notion of performance-aware fair scheduling by extending one of the scheduling classes of the Linux kernel. Metronome exploits HRM and the performance measurements it provides to drive application performance towards performance goals for co-runner applications within a multi-program workload. Metronome achieves its goal by implementing simple compute bandwidth partitioning mechanism and policy. Third, Metronome++ is a leap ahead with respect to Metronome; it adopts a split-design across the kernel- and user-space. A user-space runtime drives the kernel-space extension of the scheduling infrastructure of the Linux kernel to provide QoS guarantees for co-runner applications within a multi-

program workload by harnessing application characteristics like speedup and execution phases. Metronome++ achieves its goal by implementing compute core partitioning mechanism and policy. This dissertation additionally presents a set of minor achievements harnessing different decision-making techniques other than the heuristics Metronome and Metronome++ make use of.

On the temperature side, I present ThermOS, an extension of the Linux kernel providing DTM through formal feedback control and idle cycle injection. ThermOS addresses a shortcoming of commodity CMPs, which do not allow different cores to run at different clock frequencies when they operate in the same state. ThermOS avoids the negative effects depending on the lack of fine-grain control over hardware facilities like Dynamic Voltage and Frequency Scaling (DVFS) and improves upon state of the art.

On the performance/temperature side, I present preliminary results regarding joint adaptive performance and thermal management combining some of the aforementioned approaches in the Dynamic Performance and Thermal Management (DPTM) framework.



## ELIoT: A PROGRAMMING FRAMEWORK FOR THE INTERNET OF THINGS

**Alessandro Sivieri** - Supervisor: **Prof. Gianpaolo Cugola**

The world of embedded, communication-oriented devices has dramatically changed in the last few years, bringing new challenges and opportunities to researchers and developers of embedded, distributed and integrated systems: from messaging protocols to tools and facilities for the developers, to completely new scenarios enabled by this device revolution.

In the last ten years, several key technological and methodological changes started a revolution in the consumer embedded world: thanks to the fact that microprocessors have become cheaper and more powerful, the market has been flooded by a whole new world of mobile devices, able to run complex applications, collect data from a variety of integrated sensors, and combine these data to produce actions and to try to understand the environment in which they are located and the behavior of their owners; the so-called “Internet of Things” phenomenon started building up, as a consequence of availability of sensing and actuating devices powerful enough to be programmed as standard computers and to be integrated into the Internet, and to interact with the users through intuitive and dynamic interfaces; finally, research in the WSN and embedded fields have

produced reliable algorithms to be used in these applications, enabling low powered, wireless and distributed scenarios. As software engineers, we are interested in how these devices can be programmed, which tools are given to the developers to write applications that can comprise a wide variety of hardware, from the tiniest boards used in Wireless Sensor Networks, with 8/10 KBytes of memory and usually powered by small batteries, to smartphones and tablets, with powerful processors and GBytes of memory, powered by bigger batteries or (in some cases) power supplies.

The programming frameworks that can be used in these applications have some common characteristics: they target a usually small part of the device spectrum; in the case of WSN the hardware abstractions typical of the most famous operating systems, TinyOS and Contiki, have been conceived to support the boards available today, where a profound knowledge of the hardware is needed by the developer in order to port an application to a different platform; the situation is similar for hobbyist platforms such as Arduino. Difficulties in testing and debugging the applications are also common for these scenarios: this is often due to the unpredictability of at least

part of the interactions within the environment, that will be discovered only when the application is actually deployed; smartphone applications also rely quite heavily on Internet connections, and their behavior can become erratic if the connection is faulty, condition that can happen in several situations.

At the same time, there are profound differences: on the one hand, WSN research focuses on algorithms to collect data efficiently and to save as much energy as possible, giving the developer the programming tools to strongly tailor the application to her specific requirements (e.g., need for more communication reliability). These frameworks also consider interactions between devices to be local: data is exchanged inside the same network and decisions are taken locally. Interactions with the Internet are not common, since these devices have such low capabilities that it becomes quite difficult to add data exchange with the external world, and often the only external link is made by the sink, which is usually a computer or a more powerful device.

On the other hand, frameworks for more powerful devices focus on the ease of programming, to keep the barrier low to new developers, and the devices are usually in charge of collecting

data and sending it to remote servers, often located in today's cloud systems, which will elaborate these data and show it to the user, usually through a Web site (e.g., the music player example cited before). Local interactions are usually neglected and devices rarely communicate with each other, even if they are in communication range and integrate more powerful communication capabilities, different communication media, and more energy and power capabilities than, for example, WSN.

We think that this can be changed, and we introduced ELIoT (ErLang for the Internet of Things) to show that it is possible to offer a powerful framework that allows both types of communication. In doing so, we leveraged some of the existing features of the Erlang language and Virtual Machine, and introduced features for both types of communication: different network distribution protocols, which give more control to the developer on how the VM behaves when sending messages to remote processes. In particular, ELIoT offers specific mechanisms to make the transmission reliable or non-reliable, and the developer can decide to improve upon these options depending on the type of the application. Moreover, the framework provides broadcast communication with the possibility of disseminating computation, using context filters to select only parts of the network. We also simplified the network distribution protocol used by the VM; the capability of deploying computations

at runtime and automatically generate REST interfaces to invoke them, so that they can be easily integrated into different frameworks and used outside ELIoT (even browsed through the Internet, if necessary). This capability leveraged our knowledge of Web services in Erlang, due to the development of a previous framework to change the Web of resources to a Web of computations (the Computational REST paradigm and architecture).

At the same time, testing and debugging IoT applications is known to be quite difficult, for example to correctly integrate a model of the environment or to test algorithms scalability. This is due to the peculiarities of the applications and the lack of sound software engineering tools and procedures: for low-powered devices, the developer relies on her knowledge of the low-level aspects of the boards and the operative systems, creating a solution specific to the targeted scenario. On the other hand, applications for more powerful devices (e.g., smartphones) are developed with frameworks that offer simulators unable to reliably emulate the hardware platform, especially for those operating systems (e.g., Android) offering a wide variety of different devices (smartphones, tablets, TV accessories).

ELIoT provides a simulator tool that tests the code as is, emulating the network and the environment using known tools and allowing the developer to fine-tune the environment model (to the extent of plugging in new model implementations), leveraging her knowledge of the deployment situation.

Moreover, the simulator allows mixed scenarios, where part of the network is simulated and part runs on real devices: this avoids having to emulate the hardware of several different devices, and obtain more precise measurements from the real hardware.

ELIoT also provides a modified version of an existing model checker for Erlang (called McErlang, to reflect the peculiarities of IoT scenarios and the mechanisms that ELIoT provides with respect to “vanilla” Erlang. The developer can verify LTL properties of the application, obtaining validation proofs or counter-examples violating those properties. This tool is written in Erlang, and this means that it is able to verify ELIoT applications without changing the code or having to translate it into a different modeling language, thus producing more reliable results wrt solutions that need a model translation.

Finally, we tested the ELIoT framework on several different devices, going from small, WSN-like boards, to smartphones and more powerful devices, implementing typical IoT scenarios applications. We also tested it on special devices such as ebook readers, to show how low the energy consumption can be.



# ENERGY MANAGEMENT OF FULL HYBRID ELECTRIC BIKE: HUMAN AND ELECTRIC ENERGY INTEGRATION

**Pierfrancesco Spagnol** - Supervisor: **Prof. Sergio M. Savaresi**

The increased sensibility for environmental problems and the growth of oil price brought scientists and companies to invest significant resources on means of transport with more than one source of power. Particularly automotive field attracted an increasing interest in vehicle electrification.

Moreover in urban area environmental and traffic issues lead to enforce strict normative and to create new paradigm for urban mobility (e.g. car and bike sharing), in order to decrease pollutant emissions. In this urban context, Light Electric Vehicles (LEVs) are receiving great attention, particularly Electrically Power Assisted Cycles (EPACs). Several studies have been made in order to explain the growth of EPACs in the market, to analyze the state of the art and to try to predict their technical possible evolution based on users evaluation. These studies pointed out that main issues of EPACs are: autonomy, i.e. the number of kilometers with electric assistance; weight (related to the motor and battery pack); and, finally, duration and problems related to the recharge of the battery pack.

On the other hand, the main positive characteristic of EPACs for users is the reduction of physical effort.

The research proposes an alternative to the conventional

EPACs. The main idea is borrowed from the automotive field: the development of a full hybrid electric bike (HEB), configured as a parallel hybrid vehicle. The main objectives of the HEB are sustaining the battery charge, being independent from the grid and reducing cyclist effort.

To the best of the author's knowledge, this is the first attempt to develop this kind of bicycle.

This ambitious goal can be achieved with a correct energy management between the two power sources of the hybrid bike: the primary power source (the cyclist) and the secondary power source (the battery pack that acts as an energy storage). The two power sources are coupled by means of an electromechanical converter, i.e. the electric motor/generator as shown in the general scheme of a parallel electric bike.

The research focuses particularly on three aspects that influence the design of the energy management strategy to achieve the over-mentioned goal. First, we focused on estimation problems: the correct management of the energy fluxes to sustain battery charge needs a reliable knowledge of the State-of-Charge (SoC) of the battery. Moreover, knowledge of the force and power that are acting on the bike is crucial

to correctly set the electric assistance and recovery level to both decrease cyclist effort and sustain battery charge. In this research, we focused on SoC estimation, cyclist torque estimation and road grade angle estimation with a reduced sensor layout. Second, we tackled modeling problems.

The analysis and design of the hybrid electric strategies need a reliable model of the three main subsystems: the electric subsystem (electric motor and battery pack), the mechanical subsystem (the bike itself in relation with the environment) and the cyclist model. As for the latter, available textbooks of exercise sport physiology are analyzed, and a brief summary of human metabolic pathway is presented. A brief review of the methods to calculate human energy expenditure is discussed, focusing particularly on oxygen expenditure. All the available data in literature of oxygen expenditure and human efficiency considers steady-state points, while we are interested also in transients considering the urban usage of bikes (with frequent starts and stops). Therefore, an experimental campaign with the collaboration of Mapei Sport Service center researchers has been carried out. The aim of the experiments is to understand how much energy is consumed during transients

and during constant speed rides. Therefore, the attention has been focused on the ratio between the total developed mechanical energy and the total oxygen consumption. This ratio has been called equivalent metabolic efficiency and has been used to compare different cycling activities. During impulsive muscle effort (starts, acceleration and uphill riding) this ratio halves with respect to the case of constant speed profiles on flat road. Finally, a model of human fatigue (State-of-Fatigue) available in literature is detailed and analyzed: it confirms the observations underlined by the previous analysis. From experimental observations and the analysis of the fatigue model, and considering a city-bike common usage (frequent start-stops and cruise speed lower than 25km/h) the guidelines for the control algorithm design are the followings: first, electric assistance has to be predominant during impulsive muscle effort: starts and slopes. On the opposite, electric assistance progressively decreases until a cruise speed is reached. Second, electric recovery has to be as greater as possible during braking and downhill. Third, electric recovery can be activated during constant cruise speed when metabolic efficiency is greater. Finally, two

different strategies to sustain the SoC have been studied. One indirectly influence the electric assistance, while the other proposes to directly influence the assistance and recharging intensity with respect to the SoC of the battery. Simulation results show that it is possible to reduce the peak of State-of-Fatigue up to 30%, while penalized average value can decrease of more than 15% maintaining the SoC around the desired level. Finally, a comparison between the behaviors and performance of two causal strategies with respect to system mass variation or initial SoC different from the desire value has been presented: on one hand, the algorithm is able to sustain the charge also with masses 30% different from the nominal. Moreover, the algorithm is able to recover the SoC also from a complete discharged initial condition (10% SoC), without degrading State-of-Fatigue performance. The final part of the research is devoted to experimentally validate the proposed algorithms. First we show the benefits of HEB in a controlled environment: a cycling track. This experiment evaluates the differences in terms of oxygen consumption between the bike with and without the electric assistance, with the perfect sustaining of the SoC of the battery. The experiments

are performed by the same cyclist following in both cases a predefined speed profile at zero slope and show an increase of more than 28% of the equivalent metabolic efficiency when cyclist perform the experiment with the electric assistance.

The second set of experiments takes place in Paris, in three different urban cycling profiles with different characteristics using the adaptive strategy without changing any of the parameter between the different experiments. For two of the three profiles, some cyclists used heart-rate monitors repeating the same trip at almost the same speed with the HEB prototype and a traditional bike. Results show the benefits in terms of reduced maximum heart-rate peak (difference more than 30 beats per minute) and variance. The prototypes of the hybrid electric bike have been presented to the press and the herein control strategy has been patented. The first part of the research has been developed in collaboration with two industrial partners, SEMS S.r.l and Trenord S.p.A, while the second part is the result of an industrial project commissioned by JC Decaux, the operator of the well-known large-scale public bicycle sharing system in Paris, i.e. Velib'.

## TIME-OF-FLIGHT CAMERA FOR LONG-DISTANCE 3D RANGING

Federica Alberta Villa - Supervisor: Prof. Carlo Fiorini

In my Ph.D. major research, I conceived, designed, developed, tested, characterized and validated a complete detection system able to acquire at the same time 2D and 3D images and movies with single photon sensitivity. The sensor chip relies on Single-Photon Avalanche Diode (SPAD) detectors and on Time-of-Flight (TOF) measurements through microelectronic Time-to-Digital Converters (TDC), exploited at best to obtain 3D information from each single photon detected by each pixel of the imager and reflected from the objects of the scene under observation.

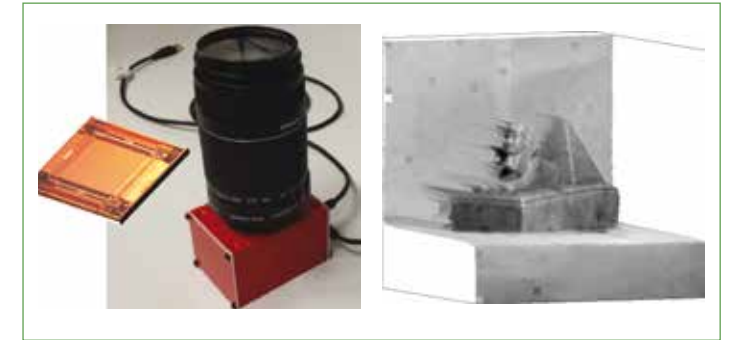
The activity was developed within the European project "MiSPIA" (Microelectronic Single-Photon 2D and 3D Imaging Arrays for Safety and Security applications), aimed at developing 2D imaging cameras and 3D ranging cameras for safety and security applications. Beyond these applications the developed system will be employed in Near Infra-Red Spectroscopy (NIRS), Fluorescence Lifetime Imaging (FLIM), Diffuse Optical Tomography (DOT) and other measurements based on the Time-Correlated Single-Photon Counting (TCSPC) technique. The core of the entire project has been the design of the microelectronic sensor chip,

which is an array of smart-pixels, with a SPAD and a TDC each. Arrays with different sizes (32x32, 16x32, 60x1), SPAD diameters (30  $\mu\text{m}$  – 135  $\mu\text{m}$ ) and fill-factors have been fabricated in a 0.35  $\mu\text{m}$  CMOS technology, by Fraunhofer IMS Institute of Microelectronics, in Duisburg (Germany). Planar arrays have been designed specifically for 3D ranging applications, while linear arrays are more suitable for bioimaging, microscopy, and spectroscopy applications, since they can reach higher frame-rate and provide higher fill-factor. More specifically linear arrays will be investigated in portable time-resolved Raman Spectroscopy systems, becoming the enabling technology to distinguish Raman stimulated emission from fluorescence background, by exploiting the arrival time of emitted photons. Both photon-counting and photon-timing electronics has been implemented within the sensor. In photon-counting mode, for 2D imaging, a 6-bit counter provide a result proportional to the intensity of the light emitted/reflected by the objects in the scene or the biological specimen under observation. In photon-timing mode, for 3D distance-resolved ranging, the integrated TDC is split in coarse counter (for providing a long dynamic range) and a fine interpolator (for sub-

nanosecond time stamping) in order to measure the accurate time-of-flight or fluorescence decay of the sample. The time bin is adjustable from 312 ps (with 10 bit resolution) to 10 ns (with 6 bit resolution), i.e. from 5 cm to 3 m depth resolution, using a 100 MHz reference clock. The power consumption of the chip at the maximum resolution and speed is lower than 1 W, whereas the total power consumption of the entire camera is 2.5 W. Compared to other arrays of SPADs and TDCs presented in literature, this sensor shows better linearity and state-of-the-art SPAD performance, though timing resolution was kept lower to match. The depth precision in 3D ranging applications is improved by accumulating a large number of conversion (> 500) and then computing the centroid of the acquired histogram of photons arrival times, thus reaching millimeter precisions. The complete system comprises, the microelectronic sensor chip, an external FPGA electronics to interface the imager with the computer, and a graphical user interface to set camera parameters and display acquired images and movies. Many measurements have been performed in order to characterize the SPAD quality (uniformity, PDE, DCR,

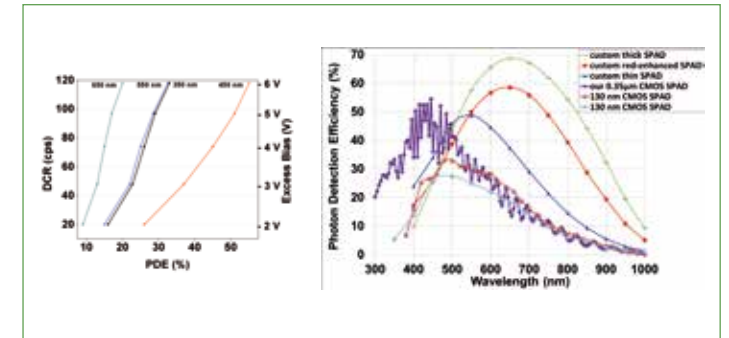
afterpulsing, timing, crosstalk) and TDC performance (single shot precision, DNL and INL) and to test the overall system. 3D movies with frame rates up to 200 fps and centimeter precision have been acquired, in collaboration with Heriot-Watt University and with the Physics Department of Politecnico di Milano.

Fig. 1 shows the developed camera (left) and an example of target acquired with a scanning set-up (right).



1. Developed camera for 2D imaging and 3D ranging, with the 32x32 pixels array chip containing 1024 SPAD detectors and 1024 TDCs (left). Example acquired image, with 2D and 3D data overlaid.

Besides the 30  $\mu\text{m}$  SPADs laid out in the imager arrays, also large area SPADs (diameter up to 500  $\mu\text{m}$ ) have been developed, for many demanding applications such as Near Infrared Spectroscopy (NIRS), in which extremely faint and defocused light reflected and/or scattered by biological tissues must be detected over a wide collection area. Furthermore other applications that require to properly align the optical path with the detector take advantage of employing such large area SPADs. Chips have been designed as either stand-alone SPADs with integrated quenching circuit or as complete single-photon detection channel smart pixels able to be operated in both photon-counting and photon-timing measurements. The produced CMOS SPADs provide very high performance, especially in terms of Dark Counting Rate (DCR), lower than 50 cps at room temperature for 30  $\mu\text{m}$  diameter SPADs and lower than 100 kcps for 500  $\mu\text{m}$  devices. Very good results have been achieved also in terms of Photon Detection Efficiency (PDE) from the Near-UV (20% at 300 nm wavelength) up to



2. Dark-Counting Rate vs. Photon Detection Efficiency of the developed SPADs (30  $\mu\text{m}$  diameter), at four different wavelengths and excess bias (left). Spectral Photon Detection Efficiency of the developed CMOS SPADs, compared to the state-of-the-art custom-process, reach-through thick SPAD by Excelitas, red-enhanced SPAD by Polimi and MPD, planar silicon thin SPAD from MPD, and CMOS SPADs in 130 nm technology.

Near-IR (5% at 850 nm) with the maximum at 420 nm (55%). Fig. 2 left shows the DCR of 30  $\mu\text{m}$  diameter SPADs vs. PDE, at different wavelength and excess biases, Fig. 2 right compares the PDE of the developed SPADs with other best-in-class custom and CMOS SPADs. Overall Fig. 2 shows that the developed SPADs reach the new state-of-the-art performance among CMOS single-photon avalanche diodes. These SPADs behave also as the ideal candidates for being exploited as the fundamental

cell in Silicon Photomultiplier (SiPM) detectors. In fact some testing structures of SiPM (8x8 cells, 60% fill-factor) have been conceived and manufactured and show very good performance in terms of photon-number resolution. In the near future an in-depth characterization of the produced SiPM structures will be performed and the design of larger SiPM will be carried on.

# MEASURING AND IMPROVING ENERGY EFFICIENCY OF A DATA CENTER IN A SELF-ADAPTIVE CONTEXT

Monica Vitali - Supervisor: Prof. Barbara Pernici

This thesis proposes a methodology and novel techniques to assess and improve Energy Efficiency (EE) in a data center environment, while respecting the constraints established in the Service Level Agreement (SLA) with clients. This work has been motivated by the recent attention towards green energy and renewable resources, and by the growing demand of energy coming from the Information Technology (IT) sector. This demand is so high to be comparable with the one of the major countries and it is expected to keep growing in the next years.

In this work we try to understand the current trends towards a greener asset for data centers and how it is possible to understand the current level of efficiency of a data center while suggesting some repair actions for improving it. We focus on the application perspective for improving efficiency, looking in details at applications and their IT infrastructure. We are considering the context of SOAs, where the applications hosted in the servers are services that can be represented using a process. Each service is decomposed in a set of activities connected in a work-flow. We consider that the data center uses virtualization for hosting its applications and we consider that each single activity

runs on a dedicated Virtual Machine (VM). An example of the typology of systems we aim to model can be seen in Fig. 1. A monitoring system is available for monitoring the system behavior with sensors that can work at several levels: (i) data center level: sensors collect information about the general data center; (ii) server level: sensors collect information about a specific server; (iii) VM level: the monitoring system collects data that can be obtained by the hypervisor or directly by the Operating System (OS) of the VM as if they were physical machines; (iv) application level: at this level, sensors collect information related to the application, such as response time, throughput, and availability.

We propose a goal-oriented approach where the state of the system is assessed using a set of indicators. This set is composed by some classical indicators, coming from the state of the art, and some new indicators, specifically defined to better describe the efficiency of a data center. These indicators are subjected to thresholds that are used as goals of our system. We propose a self-adaptive context-aware framework, where we learn both the relations existing between the indicators and the effect of the available actions

over the indicators state. The system is also able to respond to changes in the environment, keeping these relations updated to the current situation.

The main contributions of this work are listed below:  
**Hierarchical classification of Green IT approaches** this thesis provides a deep study of the fields of Green IT and Green IS and of the state of the art of these fields. Instead of limiting this analysis to a simple quote of available results, we propose a hierarchical classification of the different perspectives and we also propose a methodology for combining them in order to have a comprehensive approach towards EE.

**Metrics selection and proposal** Several metrics are analyzed in order to find a set which is representative of the QoS and EE of an IS. We also evaluated metrics that have an indirect impact on these two aspects, affecting the system state. Also new metrics have been introduced and tested, focusing attention on the application usage perspective.

### A goal-oriented model for efficiency management

A model is proposed and tested based on a goal-driven approach. The aim of the model is to provide an adaptive environment able to automatically adapt to suboptimal situations by

enacting adaptation strategies that impact directly or indirectly over efficiency and quality.

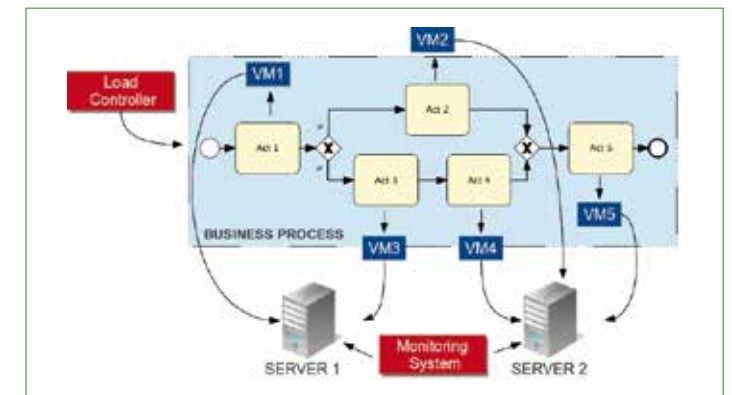
### Representation and learning of indicators relations

Using a Bayesian Network representation, we express the relations occurring among the various metrics used for monitoring the system state. These relations are very important for predicting the future behavior of the system and for conducting what-if analysis reasoning. The network has been learned from monitored data, without using any kind of knowledge coming from experts. Using machine learning techniques, we were able to learn the structure of a BN starting from continuous variables.

### Adaptive strategy selection

Effects of actions over EE and QoS are automatically learned providing an adaptive and dynamic environment for managing these aspects of an IS. In this way, the framework is able to suggest the best adaptation action without any intervention from experts and to automatically adapt to modifications in the environment that can modify the impact of the actions enactment.

**Modeling service execution in data centers** A mathematical model is provided for simulating the behavior of a service oriented data center. This



1. An example of the considered class of systems

model includes features typical of the considered domain (the service oriented data center), unavailable in other models. The model enables to run simulations of the data center behavior under several loads and with different configurations. This feature is particularly important during the learning phases. In fact, in this way it is possible to test new actions and new configurations in a protected environment and to observe possible outcomes on the system, avoiding to undermine the performance of the real system.

Results have shown that the proposed methodology is able to create a network of relations between indicators and to propose an effective

repair action to contrast suboptimal states of the data center. We also propose a model for simulating the data center behavior which can be used to conduct what-if analysis, allowing testing new configurations in a safe environment and predicting future states of the system. The proposed framework can be an important tool for assisting the system administrator in the management of a data center oriented towards EE, showing him the connections occurring between the sometimes contrasting goals of the system and suggesting the most likely successful repair action(s) to improve the system state, both in terms of EE and Quality of Service (QoS).

## FEMTOSCOPE ARRAY FOR CORRELATION AND SPECTROSCOPY (FARCOS): SIMULATION AND CHARACTERIZATION OF THE DETECTION LAYERS

**Pietro Zambon** - Supervisor: **Prof. Chiara Guazzoni**

FARCOS (Femtoscope ARay for COrrelation and Spectroscopy) is a novel modular and versatile detection system constituted by telescopes featuring high angular and energy resolution and able to reconstruct the Light Charged Particles (LCP) and Intermediate Mass Fragments (IMF) momentum at high precision for different physical cases in heavy-ion collision nuclear physics experiments at intermediate energies (10-200 MeV per nucleon), with stable and radioactive beams, which are the unique tool available on earth to study the properties of nuclear matter under extreme conditions. Each telescope is made of a couple of Double Sided Silicon Strip Detectors (DSSSD), the first 300  $\mu\text{m}$ -thick and the second 1500  $\mu\text{m}$ -thick, both with active area 6.4 cm  $\times$  6.4 cm, 32  $\times$  32 strips 2 mm pitch orthogonally arranged on the two detector sides, and as third detection stage thick CsI(Tl) scintillation crystals read-out by photodiodes with the purpose to stop even the most energetic particles and to allow particle identification. One of the ambitious FARCOS goal is to achieve a full particle identification and correlation by exploiting, together with the standard identification techniques ( $\Delta E$ -E on the several detection stages, Time of Flight (ToF), Pulse Shape

Analysis techniques both on the CsI(Tl) scintillation signals and, more challenging, on the DSSSD in order to lower the identification energy threshold and thus exploring wider energy ranges. The possibility to use our custom developed digital Data Acquisition system would be beneficial to this purpose. Indeed the waveforms shape retains – in line of principle – all the information about the interaction and the evolution dynamics of the impinging particle into the detector and a full signal digitization allows investigating easily all the desired parameters on a flexible software basis. In order to optimize the DSSSD performances, their coupling with the hybrid custom developed front-end electronics and to extract the values of entities needed for a careful device simulation (depletion voltage, actual doping concentration in the bulk region), we carefully investigated the fundamental physical parameters strip by strip such as leakage currents, total strip capacitances separating the contribution of the bulk capacitance (one strip vs. the whole other detector side) and of the inter-strip capacitance (one strip vs. the neighboring one) as a function of the applied reverse bias and of the test signal frequency. In addition, we

performed a detailed amplitude and position detectors response mapping exploiting the pulsed ion beam available at the Laboratori di tecniche nucleari per i Beni Culturali of the INFN (Firenze), DEFEL beamline. Such beamline can indeed provide bunches of monochromatic protons/light ions down to an average number of particles per bunch well below one single particle, in the energy range 1-6 MeV with a high degree of accuracy. Protons range within this energy interval matches very well with the common detector thickness, making them extremely useful probes to investigate the effect of the penetration depth on the signals shape. Nevertheless, the ion multiplicity per bunch intrinsically allows an accurate energy calibration and the estimation of the linearity of the overall read-out system. Special attention is also given to the impact of inter-strip beam incidence on the shape of the induced signals, among the main causes that can greatly spoil the overall particle identification capabilities. The systematical investigation we performed, together with dedicated numerical simulations, allowed clarifying a set of inter-strip phenomena already reported in literature but whose origin has been up to now misunderstood, namely the occurrence of bipolar

signals or signals with polarity opposite to what expected in neighboring strips.

The experimental work has been complemented with theoretical studies. We indeed developed a novel simulation tool for the 3D electron-hole transport computation and signal formation in 2D (easily expandable to 3D) semiconductor detectors, with linear or cylindrical coordinates, as a powerful and reliable tool to investigate the possibly complex signal formation dynamics especially when in presence of the high charge injection levels and of the high detector segmentations expected in the upcoming nuclear physics experiments (as the ones involving FARCOS) and in applications involving novel high brilliance FEL sources. In this respect we performed a preliminary careful code qualification with dedicated test simulations of different physical conditions, checking the results validity also with semi-empirical models found in literature. Exploiting custom developed advanced numerical techniques (charge clustering, adaptive time step size control, ...) and making extensive use of parallel calculus it has been possible to obtain reliable and accurate simulations of the charge cloud space-time evolution down to the ps time scale and up to the high value of  $10^8$  electron-hole pairs within reasonable times. We then simulated several realistic cases obtaining always a perfect match between simulations and experimental data. It is worth to mention how a set of simulations specifically targeted to study the aforementioned inter-strip effect

allowed us to put light on the precise formation dynamics of the bipolar or opposite polarity signals arising in adjacent strips, namely the peculiar shape of the strip weighting potential below the inter-strip area; the excellent agreement between simulations and experimental data gives nevertheless another proof of reliability.

To further improve our knowledge even of the CsI(Tl) crystals scintillation properties – third and last FARCOS detection stage –, we developed a specific data analysis method able to extract the time constants and intensities defining the scintillation light response by means of a direct fitting of the fully digitized waveforms. The dependence of the scintillation light parameters on the energy, charge and mass of the impinging particle is indeed at the basis of the particle identification techniques exploiting CsI(Tl). Thanks to our analysis we have been able to characterize these fundamental parameters up to the  $Z=6$  within the energy range 10-250 MeV, and the gained knowledge has been used both to optimize the already existent particle identification techniques (Fast vs.  $E_{\text{CsI}^r}$ , Slow vs.  $E_{\text{CsI}^r}$ , Fast vs. Slow) and to individuate novel possible identification matrixes (Slow/Fast vs.  $E_{\text{CsI}^r}$ , Rise Time vs  $E_{\text{CsI}}$ ); their performances have been assessed and in particular at relativistic energies their figures of merit show higher value with respect to the standard aforementioned identification matrixes.

The first on-beam test for four FARCOS telescopes took place at the Laboratori Nazionali del Sud of the INFN (Catania), with

the main purpose to test the overall system functionality and to probe the light yield non-uniformity of the CsI(Tl) crystals. Several known reaction were exploited; the accelerated beams were 62 MeV/u protons and  $\alpha$ -particle and 20 MeV/u  $^{20}\text{Ne}$  ions on different targets Au,  $\text{Al}_2\text{O}_3$ , Pb, Al,  $\text{CD}_2$ . Thanks to our digital DAQ we have been able to record and process off-line the detectors signals. Using a mixed nuclei  $\alpha$ -source located inside the experimental chamber it has been possible to perform a strip-by-strip energy calibration and the extraction of the energy resolution, which for the  $^{239}\text{Pu}$  line at 5156.6 keV show an average value for the front side strips of 27.1 keV FWHM, while for the back side strips 44.7 keV FWHM. Exploiting the pixelation of the DSSSD and the elastic scattering peak of the beam it has been possible to map the CsI(Tl) light yield non-uniformity – always below 1% – and to correct the identification matrixes thus increasing the CsI(Tl) energy resolution (0.65% FWHM of the elastic scattering peak of the 62 MeV/u  $\alpha$ -particle). The overall performances, though susceptible of further improvements, are already excellent.



# EXPLORING POWER RELIABILITY AND PERFORMANCE ASPECTS IN ON-CHIP NETWORKS FOR MULTI-CORES

Daive Zoni - Supervisor: Prof. William Fornaciari

## Extract

For the last decade in the 20th century and early years of the 21st century, the processor design trends were dominated by increasingly complex feature sets focused on single core processors, exploiting instruction level parallelism (ILP), thread level parallelism (TLP) and higher clock speeds. Such an increase in performance for single core processors has been fueled by the technology scaling and manufacturing advantages that have driven the increasing integration capability of CMOS devices. Predictions attributed to the well-known Moore's law allow nowadays chips to host up to billion transistors and more, thus allowing single core processors to provide enough architectural abstractions to run multiple applications in parallel. However, the continuous integration, motivated by the request for even more performance, had severe drawbacks, posing constraints on device power, heat and reliability, that most of the time turn in architectural trade-offs. While performance has been the major design metric in last decades, power consumption cannot be neglected in current designs. Thus, computer industry has been forced to shift to multi-cores as one of the best ways to address the power issues, while maintaining

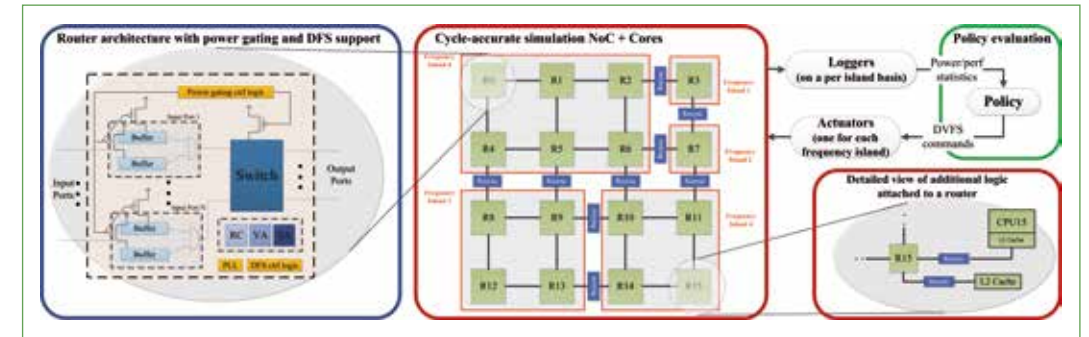
performance. In this scenario, multi-core architectures seem to be the only plausible solution to meet the performance, power and reliability requirements of a wide variety of applications, targeted for a general-purpose chip multi-processor (CMP) and the embedded system-on-chip (SoC) environments. The multi-core revolution allows to better use the computational power providing highly parallel architectures to run multiple applications. Moreover, the *International Technology Roadmap for Semiconductors (ITRS)* projects an increasing demand for higher performance, which in turn would require chips with 100x more cores than the current state-of-the-art designs.

*In this scenario, the need for huge and flexible interconnection resources to properly interconnects all the chip cores highlights the limits of classic bus-based architectures and Networks-on-Chip (NoCs), also called On-Chip Networks (OCNs), emerged as a viable solution to deliver such requirements. All in all, the multi-core interconnection fabric design is a critical component in current and future designs due to its non negligible power consumption, its strict performance dependence with*

*overall chip performance and the reliability properties due to its intrinsic distributed paradigm. At the same time the NoC design require a fresh view, where the interdependencies between cores memory hierarchy and interconnection must be jointly accounted at each design stage.*

The interdependence between the NoC and the other subsystems, i.e. cores and memory hierarchy, as well as the orthogonal power, reliability and performance constraints pose severe challenges in the proper design and optimization of the multi-core interconnection fabric. Moreover, the design of novel microarchitectural solutions to address the trade-off between the orthogonal metrics are usually coupled with hardware actuators or mechanisms such as Dynamic Voltage and Frequency Scaling (DVFS) and Power Gating. However, the new microarchitectural solutions and the hardware mechanisms introduces both timing and power overheads, which must be carefully evaluated to avoid the opportunity that they shadow partially or even totally the benefit of the proposed methodology.

The need to explore such a huge design space considering multiple orthogonal metrics,



1. Proposed research flow for the NoC power-reliability-performance exploration and optimization.

different chip components and timing and power overheads due to actuators requires accurate frameworks to support NoC evaluations and optimizations. Figure 1 comprehensively details the three main logical blocks of the proposed research which also represents the three main contributions of the proposed research.

### • HANDS framework –

An accurate simulation framework, the Heterogeneous Architectures and Network-on-Chip Design and Simulation (HANDS) flow, have been proposed starting from a collection of third-party state-of-the-art tools and in-house developed ones, it addresses the design-time and run-time *exploration/estimation* of power/performance and reliability/performance aspects in the multi-cores and NoCs. (See red panel in Figure 1).

### • Accurate actuator models

- Power gating and DVFS simulation support have been integrated in the HANDS starting from their accurate analytical models. Furthermore, the developed SPICE models allowed for each of them allowed to integrate accurate

timing and power overheads. The proposed actuators have been coupled with the NoC model enabling the possibility to use them on a per router basis up to islands of routers of different sizes, thus enabling the *exploration/optimization* of power, reliability and performance metrics (See blue panel in Figure 1).

### • Power-performance-reliability methodologies

- Three different NoC optimization methodologies have been proposed focusing on power, reliability and performance metrics. In particular, the first proposal focuses on the dynamic-power/performance trade-off in NoC routers exploiting the DVFS model by implementing a control-based methodology. The second proposal concerns static-power/performance optimization of routers' buffers, which is the main static power source. It effectively exploits the power gating support offered by the implemented simulation flow. Moreover, the proposal takes into account the NBTI aging mechanism, that is a physical phenomena which tends to degrade the CMOS

logic performance over the time. The last methodology focuses on static operating frequency assignment in a multi-core to reduce the power consumption, with the final objective to control the thermal profile. A linear model of the system has been created to optimally set the frequencies considering several figures of merit. (See green panel in Figure 1).

The proposed research addressed the optimization of different orthogonal metrics in the NoC design focusing on both simulation framework and methodologies. Moreover, a careful evaluation of the timing and power overheads due to DVFS and power gating mechanisms as well as them of the proposed new microarchitecture is provided, thus delivering a complete flow which can be used as a starting point for further investigations.