



## DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY

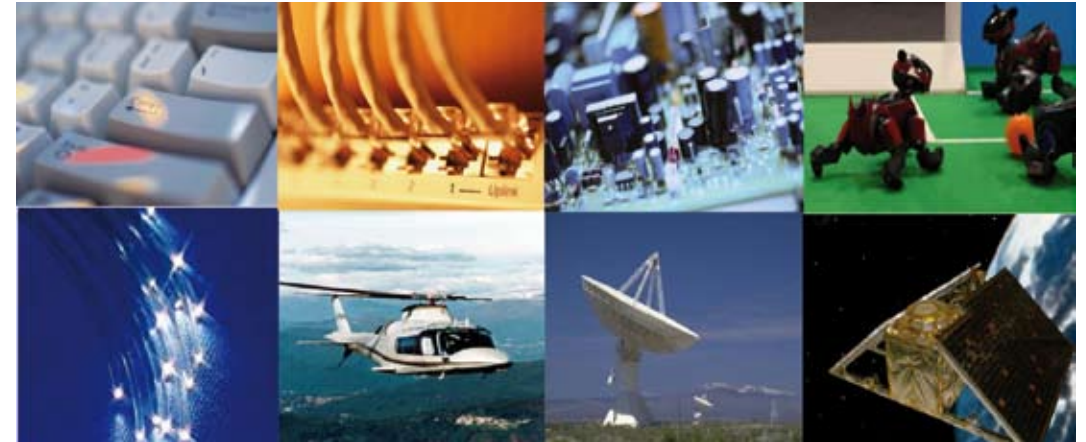
Chair:  
**Prof. Barbara Pernici**

The PhD program in Information Technology 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 e Informazione (DEI). 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 School is the largest at the Politecnico in terms of number of students. There are more than 50 first year students and therefore about 150 total. The students are subject to an examination every year to evaluate the progress achieved in their research and course work. Before the end of the second year, the students have to submit a minor project (whose topic has to differ from that of the major) evaluated after a public seminar.

### Topics

The research carried out in the department (including 35 between computing or experimental laboratories) can be subdivided into 4 main areas:

- *Computer Science and Engineering* (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* (coordinator Prof. Carlo Fiorini): 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* (coordinator Prof. Paolo Rocco): 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* (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 liason

Due to its truly 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 DEI 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 an extra budget offered by 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 the Wuhan University, China, that includes collaborative research, lectures, symposia and exchange of researchers, students and information; the cooperation with the Center for Communications and Signal Processing of the New Jersey Institute of Technology aimed at PhD co-tutoring in the area of wireless communication systems; the agreement for a joint Doctorate in Information and Communication Technology

between DEI and the TELECOM Paristech (former ERNST), Paris, France; double PhD Program with Drexel University. The PhD program in Information Technology participates in ICO-NEH (International Curriculum Option of Doctoral, e 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 Spa)

Marc Dacier (Symantec Research Labs Europe)

Alberto Daprà (CRISP, Interuniversity Research Center for Public Services)

Riccardo De Gaudenzi (European Space Agency)

Stefano De Panfilis (Engineering Ingegneria Informatica Spa)

Stefan Debald (Microsoft FAST)

Giuseppe Fogliazza (MCM Spa)

Flavio Gaj (Unicredit Spa)

Jana Koehler (IBM Zurich Research Laboratory)

Nathan Krishna S. (IBM Zurich Research Laboratory)

Luca Tonelli (LMT Europe Srl)

### Prizes and awards

In 2010 the following awards have been obtained by PhD students:

**Chorafas Foundation Award:** Alberto Dalla Mora, Federico Maggi

# METHODOLOGY AND ADVANCED ELECTRONIC ARCHITECTURES FOR HIGH PERFORMANCE DIGITAL PROCESSING

Andrea Abba

Today in many areas of modern science, from medicine to astronomy, from biotechnology to security, from climatology to environmental monitoring, from physics to communications, measuring signals with very high efficiency is a primary task for processing systems. In this, the term efficiency is linked to the specific application, such must be understood as a resolution when measuring signals from radiation sources in medicine and physics, and instead as bit-rate and bit-error-rate in communication systems. A fortiori, the methodology by which we proceed to implement the measure is closely related to the application. This makes it impossible to set a research of digital signal processing in general terms of methods and techniques as that would confine the work to an abstract level. Conversely should embrace a bottom-up process starting from case studies that allow applications to induce and focus in general terms the followed methodology. Therefore, my research activity dealt with different application scenarios. Surely, as a paradigm of how to develop this inductive process, I mainly focused my research in a strategic application area in modern physics and medicine, namely the high resolution measurement of signals from radiation sources.

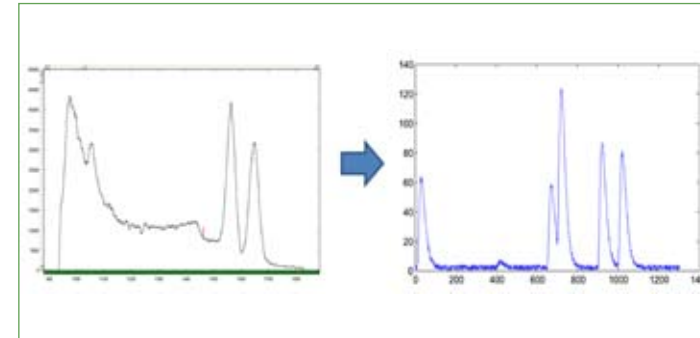
In recent years, digital signal processors have been impulsive alternatives to traditional analog systems for implementation of research progress in many fields of use of radiation analysis, from chemistry of matter to medical and astronomical investigations, from environmental monitoring to security systems. The massive interest in these systems stems primarily from their inherent adaptivity, the versatility of calibration and the potential to achieve signal-to-noise ratio very close to the optimal value predicted by the theory. The use of devices for "digital signal processing", including DSP and programmable logic FPGA, maximizes the measurement accuracy associated with the events detected and minimizes the power dissipation with respect to analog solutions. Moreover, digital processing allows adaptive data elaboration according to noise spectrum, disturbance presence and position of each pulse in the sequence to be processed. The types of signals and their distribution characteristics in amplitude and time require the development of innovative systems both at hardware and software level. The massive evolution of digital processors for radiation measurements has put in evidence the extreme convenience to develop more and more sophisticated



1. Prototype of the emulator of detection system

techniques for emulating acquisition systems up to simulate the physical radioactive source. Several advantages trigger this trend, starting from improvement of experimental conditions in absence of radioactive source and detecting apparatus, which means health safety of experimenters and possibility to perform remote experiments independently from the presence of the radioactive source and detection system. Also quality of the experiment is positively affected. In fact, the availability of the configurable virtual signal source simplifies testing of processors, allows absolute and fair comparison among different processing techniques, permits to directly evaluate algorithms and adjust the processing flow, opens the way to the existence of central banks of experimental data to be accessed remotely. But, radiation source emulation is the first step in the concept of these

emulation systems. I present a digital fully configurable architecture that performs the function of signal generation for emulation of radiation detectors and front-end electronics and the function of signal processor. The combination of a complex



2. The processor converts an energy spectrum into an analog signal similar to the output of a real detection system

and dedicated hardware architecture with a firmware environment that combines heterogeneous computing scenarios, allows to create in hardware a virtual platform that faithfully represents the complete detection-acquisition-processing chain (see Figure 1 and Figure 2).

Building on the analysis of signals from radiation detectors, several general and innovative techniques in sampling, interpolating, triggering signals have been developed. I have presented a method conceived for measuring time intervals with accuracy of few picoseconds, which is based on phase measurements of oscillating waveforms. The oscillation is generated by triggering an LC resonant circuit, whose capacitance is pre-charged. By using high Q resonators and a final active quenching of the oscillation,

it is possible to conjugate high time resolution and a small measurement time, which allows a high measurement rate. Experimental tests show the feasibility of the method and a time accuracy better than 4 ps rms.

for small animals studies. By the way, this is also a fundamental scenario of pharmacology research. The activity has been carried out both in terms of hardware and software (see Figure 3) taking into account also the engineering aspects of development strategies.



3. Multi-channel processor for medical imaging purpose

Medical imaging has been a real reference task of my research that was in charge to develop signal processing electronics of a new Anger camera to be used in clinical and research environments in specific applications where high overall spatial resolution and system compactness are required. The camera is based on the use of a matrix of Silicon Drift Detectors (SDDs) coupled to CsI(Tl) crystals. Two prototypes were realized during the project, with 5x5 cm<sup>2</sup> and 10x10 cm<sup>2</sup> FOV. They offer a high intrinsic spatial resolution (< 1mm), an overall spatial resolution of ~ 2.5 mm @ 5 cm and appropriate sensitivity. The developed camera is compact, very versatile and has a potential to be employed in several imaging applications, for clinical studies on humans, for small organs imaging in adults and infants, and could be incorporated in systems for both planar and SPECT acquisition

## A DATA QUALITY BASED METHODOLOGY TO IMPROVE SENTIMENT ANALYSES

**Donato Barbagallo**

Most companies concur that the Web has become an invaluable source of marketing information, as a very large, rich, and constantly updated knowledge base. Moreover, word of mouth, that found its great renaissance in Web applications such as social networks, is the primary factor behind 20 to 50 percent of all purchasing decisions. Its influence is greatest when consumers are buying a product for the first time or when products are relatively expensive, factors that tend to make people conduct more research, seek more opinions, and deliberate longer than they otherwise would. And its influence will probably grow: the digital revolution has amplified and accelerated its reach to the point where word of mouth is no longer an act of intimate, one-on-one communication. Today, it also operates on a one-to-many basis: product reviews are posted online and opinions disseminated through social networks. Some customers even create Web sites or blogs to praise or punish brands.

Monitoring the Web is seen as a real-time alternative to costly paper-based marketing surveys. Furthermore, automated monitoring can provide continuous as opposed to occasional feedback.

Unfortunately, managers also believe that existing tools are immature and, since critical decisions would be taken on the basis of Web monitoring information, they are admittedly cautious. In particular, they have pointed to a need for objective evaluations of the quality of the data produced by automated Web reputation analyses and for evidence of their dependability for decision making. The literature lacks such evidence. In particular, there is a lack of scientific approaches providing vendor-independent evaluations of data quality. The main reason for this literature gap is a technical difficulty of data quality analyses in this domain. In short, assessing the quality of Web-based reputation analyses requires algorithms to assess the quality of the input data, an indepth knowledge of the technical architecture of the software tools supporting such analyses and, ideally, a reference tool that provides a high-quality output that can be considered a benchmark for comparisons. The tools supporting the automated analysis of Web data are very complex and building a reference tool represents an open research challenge.

It is necessary to observe that data quality is a key variable for these tools, since these analyses are subject to errors

due to the non-structured and semi-structured nature of Web data. In particular, extracting an opinion from a text and assessing the degree to which the opinion is positive or negative is technically error prone. Anyway, the automated interpretation of natural language is still subject to error, but has reached enough quality to find broad application in document management and knowledge management applications. In this respect, Web information raises new challenges related to the low quality of text of Web conversations. Past contributions focus on web data reliability, based on the age of data, and do not consider the quality of content.

If text comes from official information, such as online newspapers, it can be assumed to be reliable, mostly well-written and easy to interpret. If it comes from Web 2.0 sites, such as microblogging, it must be cleaned before interpretation. This data cleaning component represents the first critical component that is missing in most existing tools and will be implemented and tested as part of this research work.

Furthermore, as well pointed out in a special report published by the Economist on February

27th 2010, one of the biggest challenges for future ICTs is overcoming the massive information overload due to the increasing availability of data acquisition and exchange. As a consequence, the definition of feasible methods able to both assess the quality of the information and to evaluate its relevance to specific tasks is crucial.

In the relevance evaluation, semantic tools are assumed to provide more dependable evaluation compared to non-semantic tools. However, this research will test this assumption. Testing has been performed with and without the data cleaning module. The software necessary to activate the semantic interpretation of natural language and perform related quality tests represents the second component that is missing in current literature and will be implemented and tested as part of this research work.

A third component is missing in both the literature and existing tools and has been implemented as part of this research to test the impact of the reputation of Web sources on the quality of the results of our analyses. To the best of our knowledge, there is a lack of literature dealing with the dependability of the data that are being analyzed and on the weighing functions that balance the different trustworthiness and reputation of Web sources.

Trustworthiness on the Web is often identified with popularity: this equation led to the success of the PageRank algorithm even if it does not necessarily conveys

dependable information since highly ranked Web pages could be spammed. To overcome this issue, new algorithms are based on hub and authority mechanisms in the field of Social Network Analysis (SNA). Especially when considering services such as forums, in our approach we assume that it is important to evaluate even a single contribution: SNA can be used to evaluate each author's trustworthiness. The selection of sources providing dependable information has been scarcely based on the definition of methods for assessing both software and data quality. However, the concept of reputation is the result of the assessment of several properties of information sources, including correctness, completeness, timeliness, dependability, and consistency. It should be noted that the data quality literature provides a consolidated body of research on the quality dimensions of structured data, their qualitative and quantitative assessment, and their improvement. The analysis of the quality of semi-structured and unstructured data and of trust-related quality dimensions, and, in particular reputation, are however still an open issue. The three software components that have been developed as part of this research can provide interesting insights from both a practical and scientific standpoint. The output of this thesis is a tool, called SentiEngine, that performs sentiment analyses considering all the variables discussed. The tool presents several innovative aspects:

- a methodology to design semantic models that are

analyzed by sentiment analysis tools.

The reuse of well-established techniques and the design of an effective model are two fundamental steps for the quality of the final analyses.

- A set of algorithms that perform sentiment analysis in the case of Web 2.0 data, whose performance have been compared with a market benchmark tool, which is outperformed in terms of precision in several experiments.
- A framework to assess Web information sources. The analyses presented in this chapter have shown how research result performed by the well-known Google search service present some mismatch with metrics indicating the participation of the users, while metrics concerning traffic and timeliness are concordant.
- The design of a mashup-based dashboard that support users in the selection of relevant content, by assessing the reputation of Web information sources.

## ENABLING REASONING UPON STREAMING DATA

**Davide Francesco Barbieri**

Streaming data is an important class of information sources, which is gaining more interest as the Web increases in size in terms of data amount and traffic generated per second. Examples of data streams are Web logs, click streams, locations of mobile users, feeds, sensor data, stock quotations, and so on. Streaming data is received continuously in real-time and can not be stored due to the amount of memory the recording process would need. The data items in a stream can be either implicitly ordered by arrival time, or explicitly associated with timestamps. Streams can not be treated as persistent data to be stored and queried on demand, but rather as transient data to be consumed on the fly by continuous queries. Continuous queries, after being registered, keep analyzing such streams, producing answers triggered by the streaming data and not by explicit invocation. The problem of managing and querying this particular kind of data has been solved in the last decade through the development of a particular class of database systems, known as Data Stream Management Systems (DSMSs), which are capable of querying portions of simple relational streams, applying a selection criteria to them. DSMSs and Complex Event Processors (CEPs) cannot perform complex reasoning tasks. Reasoners,

on the other hand, can perform complex reasoning tasks, but they do not provide support for managing *rapidly* changing worlds. While reasoners scale up in the classical, static domain of ontological knowledge, reasoning upon rapidly changing information has not been deeply investigated so far. In this thesis, data stream and reasoning technologies will be combined, starting from the terminology. Database (DB) and Knowledge Engineering (KE) communities often use different terms to indicate the same concepts. DB community distinguishes among schema and data, whereas KE community distinguishes among *factual*, *terminological*, and *nomological* knowledge. The notion of data is close to the notion of factual knowledge, and similarly the notion of schema is close to the notion of terminological knowledge. Nomological knowledge is information about rules defining actions and action-types and governing relationships in a given culture or society (e.g., when it rains, traffic gets slower); this notion is somehow captured by constraint languages for DBs, but it is mainly peculiar of KE. From now on, the concept of “*knowledge*” will refer both to terminological and nomological knowledge (thus the notion of database schema is included in

the terminological knowledge), while “*data*” will be used as a synonymous of factual knowledge. Knowledge and data can change over time. In order to classify knowledge and data according to the frequency of their changes the notion of “*observation period*” is introduced, defined as the period when the system is subject to querying. In the context of this work, knowledge is considered **invariable during the observation period**; only streaming data can change. Of course, knowledge is subject to change, but this part of evolving knowledge is not subject to observation. The combination of reasoning techniques with data streams gives rise to **Stream Reasoning**, an unexplored, yet high impact, research area. Central to the notion of stream reasoning is a paradigmatic change from persistent knowledge bases and user-invoked reasoning tasks to transient streams and continuous reasoning tasks. For example, in the context of traffic monitoring, invariable knowledge includes terminological knowledge (like address data - e.g. street name, civic number, city name, ZIP code, etc.), and defines the *conceptual schema* of the application. The description of how the world is expected to change (e.g., given traffic lights

are switched off, some streets are closed during the night, traffic jam happens more often when it rains or when important sport events take place), also falls in the invariable category of knowledge. Traditional databases are suitable for capturing a relatively small quantity of knowledge in their schema and huge datasets of both invariable data and event driven changing data whose mean time between changes is slow or medium. Periodically changing data can be modeled by means of triggers performing the updates. In the database world Data Stream Management Systems (DSMS) have represented a paradigm change because they move from persistent relations to transient streams, with the innovative assumption that streams can be *consumed* on the fly (rather than stored forever) and from user-invoked queries to *continuous queries*, i.e., queries which are persistently monitoring streams and are able to produce their answers even in the absence of invocation. DSMSs can support parallel query answering over data originating in real time and can cope with burst of data by adapting their behavior and gracefully degrading answer accuracy by introducing higher approximations. Current reasoners are suitable

for capturing large and complex knowledge, but at the cost of small datasets. Complex forms of periodically changing data can be modeled by means of rules. However, standard reasoners cannot cope yet with dynamic knowledge, as they consider static RDF graphs to carry out reasoning tasks, even though there is a potential interest for giving up with one-time semantics in RDF repositories and to explore the benefits provided by continuous semantics. One of the main targets of this thesis is to provide reasoners an access protocol for heterogeneous streams. As RDF is the most accepted format to feed information to reasoners, this thesis will introduce *RDF streams* as an extension of the RDF data model. An RDF-based representation of heterogeneous streams represents a paradigm shift that can facilitate interoperability between different systems, since SPARQL (the the standard language defined by W3C for querying RDF graphs) can query multiple sources of RDF graphs. In this thesis a new language named *Continuous SPARQL* or simply *C-SPARQL* will be introduced, defined as extension of SPARQL. The purpose of this extension is to provide a facility to seamlessly select RDF triples from both streams and RDF

repositories, providing support for continuous querying (and, therefore, reasoning) over data streams and rich background knowledge. The implementation of an execution environment for C-SPARQL is needed in order to evaluate the system performance and to validate the techniques proposed. In this thesis, a C-SPARQL execution engine will be designed, taking query optimizations into account. In particular, a materialization technique will be studied in order to prevent useless computations of the entailments already computed at a previous step of the continuous query evaluation. To evaluate the system real-world scenarios and data will be used whenever possible. The streaming data used for the tests will be also used to show some examples of RDF stream generation starting from a raw or relational data stream.

# DEVELOPMENTS IN SIDE CHANNEL ATTACKS TO DIGITAL CRYPTOGRAPHIC DEVICES: DIFFERENTIAL POWER AND FAULT ANALYSIS

Alessandro Barenghi

Due to the growing employment of computing platforms to manage, store and elaborate sensitive data, the issue of warranting confidentiality while these procedures are performed is now of prime interest. The employment of cryptography is the main means through which the aforementioned confidentiality is ensured, thus proper attention must be devoted to the study of reliable, practically implementable cryptosystems, together with the techniques required to embody them in physical devices. One of the great challenges of deploying a cryptographic algorithm on the field, is to take into account the possible attack channels which derive from its physical realization: in fact, despite the theoretical soundness of the standardized primitives, it is possible to breach the security provided by the most sound algorithms, in the case their implementation leaks information on a side channel, be it timing, power consumption or a malfunctioning of the computing circuit. The interest throughout the academical community is high, since the exposure of hardware and software platforms to side channel attacks has been considered very marginally, and the industrial interest is high as well, since current mobile devices are largely exposed to this kind of threats.

The so-called side channel attacks are usually split into two categories, active and passive, depending on whether the attacker needs to actively disturb the execution of the cryptographic primitive on the device (hence these attacks are also called fault attacks) or relies only on the observation of the leaked information without any physical intervention. The overall contents of this thesis is to overview the state of the art in the field of active and passive side channel attacks, and to propose some advancements in both directions. The final purpose is to give a better understanding of this quickly evolving research field, to fill some still rather unexplored areas, to exploit unemployed techniques and possibly to suggest viable countermeasures to defeat the attacks.

The first part of this thesis will present two practically feasible attack techniques on the two most widely adopted cryptographic primitives, the RSA asymmetric key cryptosystem and the AES symmetric key cryptosystem.

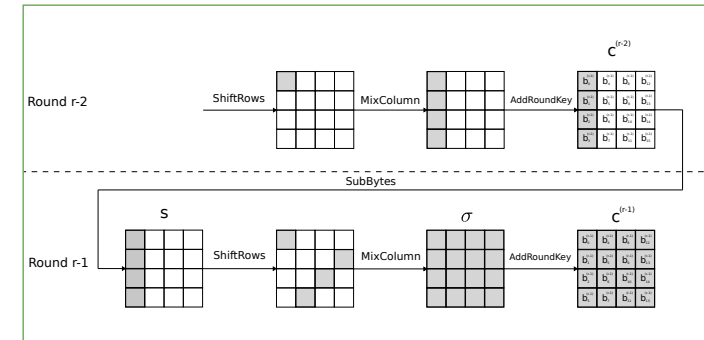
The new attack based on injecting faults during the RSA encryption primitive, is able to successfully decrypt a message without any knowledge of the secret key whatsoever.

The attacker only needs to obtain a faulty and a correct encryption of the same message and the plaintext can be recovered in a few minutes computation on a common desktop.

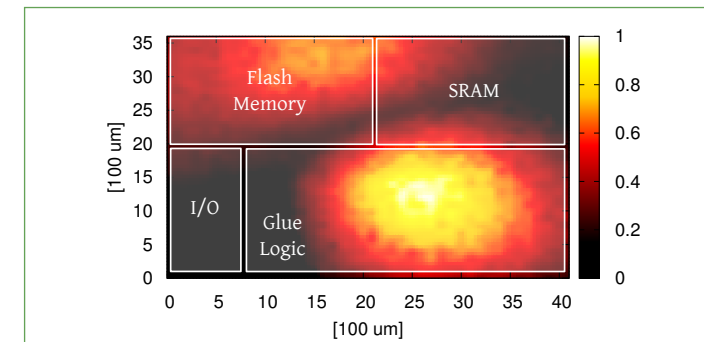
This thesis also presents a new attack technique able to practically recover the whole expanded key of any instance of the AES, regardless of the length of the key schedule, the number of rounds of the cipher and the key scheduling strategy employed to expand the original cipher key.

A key point to deem these fault attacks a realistic threat even when considering a possible attacker having access only to limited resources, is to develop a practical, non-invasive, non-tamper evident technique. This thesis presents the results of the research on the employment of a reduced power supply voltage in order to gracefully induce faults in a computing device. The results obtained show that it is possible to induce well controlled single bit faults in an ARM926 processor, regardless of the system on chip it is embedded in, provided accurate enough supply voltage control is available to the attacker.

In the case a completely passive attack is desired,



1. The spread on a single fault caused by the AES cipher it is possible to successfully which is useful in order to lead



2. Map of the intensity of the EM emissions of a chip while computing the AES cipher. Clearer zones show where the emissions are stronger.

infer the value of the secret key, stored in a non-extractable way, from the analysis of the power consumption of the cryptographic device that is employing it. This kind of analysis relies on correlating the power consumption during the execution of an operation of the algorithm under attack involving directly the secret key. Knowing the switching activity that will be triggered by a specific value of the key, it is possible to infer the actual value employed in the circuit through collecting a reasonable number of measurement from different runs of the algorithm, in order to compensate for the measurements variations.

The new contribution in this thesis is to determine that the correlated consumption,

the attack, is carried over a very narrow frequency band of the spectrum of the recorded consumption signal. This in turn has made possible to employ digital signal processing techniques aimed at suppressing a vast amount of measurement noise, thus leading to a tenfold reduction in the number of measurements needed to perform a successful attack against a 32-bit CPU running a software implementation of AES, a scenario widely recognised as difficult to attack.

The effectiveness of side channel attacks is better justified if they are compared to mathematical cryptanalytic techniques. To the purpose of completing the evaluation of the strength and performance of the most commonly employed ciphers,

this thesis also presents contributions in the field of efficient implementation of cryptoschemes for both common use and brute force attacks.

The protection of data at rest is one of the most common scenarios where high throughput and low cost cryptographic accelerators are desired. Through exploiting the properties of the IEEE P1619 standardized mode of operation for data encryption, it is possible to utilize common graphical accelerators in order to speed up the encryption of large quantities of data.

This thesis presents results regarding the obtained speedup in hard disk encryption using nVidia GT200 GPU cores, obtaining beneficial effects on the encryption throughput at a low cost.

It is also possible to exploit the large amount of computational power provided by GPUs to perform fast brute force attacks to weak ciphers. In this work it is presented a fast, bit-sliced implementation of the DES cipher, able to reach several millions of encryptions per second on a single graphic card, thus bringing a practical brute force attack to DES in the realm of practical feasibility, with only a small amount of commodity hardware.

## PILOT-AIDED CARRIER RECOVERY IN THE PRESENCE OF PHASE NOISE

Luca Barletta

In digital communication systems carrier recovery is required if the signal is detected coherently. Basically, two different approaches help to deal with the carrier recovery at the receiver: the first one is represented by the insertion of a pilot tone in the transmitted signal that allows the receiver to reconstruct, employing a PLL, the carrier waveform; the second one consists in the estimation of the carrier parameters exclusively based on the modulated signal, and takes the name of suppressed carrier transmission. We focus our attention on the latter approach, as this is prevalent in practice.

Noise in oscillators represents a major issue to deal with, because of its dramatic impact on the frequency spectrum of the oscillator. This phenomenon is known as *phase noise*, and it deeply affects the BER performance of communication systems, acting as a multiplicative noise in single-carrier systems and as an inter-channel interference in multi-carrier systems. In the high signal-to-noise ratio regime, the phase noise dominates the performance causing a floor in the system's BER performance, thus the synchronization problem is dominated by the compensation of the oscillator's phase noise. Typically, the synchronizers

adopted in suppressed carrier systems fall into one of the following categories: in data-aided synchronization known symbols (referred as *pilot* symbols) are time-division multiplexed with data symbols to help the carrier estimation; in non-data-aided synchronization (known also as *blind* synchronization) the receiver operates a non-linearity prior to the phase estimation in order to remove the phase modulation carried by the data sequence; finally, in decision-aided synchronization the estimator operates either a hard or a soft decision on the partially synchronized data sequence.

In communication systems working with high spectral efficiency, the growing of the constellation cardinality causes the narrowing of the *S*-curve of the phase detector of the conventional PLL scheme. The use of a decision-directed scheme makes the receiver prone to the phenomenon of *cycle slips*. In order to combat the cycle slip phenomenon, in this work we adopt a pilot-aided modulation: pilot symbols are inserted in the data stream to provide an absolute carrier phase reference, that helps to restore the correct working point of the *S*-curve of the phase detector and consequently to lower the probability of having a cycle slip.

Data-aided parameter estimation, known as pilot symbol assisted modulation (PSAM), has been originally analyzed by Cavers in the context of channel amplitude estimation in the presence of Rayleigh fading. Clearly, although this work focuses on carrier phase estimation, one can advantage the use of pilot symbol to jointly estimate many parameters, such as channel amplitude, frequency offset, and so forth.

The key issue we address is the derivation of the optimal filter that estimates the actual phase of the carrier by taking advantage of the available pilot symbols. We analyze the *mean-square phase error* (MSPE) performance of the single-carrier system when a continuous sequence of equally spaced blocks of pilot symbols is multiplexed with the payload data stream. We mostly focus on the basic case where pilot symbols are isolated, and then generalize our results in the case where blocks of pilot symbols are transmitted maintaining the pilot rate fixed.

One of the main result of this work is the full comprehension of the role played by the aliasing of the carrier spectrum, induced by the down-sampling of the carrier phase implemented by pilot symbols. This result puts light on the design of the pilot rate and on the cyclostationary

nature of the recovered phase, hence of the MSPE performance. The major shortcoming of the article by Meyr et alia is that of neglecting the aliased components of the carrier to be recovered, making its results useless in the regime of bad carrier and/or good channel.

With reference to the AWGN channel, for the case of purely phase modulated (PM) pilot symbols, we propose an analysis based on the discrete Wiener-Hopf method, whose time-continuous counterpart descends from the paper by Bode and Shannon, that serves to derive the optimal pilot filter in the MSPE sense. We explicitly derive the optimal pilot filter by exploiting the cyclostationary nature of the underlying system model. Then we find a suboptimal and low complexity solution that approaches the optimal MSPE performance. When pilot symbols are both phase and amplitude modulated (PM-AM), we show that the system model is no longer cyclostationary and then we resort to the Kalman theory for deriving the optimal time-varying pilot filter. We demonstrate how the Kalman and the Wiener-Hopf approaches lead to the same results in the cyclostationary case. We also propose a suboptimal time-invariant filter that approaches the optimal MSPE performance. For both the PM and PM-AM pilot symbols, we work out specific results in the popular case where phase noise is modeled as random phase walk, focusing our attention on the 1-causal and on the unconstrained-length realization of the optimal filter.

We finally take into consideration the problem of pilot-assisted carrier recovery on the ISI channel by analyzing the post-equalization carrier spectrum. Specifically, we analyze the impact of the equalization of the pilot symbols on the post-equalization carrier spectrum in two cases: the first one is where the phase noise impairment is dominated by the receive oscillator, and the second one is where both the transmit and the receive oscillators provide comparable phase noise contributions. We specialize results for the two-ray channel, that is often encountered when modeling point-to-point microwave radio links. We then argue on how this analysis allows to use pilot symbols in time domain in the problem of carrier recovery in the multi-carrier scenario.

In chapter 2 some basics on the Wiener and on the Kalman theories are recalled, in order to present the mathematical tools and the notation that are used in the subsequent chapters. In chapter 3 we firstly formulate the problem that is addressed in this work, and describe the channel impairment represented by the phase noise and its impact on the final error-rate performance. Then, by considering a transmission over the AWGN channel, we propose the solution of the problem in the following cases:

- periodic transmission of isolated and phase-modulated pilot symbols,
- periodic transmission of isolated and phase-amplitude-modulated pilot symbols,
- periodic transmission of blocks of consecutive phase-

modulated pilot symbols. Moreover, we advance the solution to a simplified model where the phase noise is assumed to be constant in each block of pilot symbols. In chapter 4 the popular case of first-order random-walk phase noise is widely analyzed. We present the model of the first-order phase noise and obtain the statistical characterization that serves to find the optimal estimator. Then we derive closed-form expressions of the optimal pilot filter and its MSE performance in the case of phase-modulated pilot symbols. The optimal MSE performance is compared with the performance of a sub-optimal filter. For the case of phase-amplitude-modulated pilot symbols, the optimal one-step-ahead predictor and the optimal smoother are derived from the Kalman theory, and their MSE performance is compared with that of a sub-optimal time-invariant filter. In chapter 5 we analyze the problem of carrier recovery on the ISI-AWGN channel. We study the impact of the ISI channel and of the AWG noise on the carrier spectrum, that is of fundamental importance in order to determine the optimal pilot filter. In chapter 6 we apply the results derived in chapter 5 to the multi-carrier modulation (OFDM). We propose a pilot scheme that allows to tackle the phase noise impairment in the time domain. Finally we compare the error-rate performance of our proposed carrier recovery scheme with an other present in the literature. Finally, conclusions are drawn in chapter 7.

## NAVIGATION STRATEGIES FOR EXPLORATION AND PATROLLING WITH AUTONOMOUS MOBILE ROBOTS

Nicola Basilico

Autonomous mobile robots are a promising technology that inspired, during the past years, different research activities devoted to address the several challenges posed by the complex interactions between the robots and their environment. Intuitively, an autonomous mobile robot can perform a task without continuous human supervision. One of the most important advantages of this technology is that autonomous mobile robots can be employed for tasks that would be difficult, dangerous, or simply boring for humans.

Designing autonomous mobile robots that can execute a task without any human supervision can be very useful in a large number of applications. Indeed, there are situations in which the human telecontrol is impossible or it is simply not convenient. Moreover, the human intervention can be subject to errors that can worsen the execution performance or even compromise the successful completion of the task. For these reasons the need for a stronger level of autonomy, that can be denoted as *full autonomy*, has become important. A fully autonomous robot integrates in its control architecture a planning system that can operate at two different levels of abstraction. At the higher level, a

global task is specified and the robot has to find out the set of actions to achieve it. At the lower abstraction level, the robot computes the set of low-level operations to perform a given action. A rough distinction between the two levels can be outlined by saying that at the higher one the robot has to autonomously determine *what* to do by making corresponding *decisions*. Differently, at the lower level it has to determine *how* to execute such decisions by computing the corresponding *plans*. This thesis is about techniques to design strategies, i.e., to equip mobile robots with the ability to make decisions at the higher planning level. An interesting approach to deal with this kind of problems is to exploit techniques from Artificial Intelligence and from Decision Theory. Generally speaking, a mobile robot can be modeled as an intelligent agent, able to interact both with the environment and with other agents populating it. Decision-theoretic models can then be applied to capture the agents' objectives, define how to measure the goodness of a decision, and compute a strategy. This dissertation focuses on the problem that an autonomous mobile robot faces when deciding how to exploit its mobility to complete a given task or mission, namely

on the problem of deciding *where to move*. We refer to this problem as the definition of the mobile robot's *navigation strategy*.

A navigation strategy can be broadly defined as *the set of techniques that allow an autonomous mobile robot to answer the question "where to go next?", given the knowledge it possesses so far*. For example, in exploration a robot's navigation strategy could be to randomly select next locations or to simply follow a pre-computed trajectory. In general, the navigation strategy significantly impacts on the task execution's performance. Therefore, the problem is to define *good* navigation strategies, i.e., strategies that allow the robot to perform its task maximizing some performance metric or criterion. Despite their importance, a general satisfactory characterization of the problem of defining navigation strategies is still missing. General methods that can address wide ranges of applications and that can simplify experimental evaluation of navigation strategies have not been exhaustively studied. However, since navigation strategies are a fundamental component for autonomous mobile robots, different works

in literature dealt with them. The mainstream approach followed so far seems to adopt *ad hoc* solutions specifically tailored for the particular situation in which the robot is deployed, without any attempt to define a more general theoretical framework. Although many solutions have been proved effective in practice, this trend presents limiting drawbacks. First, it is difficult to compare different strategies with the aim, for example, of selecting the best one for a given situation. Moreover, modifying a strategy for being employed in different contexts or for improving it can require significant efforts. This demand for comparability and flexibility encourages the study of more general application-independent frameworks where the problem of defining navigation strategies can be cast.

The objective of this dissertation is to contribute along this direction by tackling the problem of defining navigation strategies from a more general perspective. We start from the idea that defining navigation strategies is a decision-theoretical problem and that several advantages can be obtained when applying techniques coming from this field. Following this approach, we will derive new original and interesting results within

the scope of two applications, namely exploration of unknown environments (for map-building and for search and rescue) and surveillance. Decision theoretical models are characterized by established formal foundations that can enable the development of more general and flexible navigation strategies. For example, some decision theoretical models allow one to easily combine together different criteria to drive the decision-making process or to model complex scenarios characterized by some degree of uncertainty or by the interaction with other agents. Exploiting general models also simplifies the task of evaluating and comparing different strategies. From this dissertation, it emerges that the employment of decision-theoretic techniques can, from the one hand, provide a robot with effective navigation strategies and, from the other hand, contribute to develop more flexible and comparable navigation strategies.

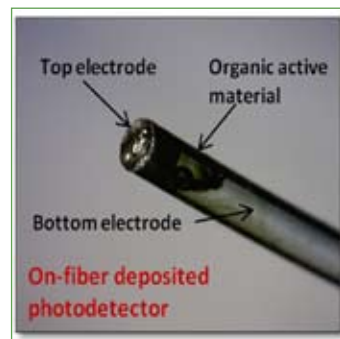


## LIGHT HARVESTING ELECTRONIC DEVICES BASED ON SOLUTION PROCESSIBLE SEMICONDUCTORS

Maddalena Binda

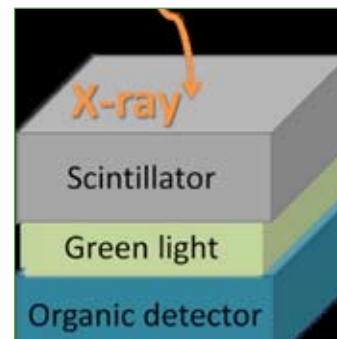
In the last few decades, the profuse research activity on solution processible semiconductors, particularly focused on organic semiconductors and quantum dots, has opened the new fascinating perspective of a low-cost electronics, easily conformable to large area applications and in principle compatible with very unconventional substrates, like light and flexible plastic ones. Nevertheless, it is in light exploiting applications that these materials show the most alluring potentialities, thanks to their typical luminescence in the visible range of the spectrum and to their high absorption coefficients. Indeed, the first profitable employment of solution processible semiconductors at the commercial level has been in the field of ultrathin displays, where OLED-based screens emerge for the ultrabright colors, the high contrast ratios and the low power consumption. Besides being exploited as light emitters, molecular semiconductors can be employed with good success as light harvesters for either signal detection or electrical power generation: in this thesis work, both these goals have been addressed. Particular attention has been

paid to the development of solution processed photodetectors and solar cells capable of efficiently operating in the near-infrared (NIR) range of the spectrum, which is not trivial given the chemical accessibility and stability issues affecting low bandgap compounds. Despite being quite similar in the architecture and in the working principle, photodetectors and solar cells are instead very different in the way they are operated (with or without a bias applied, under pulsed or continuous wave illumination,...), thus requiring the attention to be focalized on different parameters from time to time. In the first part of the thesis, the development of fast and efficient organic photodetectors for the NIR is described. Stability of the operation in air, which is a very uncommon feature when dealing with organic materials, is also demonstrated for these devices. To this stage of the work, a planar configuration for the photodetectors has been exploited: indeed, besides being more suitable in the characterization process of new materials due to their ease of preparation, planar structures possess some key features like the compatibility with the TFT technology, the freedom in the choice of the metal electrodes and the low intrinsic



1. Prototype of an organic near-infrared photodetector directly deposited onto the cut-end of a Plastic Optical Fiber (POF)

capacitance, that could make them preferable with respect to the more popular "sandwich-like" geometries for particular applications. The good results that have been achieved are the fruits of an optimization work carried on both at the chemical level, by suitably tailoring the solubility of the dye, and at the device level, by tuning the active layer composition and controlling its interaction with the substrate. The active material optimized in the first part of the work has been then exploited in vertical photodetectors, where remarkably high sensitivity, competing with the best examples from the literature, has been achieved by providing the devices with an additional organic layer for dark currents



2. Sketch of an organic-based pixel for X-rays detection. The indirect detection of X- radiation is performed by coupling an organic visible light detector with an inorganic scintillating crystal

suppression. Finally, in the last part of this section, the peculiarity of organic materials of being virtually depositable on every kind of substrate has been exploited to realize a working prototype of photodetector directly deposited on the end cut of a plastic optical fiber (POF) (Figure 1), meant for being employed in data transmission systems. In the second section of the thesis, a specific application is addressed, concerning the development of a prototype pixel for X-ray detection based on coupling the organic photodetector with an inorganic CsI(Tl) scintillator (Figure 2). The pixel is meant to be the first step towards the realization of a flat digital panel to be exploited for medical imaging



3. Organic photodetectors realized onto a flexible PolyEthylene (PET) substrate

purpose, aiming to profit from the peculiarities of organic electronics in terms of low-cost large-area processibility. Starting again from the active material optimized in the first part, a preliminary study that demonstrates the feasibility of X-ray detection through the scintillator/photodetector assembly has been carried on. Then, efforts have been dedicated to improve the sensitivity of the pixel by selecting more suitable organic dyes, better matching the emission spectra of the scintillator, and by suitably addressing dark currents suppression. The successful realization of pixel detectors both on glass and on plastic substrates (Figure 3) has demonstrated again the

feasibility of organic devices realization onto different insulating substrates, thus positively addressing large area digital panel sensors. In the third and last section, the development of solution processed hybrid organic/inorganic solar cells based on a lead sulfide nanocrystals active material is described. Infrared light collection up to  $\lambda = 1400\text{nm}$  has been demonstrated. Oxidation processes involving the nanocrystals film have been induced, studied and exploited as an effective way of improving the performance of the cells both in terms of open-circuit voltage and fill-factor. The physical mechanisms responsible for such improvement have been deeply investigated, stressing the key role of current leakage and interfacial charge recombination phenomena in limiting the efficiency of the cells.

# TIME DEPENDENT PHAENOMENA IN CHALCOGENIDE MATERIALS FOR PHASE CHANGE MEMORIES

Mattia Boniardi

In the overall Non-Volatile Memory (NVM) panorama, the current mainstream Flash memory technology is expected to reach fundamental physical standstill, preventing it from being scaled down to the next technology nodes, below 25 nm, in the respect of the Moore Law. Therefore, by the first years of the 21<sup>st</sup> century, several solid-state emerging technologies have been proposed, aiming to maintain the current scaling trends for the non volatile memory devices. Most of them rely on a storage operation mainly based on variable resistive elements, even though driven by very different physical principles and on a wide variety of materials. Among these technologies the Phase Change Memory (PCM) shows several interesting features such as fast read/write operation, low voltage operation and superior endurance properties, as well as great scaling potential. So it is claimed to successfully meet scalability, reliability and performance specifications. This doctoral work is focused on the Phase Change Memory (PCM) technology, and in particular on certain time dependent phenomena involved in the amorphous phase of the material employed in the PCM operation. Such material is a chalcogenide compound made by germanium, antimony and tellurium, the  $\text{Ge}_2\text{Sb}_2\text{Te}_5$ , or simply GST. The

PCM operation basically relies on the ability of GST to reversibly change its phase between two states by means of proper electrical pulses applied to the memory cell. Such states are the crystalline, low resistive state, associated to the logic 1 bit and the amorphous, high resistive state, associated to the logic 0 (see Fig. 1).

While the crystalline phase is microscopically and electrically stable against time, the amorphous one is metastable and gets involved with two subsequent phenomena, namely the structural relaxation and then the crystallization mechanisms. Both phenomena are classified as time dependent and thermally activated and both are involved with the electrical stability with time of the reset state of the PCM device. The deep comprehension of their fundamental physics is very attractive for the microscopic level description of the reliability mechanisms related to the amorphous state of the material that are involved in data retention and in the stability of reset and partial reset states. Despite structural relaxation and crystallization having been considered for long time as two basically different mechanisms, mainly due to their opposite effects on the memory cell electrical parameters, within this thesis work they will be

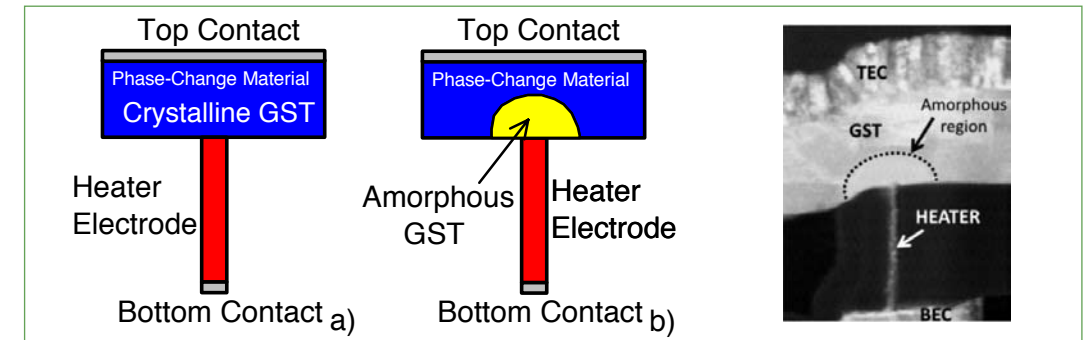
described for the first time into a unified framework, providing a microscopic common interpretation to them based on thermodynamic arguments and on multi-phonon interactions. Such interpretation will provide new insights from both the application and the amorphous physics comprehension standpoint. A temperature-accelerated investigation of the amorphous phase of GST oriented to its defect state spectroscopy, supported by its analytical modeling, will be straightforward for the common interpretation of the two phenomena in the framework of the Meyer-Neldel rule, also known as the compensation law for the Arrhenius phenomena. Such study aims to provide a microscopic picture of the thermally-activated barrier evolution during the two subsequent time dependent processes. This doctoral thesis is also oriented toward the array characterization and modeling at the statistical level. This is considered as an attractive activity for multi-megabit PCM array research and development, in which Numonyx, now Micron Technology, plays a leading role. An extensive statistical study of the structural relaxation phenomenon will be presented, aided by the full characterization of a PCM demonstrator chip of

the 180 nm technology node. A structural relaxation variability study on chip will be reported, allowing reliability predictions on the stability of the amorphous state evolution with time at the retention time specifications, on a large population of cells. The investigation, which is made

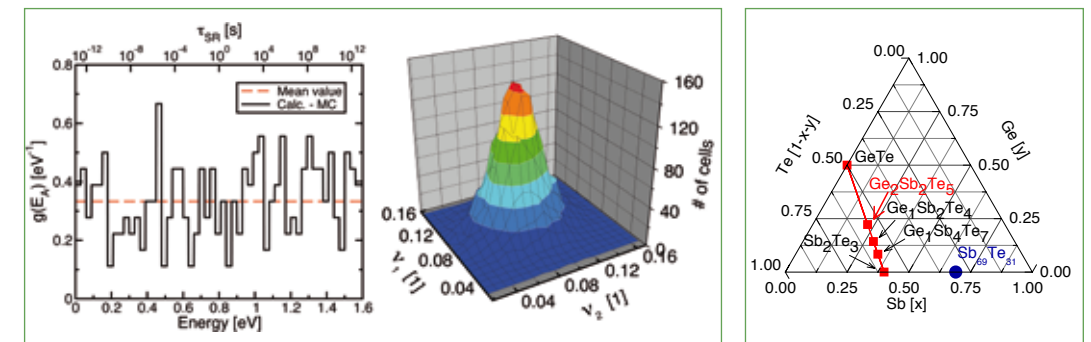
variability among the studied large population of cells.

Finally one of the most recent trend in the PCM research and development is the new material investigation: in fact it is widely recognized that the reliability and performance of such technology

and on the other hand to investigate and potentially enlarge the possible application spectrum for the PCM technology, at its debut on NVM market. A study made on different alloys will lead to identify some composition trade-offs dictated by the resulting electrical parameter



1. PCM cell working principle and TEM image



2. Simulated discrete defect distribution and non-autocorrelated SR experiment due to random defect nature in the amorphous network

3. Ge-Sb-Te compound system diagram

here on the full reset state for all cells, can be easily extended to all the possible resistance levels within the program window, aiming to be a useful physics-based guideline for resistance drift statistics investigation in multilevel cell devices. A Monte-Carlo modeling of discrete defect distribution (see Fig. 2) within the amorphous active volume of cells will be the straightforward way to the statistic description of the amorphous resistance drift

strongly depend on the particular phase change compound material.

So the last part of the work will be dedicated to the Ge-Sb-Te compound system investigation (see Fig. 3), aiming to provide on one hand some physics-based guidelines for phase-change material choice as a function of different applications, to best meet their reliability and performance specifications

trends among reliability specs, like program window, switching voltage, structural relaxation, crystallization and endurance. Among performance specs the time duration of the set operation is the most challenging parameter to study since it may allow to meet DRAM performance requirements, enabled by a high set speed material and hence to provide large replacement capabilities for the PCM technology.

## ELECTRICAL CHARACTERIZATION AND MODELING OF RESISTANCE SWITCHING NiO-BASED RANDOM ACCESS MEMORIES

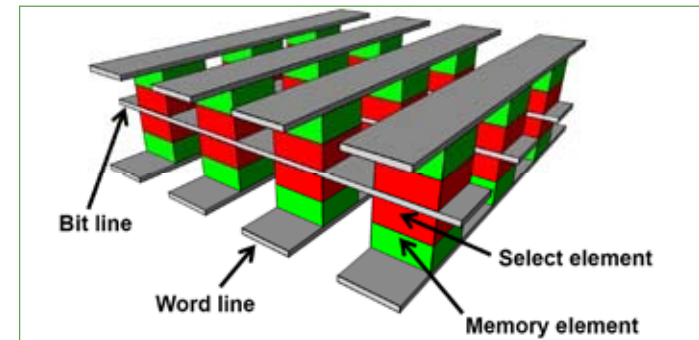
Carlo Cagli

Although in recent years there has been a huge increase in the storage capacity of integrated memories, still the technological solutions remained almost the same. If by one side SRAM and DRAM holds the record for programming speed, paying back their volatility in terms of conspicuous consumption, on the other side, flash memory represents currently the sole leader for the non-volatile storage, but it shows poor performances in terms of speed and endurance when compared with its alter-ego DRAM and SRAM. As we approach the scaling limit imposed by physics to flash technology, the research is making more and more efforts towards the development of new devices that are able to combine the characteristics of flash technology with the performances of faster memories, overcoming the scaling limits that will soon afflict the floating gate devices. Among many competitors, metal oxides-based resistive switching memories (or resistive RAM: RRAM), represent a great promise in terms of low cost and very high performance (writing time less than 1ms, low operating voltages, and high endurance). In particular the possibility to create 3D arrays, as shown in fig. 1, attracts much attention.

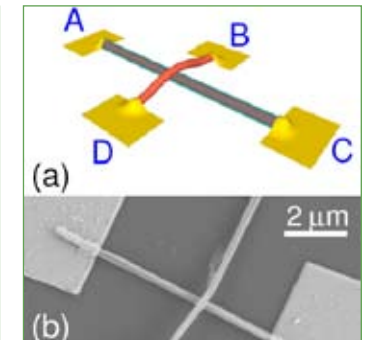
This thesis is entirely focused on the study of RRAM memories based on nickel oxide. The working principle of these devices is based on the so-called resistive switching, discovered in the late sixties. Many metal oxides (as chalcogenides, organic materials, perovskiti and others as well), including binary oxides of transition metals such as NiO or HfO<sub>2</sub> and TiO<sub>2</sub> are able to reversibly change their resistance according to proper polarization. The RRAM memory cell has an extremely simple structure, similar to that of a flat and parallel plates common capacitor: the oxide is sandwiched between two metal electrodes that act as top and bottom contacts. Applying a current ramp across the two contacts, reached a certain threshold voltage, the electrical resistance drops abruptly. This phenomenon is very similar to a breakdown, and is known as forming process. A thin conductive filament (CF) is formed through the oxide shunting the two electrodes, causing the sharp drop of resistance (SET state). From this condition, applying a voltage ramp across the cell, the high current flowing through the CF causes its partial dissolution, restoring the high resistance state (reset state). The low resistance state can be reached again polarizing the cell with a current ramp.

From this stage, the device can switch between high and low resistance (respectively the state of RESET and SET) with the same mechanism. The proper nature of the CF and its structure is still matter of debate, just as the exact mechanism of the resistive switching, however it is possible to find different models in literature that try to explain this phenomenon.

In this thesis two models, for set and reset operations, were developed. The first one is based on evidences that the set transition is enabled by the so-called threshold switching. The latter is a characteristic effect well known to happen in chalcogenide materials, and consists in the field-driven tunneling conduction activation through the trap centers of the material. This activation is induced by the electric field that lowers the potential barrier between adjacent traps. Reached a certain lowering, the electrons flow hopping through the traps, constituting a purely electronic filament. The dissipated power generates a sudden temperature increase, which in turn leads to a stable CF formation. The temperature has a fundamental role in the reset mechanism as well: the current flowing through the CF determines its increase and thus



1. Example of 3D arrays, in principle achievable with RRAM. With this architecture very high memory density could be obtained



2. a) Schematic of a NW-based memory cell, showing also the 4 contacts (A to D) b) SEM micrograph of one device

the diffusion (or oxidation) of the CF restoring the high resistance. Both models accurately reproduce the experimental measurements and provide useful tools for reliability and retention predictions. Based on these models, the thesis develops in detail these critical issues. In particular the study of the intermediate states of resistance has led to further development of modeling and description of the resistive switching, allowing for strong improvements in retention predictions. Accurate temperature measurements demonstrated that the conduction mechanism in the CF slowly drift from being metallic-like, in the set state, to semiconductor-like, in the reset states. This study allowed the evaluation of a critical microscopic parameter, namely the CF effective diameter,

showing how the time-to-fail is related to the initial CF diameter. From this relationship it has been possible to calculate the retention time as a function of the resistance. Statistical measures on a large number of samples have validated this model. Moreover telegraph noise in high resistive states was described and modeled. Its impact on retention, in terms of reading disturb, was assessed.

Along with the characterization and modeling work described so far, an accurate experimental work was performed. It consisted in the design and synthesis of self-assembled core-shell Ni-NiO RRAM memory cells. The nanowires were synthesized by electroplating of nickel in an alumina template and were magnetically aligned and assembled in cross-latch structures (see Fig. 2).

The operating principle of these devices is quite similar to that of planar cells, but the switching oxide consists of the thin NiO layer surrounding the nanowire. The point of contact between two nanowires constitutes the memory cell. With this approach, one can in principle obtain arrays of highly scaled memory, without being constrained by the lithography limits at extremely low costs. An accurate electrical characterization of these devices allowed us to demonstrate for the first time the resistive switching through the thin NiO shell layer surrounding the nanowire. This work constitutes therefore a fundamental demonstration of the feasibility of a self-assembled RRAM device.

## CONFIGURABLE HARDWARE SOLUTIONS TO COMPLEX MOLECULAR BIOLOGY PROBLEMS

**Fabio Cancarè**

Since its birth, the main goal of Bioinformatics has been to retrieve significant biological information by processing raw experimental laboratory data. If until some years ago the main problem that bioinformaticians were facing was to obtain useful raw biological data, in the last few years the main problem has become the management of the huge quantity of the relevant available data and, in particular, their processing in a reasonable amount of time. Many bioinformatics algorithms, in fact, are characterized by a high computational complexity and their execution on a generic workstation can be quickly overwhelmed by an increase in size of the input data. A solution to this problem could be the implementation of the target bioinformatics application exploiting recent digital technologies such as reconfigurable hardware (FPGA), graphical processing units applied to general purpose computing (GPGPU) and multi-core/many-core processors. These devices can, in fact, greatly improve the temporal efficiency of any algorithm characterized by a certain degree of parallelism. If, on one hand, these devices can provide speed-ups of one, two and sometimes three orders of magnitude, on the other hand

they require the modification of the application source code to make it compatible with the chosen hardware architecture. Sometimes altering the source code is equivalent to insert some keywords (pragmas) within the code; most of the times, however, the code must be partially or completely restructured in order to expose the parallelism and/or to make use of a programming language that can be compiled targeting the chosen hardware architecture. Thus, it is worthwhile to think about the tradeoff between having a faster application and the effort of coding it; this is particularly true when the bioinformatician has little or no knowledge about FPGAs, GPGPUs and multi-core/many-core processors. This thesis work analyzes the field of bioinformatics with the goal of understanding the characteristics of the algorithms employed and providing general methods to accelerate them; the analysis focuses on the basic fundamental algorithms, those that have been conceived 20-30 years ago but that are still widely used, as well as on the more recent methodologies able to derive knowledge from high-throughput DNA microarrays. As Moore's law loses validity in the microprocessors field and scientists demand for computational power increases,

domain-specific hardware accelerators will assume a primary role in most scientific researches; this is particularly true in the Bioinformatics area since computational power increases are needed not only for running new applications, but also for running the old ones that involve comparisons against the data stored in publicly available datasets whose size doubles every 12-18 months. The first innovative contribution of this thesis work is a methodology for accelerating bioinformatics applications. The methodology for accelerating a bioinformatics algorithm is composed of: (i) literature analysis; (ii) analysis of the existing algorithm implementation; (iii) design of alternative implementations based either on FPGAs, GPGPUs or many-core processors; (iv) the building of the alternative implementations based on libraries of existing basic components; (v) fine-tuning together with a biology counterparts; (vi) results gathering. The methodology has been applied to several different bioinformatics applications, grouped in two different classes: classical bioinformatics applications and high-throughput DNA microarray applications. For what concerns the former class of algorithms, the thesis

has focused on the BLAST algorithm, the CLUSTAL-W algorithm and the Chou-Fasman algorithm (both the classical and the improved version). For what concerns the latter class of algorithms, the thesis has focused on the Multifactor Dimensionality Reduction (MDR) algorithm, as well as on some variations of the MDR algorithm itself. The results obtained while implementing the classical bioinformatics applications considered are the following: the BLAST algorithm has been implemented using a Xilinx Virtex-4 FPGA and is characterized by an average speed-up of 32.14x with respect to the sequential C implementation. The Clustal-W algorithm has been implemented in CUDA-C using an Nvidia Tesla C1060 GPU; the average speed-up obtained is 30.6x with respect to the sequential C implementation. Finally, the Chou-Fasman and the Chou-Fasman improved algorithm have been implemented in CUDA-C using an Nvidia Tesla C1060 GPU. Both the implementations are characterized by a two-orders of magnitude speed-up with respect to the starting implementations. For what concerns the second class of algorithm, an explanation of what DNA microarrays are is needed. DNA microarrays are devices composed of arrays of microscopic spots called probes that, interacting with biological samples of DNA oligonucleotides, are able to detect a wide range of DNA related information like gene expression and single nucleotide

polymorphisms (SNPs). In the last few years the development of this kind of devices has led to DNA microarrays able to detect hundreds of thousand features at the same time; on the other hand, however, the statistical analysis methods employed to process such data have not kept pace with the technological methods used to generate it, and thus need to be substituted by new computational methods able to both process the new raw data efficiently and also to take into consideration different ways to process them that were not employed in the past due to their computational requirements. Since DNA microarrays are used for several different purposes, it was decided to focus on one specific application involving them: how genetics features and environmental factors determine the susceptibility to complex common multifactorial diseases. In contrast with Mendelian diseases, which are caused by a single genetic variant and which can now be early diagnosed and fought thanks to genetic testing, complex common multifactorial diseases are caused by the interaction between multiple genetic variants and environmental factors; genetic variants contribute to determine disease susceptibility by interacting among them in a non-additive fashion; environmental factors, instead, constitute disease triggers and inhibitors. The actual limit to the exploitation of the available data is mainly imputable to the lack of tools able to automatically seek models and structures of interaction among gene variants and environmental factors that

influence phenotypic expression. To overcome these limitations, research in the field of data mining led to techniques, like the Multifactor Dimensionality Reduction algorithm, able to automatically explore the space of possible non-linear interactions between multiple variables, in total or partial absence of a model of the underlying system behavior, which is the case of several complex biological systems. On the other hand, such techniques are computationally expensive and time consuming to the point of severely limiting their application in the medical genetic field. Thus it was decided to implement the MDR algorithm exploiting all the hardware technologies considered in this thesis work. The first implementation is tailored for many-core processors. It was validated using an Intel 32-core processor and the maximum speed-up obtained was 14.7x with respect to the sequential C-implementation. The second implementation is written in OpenCL and can be run by any OpenCL compliant GPU. It was validated using an Nvidia Tesla C1060 GPU and the average speed-up obtained is 71.2x. The third and last implementation is based on a Xilinx Virtex-5 FPGA and is able to achieve an average speed-up of 60.9x. During the development of the hardware accelerators a components library has been built. It contains a set of hardware cores and software routines that can be re-used while implementing bioinformatics applications.

## AN APPROACH TO DEVELOP SELF-ASSEMBLING SELF-ADAPTIVE SERVICE ORIENTED APPLICATIONS

Luca Cavallaro

Service Oriented Architectures (SOAs) are a flexible set of design principles that promote interoperability among loosely coupled services belonging to multiple business domains. Applications are typically composed out of existing services possibly made available by third party vendors. This opens a series of new scenarios that are unimaginable under the closed world assumption, which mandates that developers know a priori all the components involved in the system and can model and plan their interactions. On the one hand, an organization does not control each part of the application and, hence, failures and service unavailabilities should be taken into account. On the other hand, new interesting services can become available thus enabling new features or providing equivalent services with better quality. Therefore, an interesting challenge in this context is the possibility of building self-adaptive service oriented systems, where loosely coupled component services can be selected at runtime and replaced by others when needed. The possibility of the application to self-adapt by replacing some of its component services opens new scenarios and presents new challenges to developers, making

them relax the traditional closed world hypothesis. This assumption mandates that developers know a priori all the components involved in the system and can model and plan their interactions. The inability to rely on this hypothesis called for some research on service oriented architectures that progressively relaxed this assumption. Literature proposed in recent years many frameworks that allow to design a service oriented application with reconfiguration in mind. However more steps towards open world are still needed. In fact all of the existing frameworks still rely on the hypothesis that a self-adaptive service oriented application development process can still be divided in two main parts: a design time phase, in which a developer analyzes the application requirements, the component services that will be used to fulfill those requirements and the possible reconfigurations of the application, and a runtime phase, in which the foreseen reconfigurations can take place when needed. The hypothesis that the reconfigurations that can happen at runtime are only those foreseen at design time is limiting in the context of self-adaptive service oriented

applications. Since component services used by the application are owned by third parties, there is no guarantee that those services that were selected as possible replacements for a failed one will be still working at the time when they have to be used. Moreover, since new services may become available during the application runtime, it would be desirable to consider them in the reconfiguration process, as they may provide a better quality of service with respect to those already considered at design time. In this work we propose a solution to this limitation based on Service-Tiles, a self-assembly methodology and a runtime framework, exploiting the notion of tile-based systems for an introduction to tile-based systems, which allows a developer to cope with this uncertainty and to build effectively self-adaptive service oriented applications. Tile-based systems define computations as processes that assemble atomic units called tiles. Each tile can be composed only with certain other tiles, according to the symbols they carry. The resulting assembly process is similar to building a jigsaw puzzle with the pieces from a given box.

Working in the same line, Service-Tiles offer a self-assembly

mechanism which, starting from partial information and constraints, that can be specified at design time, builds a system composing available services that respect those constraints. The partial information specifies a subset of services, known at design time, that should be included in the system, while the constraints may specify both functional or non-functional properties the whole application should have. Given these pieces of information (which can be assimilated to tiling systems symbols), a suitable composition is assembled using available services, which may include services discovered at runtime.

The ability of using services discovered at runtime and previously unforeseen by the system designer proposes the additional problem of the lack of standardization of their interfaces. In fact newly discovered services may offer interfaces different from those that the composition designer expected to interact with. To cope with this problem our work features an approach for automatically enabling the dynamic replacement of conversational services by adapting the interfaces a client expects to the interfaces actually available. The approach is based on the synthesis of {lem

mapping scripts}, finite length sequences of instructions, which translate operations that the client is assuming to invoke on the expected service into the corresponding sequences made available by the service that will be actually used. Such mapping scripts are then interpreted by adapters that intercept all service requests issued by the client and transform them into the requests the services are able to fulfill. The aforementioned approaches are coherently integrated into a framework that allows an efficient automatic replacement of services. We show that the resulting solution does not assume, as many other approaches in literature do, a complete knowledge of the application can be available during its development. Instead, it allows to build a system specifying a subset of the needed information at design time and demanding the specification of the rest at runtime, when it is available. Moreover it fully exploits the Service-Tiles possibility of integrating services unforeseen at design time into the application, by overcoming the unrealistic assumption about the standardization of service interfaces, often found in literature, and introducing the possibility

of dynamically generating mapping scripts. We implemented the framework in a prototype that opportunely combines the service assembly generative phase and the mapping script construction, offering a terse model of the composition, which results in a very efficient computation. This efficiency allows us to use the composition model and framework also to enable the application to self-adapt to component services failures, unavailabilities and to changes in the application context recomputing, partially or totally, a new solution when needed.

## ON COORDINATED CUTTING PLANE GENERATION AND MIXED-INTEGER PROGRAMS WITH NONCONVEX 2-NORM CONSTRAINTS

**Stefano Coniglio**

This thesis addresses two fields of mathematical programming: integer linear programming and mixed-integer nonlinear programming. In the first part, we address cutting plane generation, which is among the state-of-the-art techniques for solving integer linear programming problems to optimality. Specifically, we propose a novel cutting plane generation scheme to improve the practical performance of the method. We first give a brief introduction to cutting plane methods, outlining the main ideas on which such techniques rely, and highlighting their practical relevance in modern codes for solving mixed-integer linear programming problems. Then, we review some of the most relevant papers on cutting plane generation, describing which techniques are commonly applied when implementing them. We present the typical approach for cutting plane generation, i.e., that of first generating many cuts that maximize a certain objective function (e.g., cut violation), which are then selected according to different criteria. Among those, we focus on cut diversity. We describe how, in the literature, many authors have observed the benefits that are obtained, in terms of the speed of convergence, when generating a set of diversified cuts.

Motivated by this analysis, we consider two cut selection criteria: cut violation and a measure of diversity between successive cuts, and propose a cutting plane generation scheme that generates a cut that simultaneously meets both criteria. Specifically, we propose a multi-objective separation problem that generates a maximally violated cut which also maximizes a measure of diversity with respect to the cuts that were previously generated. We refer to this new scheme for cutting plane generation as sequential cut coordination since, by generating cuts which directly depend on the sequence of the previous ones, a form of coordination among them is introduced. Different versions of the multi-objective separation problem are considered and investigated. Experiments on two relevant combinatorial problems show that, when comparing our scheme to the classical one, we manage to close a given fraction of the duality gap in a considerably smaller number of rounds, generating fewer cuts, and obtaining a final relaxation which is numerically more stable. In the second part, we address mathematical programs with one or more nonconvex 2-norm constraints. Such problems typically occur in data mining applications that

involve orthogonal distances, i.e., distances from a point to its orthogonal projection onto a hyperplane. We present two relevant data mining problems of such type, namely Exact 2-norm Linear Classification and k-Hyperplane Clustering. We consider mathematical programming formulations for both, and describe the common properties that allow for reformulations where one or more nonconvex 2-norm constraints are involved. We prove that, by defining and solving both problems using different norms, we obtain lower and upper bounds that are within a factor of approximation of the optimal value of the original problem. We then consider spatial branch-and-bound methods to tackle the type of problems under study, starting with an introduction to this type of algorithms in the context of nonlinear optimization. We address the aspects of the standard method that are critical with respect to our problems. Then, we discuss the application of different convex relaxations and branching rules from the literature, highlighting their drawbacks, and propose a novel spatial branch-and-bound method that is based on iteratively enriching a disjunctive programming formulation which approximates the nonconvex

feasible region by approximating its complement with an iteratively enlarged polyhedron. The efficiency of our method is assessed via computational experiments on the Exact 2-norm Linear Classification problem. We conclude the work by proposing an efficient metaheuristic for the k-Hyperplane Clustering problem to tackle large-size instances. The algorithm is based on iteratively trying to improve on a current solution by reassigning a set of promising points to different clusters, relying on a criterion for identifying a set of good candidates. The number of such points is adaptively decreased or increased along the iterations, based on the quality of the current solution, thus allowing to escape from local minima, also applying two features of Tabu Search. Computational experiments on randomly-generated medium and large-size instances confirm the efficiency of the algorithm. Experiments on real-world data mining and image processing problems are also reported and illustrated in detail.

# GUIDANCE AND CONTROL FOR PLANETARY LANDING

**Delia Desiderio**

General architectures for autonomous entry, descent and landing of a probe on another planet usually consist of three main components, namely a guidance law, a navigation system and a trajectory tracking controller. Guidance is the process of deciding how to steer to the desired target. Navigation is the process of determining the vehicle's position, velocity, and attitude in space. Control is the process of implementing the guidance commands to achieve changes in thrust vector. In the first part of the thesis a particular guidance navigation and control (GNC) system is considered: the one used in ExoMars, a future ESA mission for the exploration of Mars. The study and the analysis of the terminal descent phase of that mission have been realized in the framework of a collaboration project with Thales Alenia Space Italy. In the second part of the thesis the attention is turned to some aspects of the GNC, in particular to the guidance and control problems associated with the descent dynamics of a vehicle during the landing phase. The tradeoff between analytical and numerical solutions to the guidance problem has been discussed exploiting the flatness property of the system. Moreover a flatness based approach is proposed also for the design

of the trajectory tracking controller. In the third part of the thesis the analysis of the static and dynamic performance is considered. To this aim the Bode's sensitivity integral and an its analogous for time-varying systems have been considered.

## 1.1 Aim and results of the thesis

The research work described in this thesis deals with some aspects of the Guidance, Navigation and Control (GNC) problem, that is the design and implementation of a suitable guidance law, the development of a navigation system and the design of a trajectory tracking control law. The GNC problem plays a leading role in the framework of interplanetary missions for the entry in atmosphere, the descent and the landing of a probe on another planet. In fact the integrity of the payload depends on the GNC system, and hence the outcome of the mission too. The main problems concerning the GNC systems are: the definition of a guidance law that takes into account the mission constraints but that is easily to implement on board with a low computational cost; the design of control laws that allow the tracking of the trajectories returned by the guidance system in non stationary conditions; the design of a dedicated

filter providing the position and attitude in space of the vehicle, useful for the controller operations (navigation system). The aim of the thesis, after the analysis of the state of art and the study of the current methodologies, is the development of new tools in order to improve the existing results in the GNC framework. To this aim three principal work steps has been done:

1. The study of the GNC system of an ESA mission (ExoMars) considering, in particular, the landing phase.
2. The development of guidance and control laws, considering the descent dynamics of a vehicle during the landing phase.
3. The analysis of the dynamic performance and the tradeoff between the static and the dynamic performance of the closed-loop system.

To sum up, for each work step, the following results have been obtained:

1. The nominal and robust analysis of the descent and attitude loops of the vehicle in ExoMars.
2. The design of suitable guidance laws, the tradeoff between the analytical and the numerical solutions to the guidance law and the design of trajectory tracking control laws.
3. The computation of a new

index for the analysis of the dynamic performance, starting from an analogue version for time-varying system of the Bode's sensitivity integral.

## 1.2 Outline of the thesis

The three principal issues, before described, form the three parts of this thesis.

Hence in the first part the ExoMars mission is considered as a case study of a GNC system, in the second part the design of the guidance and control system for the landing phase are provided and in the third part the performance analysis is described. An overview of the contents of the chapters is given below.

### Part I: ExoMars.

- In Chapter 2, after a literature review of the GNC problems, the ExoMars mission and the terminal descent phase are described. In particular the on board instrumentation and the GNC architecture are presented and some simulation results are reported.
- In Chapter 3 the analysis of the terminal descent phase of ExoMars is realized. First of all, the nominal analysis, considering both the attitude and the control loop, is carried out. Secondly the robustness analysis of the both control loops is realized by means of Linear Fractional Transformations and  $\mu$ -analysis. Finally the worst case analysis is provided thanks to the differential evolution method.

### Part II: Guidance and control for planetary landing.

- In Chapter 4 the mathematical model for the descent dynamics of the vehicle during the landing phase is computed. In order to facilitate the solution of the guidance and control

problem the flatness property is introduced and applied to the descent dynamics. Finally the mathematical model of the tracking error system is computed.

- In Chapter 5 the guidance problem is considered. After a literature review, an analytical solution to the guidance problem is computed. Numerical schemes with pseudo-spectral and polynomial approximations, that involve the minimization of a suitable cost function as index of static performance, are also computed. Simulation results show first the comparison of the numerical methods and then the comparison between the analytical and the polynomial numerical solutions. The aim is to find conditions in which the analytical solution imply static performance comparable to that obtained with the numerical methods. This is an important issue because the analytical solution is very simple in term of implementation and can allow a possible re-targeting of the landing site on line.
- In Chapter 6, after a literature review about the two degree of freedom controllers, the two designed trajectory tracking controllers are presented. The nonlinear one consists of a feedforward part based on differential flatness and a linear feedback part of extended PID type. The linear one is designed on the linearized tracking error dynamics and then applied to the nonlinear system. The simulation results are realized in different conditions in order to appreciate the difference of the two controllers and, in particular, the robustness of the nonlinear one.

### Part III: Performance analysis.

- In Chapter 7 a literature review about the analysis of the performance limitations by means of the Bode's sensitivity integral is provided. The principal theorems about this topic and suitable examples are then reported. Moreover the problem of find a tradeoff between the dynamic performance, computed thanks to the Bode's sensitivity integral, and the static performance, expressed, for example, by an optimal cost obtained by the minimization of a suitable cost function, is discussed with other appropriated examples.
- In Chapter 8, after a brief literature review about a possible extension of the Bode's sensitivity integral for time-varying systems, the stability properties and the notion of relative degree for time-varying systems are provided. In particular the linear time-varying systems are considered and the basic definitions and some mentions to the Lyapunov exponents theory are reported. An analogue of Bode's sensitivity integral for time-varying systems, present in the literature, is recalled and a new result on finite horizon is shown. Finally, considering the guidance and control system designed in Part II, the analysis of the static and dynamic performance is provided.
- In Chapter 9, finally, some concluding remarks on the work performed are reported and some possible future developments are provided.

## SYNCHRONIZATION AND DATA DISTRIBUTION OPTIMIZATION FOR DISTRIBUTED SHARED MEMORY MULTIPROCESSORS

**Andrea Di Biagio**

In the last decade, the trend in processor design moved from complex single processor systems to chip multiprocessors (CMP) exposing a high degree of hardware parallelism. Starting from the early 2000's, the costs involved in the design and fabrication of new generations of complex uniprocessors eventually became too high. Exploiting the intrinsic instruction level parallelism (ILP) required more and more complex out of order logic to be implemented on chip, thus leading to an unacceptable increase in the number of transistors and consequently to a dramatic increase in power consumption. Eventually, processor vendors moved from solutions based on complex and power-costly out-of-order single processor systems, to solutions based on chip multiprocessors implementing multiple levels of shared caches thus giving birth to a new trend in hardware design.

Recent trend in highly parallel architectures has been towards a class of machines known as Non-Uniform Memory Access (NUMA). A NUMA machine is implementing more than one system bus, each serving a group of processors called node. In such architectures, each processor has faster access to local memory (i.e., memory

associated to its node) than to remote memories. Since access times depends on the memory location relative to a processor, the performance of parallel programs strongly depends on the amount of bus traffic due to remote accesses.

On the other hand, the increased programmability and hardware parallelism of today's modern general purpose graphic processing units (GPGPUs) eventually lead to a raise in the interest for this class of multiprocessor architectures as low-cost accelerators for a wide range of high-performance applications. Modern graphic devices are able to support computational intensive data parallel applications traditionally handled by CPUs or dedicated coprocessors. A general purpose GPU consists of a cluster of independent multicore processors each implementing a fast explicitly managed local memory (EMM). In order to achieve good performances programmers have to explicitly manage data locality, choosing whether to place data in local memories or leave them on a global memory which is usually slower but shared among all multiprocessors in the system.

In recent years, languages like Fortress, X10 and High Performance Fortran have

been proposed to improve the programmability of NUMA machines. On the other hand, languages like Compute Unified Device Architecture (CUDA) and OpenCL have been proposed by the industry to support the programmability of modern general purpose GPUs. All these languages provide a set of high-level functionalities to both guide the runtime dissemination of threads over a set of available processing elements and specify the memory distribution for array data structures.

However writing parallel programs for NUMA machines and general purpose GPUs is in general more complex than writing parallel code for generic shared memory multiprocessor since they fail at hiding the complexity of the underlying multiprocessor systems exposing too many hardware details to the programmer. In the case of NUMA machines, writing efficient parallel programs requires that programmers explicitly manage the physical distribution of memory in the underlying system. A similar problem arise in the case of general purpose GPUs in which programmer has to explicitly manage multiple address space and complex memory hierarchies.

The main goal of this dissertation is to improve the programmability of both general purpose GPUs and NUMA machines simplifying the complexity of writing parallel programs for such architectures. Rather than focusing on one of the aforementioned languages which in general expose too many hardware details to the programmer, this dissertation focuses on OpenMP which is a simple and well known parallel programming standard for shared memory multiprocessor systems. In addition, OpenMP has been recently proved to be a viable solution to program NUMA machines when suitable extensions are employed. Hence, the dissertation provides the following contributions towards the proposed goal: An OpenMP to CUDA translation scheme to improve the programmability of general purpose GPUs; A technique to improve the performance of synchronizations due to data aggregation in parallel programs; A technique to improve memory locality in OpenMP applications that benefit NUMA machines.

Moreover, this dissertation provides an experimental study of the programmability of general purpose GPU architectures by means of application scenarios from the cryptographic field.



## SELF-ORGANIZING METHODS AND MODELS FOR SOFTWARE DEVELOPMENT

Daniele Joseph Dubois

The current research trends in Software Engineering are focusing on the development of new techniques to deal intelligently and efficiently with the design of systems that are able to evolve overtime and adapt to rapid changes of their requirements. Typical solutions that are being investigated are systems that are able to self-adapt at run-time to unpredictable changes in their execution context, unexpected behavior, changes in the application requirements, and so on. These operations have been traditionally done manually, but, due to the increasing complexity of modern systems, more efficient and automated alternatives must be necessarily investigated. In this context, considerable research efforts have been already done to transfer to complex software systems some characteristics of natural systems. An example of natural system of inspiration is the Autonomic Nervous System: in a human body there are many processes that a person is unaware of and they are essential to provide a constant maintenance of the body. Another example is the organization of insect colonies such as the search of the optimal path from the anthill to the food source in ant colonies, or the blinking

synchronization phenomenon in fireflies colonies. These ideas have been then transposed as a new computing paradigm called *Autonomic Computing* in which the maintenance and the decisions of a generic system occur spontaneously, with no or minimal manual intervention, thus making possible to deal with situations characterized by a high intrinsic complexity. According to the definition of Autonomic Computing proposed by IBM all autonomic systems should exhibit the so-called self-\* properties, that are: *self-configuration*, which represents the capability of the system to re-configure itself if necessary; *self-optimization*, which represents the capability of the system to modify its behaviors in order to improve the efficiency in reaching its goals; *self-healing*, which represents the capability of the system to automatically detect a malfunction and to possibly fix it; *self-protection*, which represents the capability of the system to detect security violations and to possibly block them. To achieve these properties several research lines have been developed, each one focusing on different aspects that are required to engineer these systems. The two basic aspects we have identified are the following. First, an architectural framework that may be used to

give autonomic capabilities to existing applications or that is able to support the development and the deployment of new autonomic applications regardless their behavior. Second, the autonomic logic that instruments the application: it can be an algorithm or a policy in terms of rules that may trigger system-level changes and represents the autonomic behavior of the system. Adaptation in autonomic systems may be obtained using two different approaches. The first one is the *top-down self-adaptation* in which an additional supervision layer is added to the system. This way the system is conceptually separated into a *monitor* (supervision layer) and a *managed resource* (the former system) organized as a control loop: the monitor analyzes the state of the resource using sensors, and, according to a policy that depends on the application, performs corrective actions using an actuator. The second approach is *bottom-up self-adaptation*, often called *self-organization*, in which the adaptation is obtained through interactions of system elements, without adding any additional layers to the system. With this approach system elements cooperate at the same level using only local information and inter-element communication.

The main difference between these two approaches is that in top-down adaptation the system elements have well-defined roles (monitor and resources) at different levels, meaning that they have access to different knowledge and have different capabilities; while in the bottom-up adaptation each element may behave both as a monitor and as a resource, but at the same level, meaning that different elements have the same importance and the same type of access to the environment and to the system knowledge. In this thesis we will focus on algorithmic and architectural aspects of autonomic systems that adapt themselves using self-organizing techniques, where self-organization is defined as “*An increase of order which is not imposed by an external agent (not excluding environmental interactions)*”. A characteristic of self-organization that is interesting for software systems is that the entities participating in this activity usually make decisions and perform actions autonomously, based on local knowledge, without an explicit coordination. The re-organization of the whole system *emerges* from these local actions, thus showing a scalability that holds the promise of significant progresses in the way software systems could be designed and operated. For this reason we assume to deal with software architectures that have the following characteristics: *distributed*: composed of different interacting elements; *high number of elements at the same hierarchical level*:

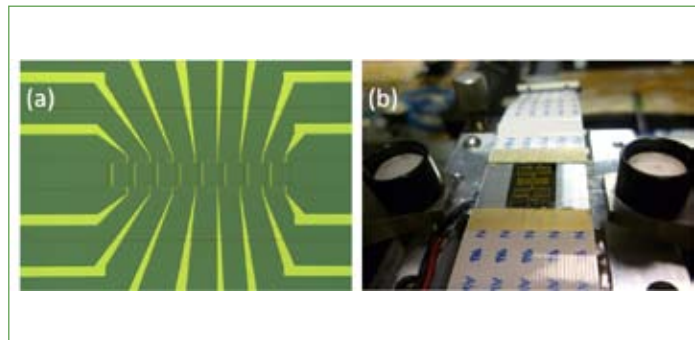
the system should have at least one level which is composed of hundreds/thousands of elements; *high uncertainty and dynamism*: the system is heavily perturbed due to internal/external changes that are difficult to predict; *no elements alone are essential*: the system keeps working if a small fraction of random elements at the same level is removed. With respect to self-organization our first research hypothesis is to demonstrate that self-organization is able to provide good solutions to selected problems in computer science under the architectural assumptions stated above. In particular, we want to show that they inherit the advantages of self-organization that are observed in the natural world. The second hypothesis of this work is to demonstrate that in selected cases the self-organization design guidelines we propose simplify the adoption of self-organizing approaches in the development process of real systems. These design guidelines consist of a software engineering approach to use self-organization in the development of real systems to identify typical commonalities of self-organizing algorithms and a way to map them into existing autonomic systems. In fact, right now most of self-organizing algorithms are described in a generic way that does not point out at their self-organizing essence; moreover most descriptions tend to be fairly generic, meaning that a system designer has to face with implementation problems that are not captured by the original algorithmic description.

Possible examples include synchronization issues, race conditions, interaction timings, and so on. The final contribution that this work shows is the design, development, and simulation of several self-organization algorithms that may be used to solve several problems in autonomic systems such as load-balancing problems, traffic optimization in publish-subscribe systems, and energy optimization in data centers. Moreover, an analysis of their constituent self-organization aspects in the form of self-organizing principles will be proposed. These principles will be used as building blocks to model existing self-organization algorithms as well as for supporting the design of new ones. The most important principles will then be mapped to some design patterns that will be used to help bring such algorithms into real software architectures. This has the advantage of easing the design and development of self-organizing approaches since different instances of the same pattern may share the same code and contain only information of a single self-organizing aspect of the whole application. Finally, the proposed contribution is completed by a presentation of some guidelines to support the modeling, validation, and implementation phases of the development process.

## “VLSI PHOTONIC DEVICES FOR ALL-OPTICAL SIGNAL PROCESSING”

Carlo Ferrari

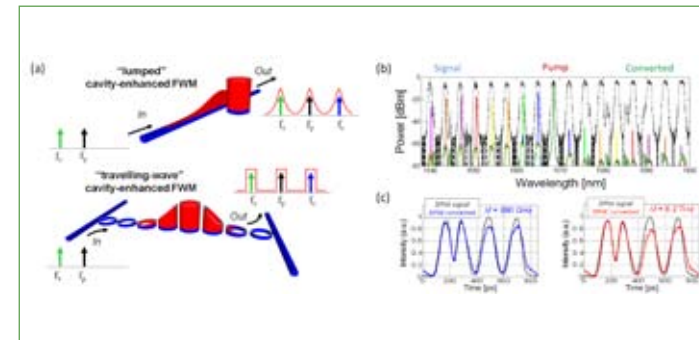
This doctoral dissertation concerns the identification and analysis of photonic integrated circuit architectures for optical signal processing devices in telecommunication and interconnect networks. The development of optical components capable of processing information in the optical domain is strategic to overcome the energy and interface capacity bottlenecks, which are increasingly constraining the performance scaling of electronics. To this aim, photonic devices with very large scale of integration (VLSI) are necessary to provide the link capacity and density required by optical interconnection, to aggregate different functionalities within the same device and to ensure an efficient physical and logical integration with electronics. The purpose of this work is to spot innovative integrated signal processing devices and single out key signal processing tasks in which photonics can really assist electronic processing. Moreover, efforts are directed towards demonstrating that novel functionalities and applications can be introduced in conventional devices, without modifying their architecture, to enrich the range of processing tasks that can be achieved by the same component. Integrated devices for all-optical buffering,



**1. Integrated ring resonator optical waveguides for all-optical buffering and wavelength conversion. (a) Top-view microscope photograph of a fabricated silicon coupled resonator optical waveguide with embedded thermal control; the footprint of each ring is  $40 \times 56 \mu\text{m}^2$ . (b) Photograph of a silicon optical chip on the mounting fixture used for characterization.**

optical signal equalization and nonlinear signal processing are presented and discussed, with particular focus on slow-light devices, which hold great potential for time-domain processing of signals and for enhancing the efficiency of nonlinear interactions. The development of optical buffers capable of temporarily storing and adjusting the timing of data streams or packets for flexible and transparent switching nodes, optical packet routers and clock distribution in optical interconnects is the first topic addressed in the dissertation. Solutions for optical buffering can become of real interest only if large delays are associated with wide bandwidth, full reconfigurability, low loss and tolerable signal distortion. High storage density, integration

and compliance with electronics are also key requirements for efficient buffering. In this perspective, a novel integrated slow-light delay line architecture based on directly coupled ring resonators is demonstrated. It is shown that coupled resonator optical waveguide (CROW) delay lines offer continuous and hitless control on the delay of a modulated data stream by exploiting a gating mechanism, whose activation can be achieved by an electrical control circuit embedded on chip. Different fabrication technologies for optical buffers, namely medium index contrast glass and silicon, are demonstrated to be compatible with a wide range of applications and signal bit-rates. The experimental demonstration of variable optical buffers with



**2. Enhanced wavelength conversion in coupled resonator optical waveguides. (a) Schematic of signal field distribution in resonant wavelength conversion in a single resonator and in a chain of ring resonators. (b) Spectrum of the output optical field after wavelength conversion over 7 THz. (c) Measured time traces of a converted phase-modulated data pattern at 10 Gbit/s.**

storage capacity exceeding 1 byte, working at bit-rates up to 100 Gbit/s and integrated on a silicon wafer is reported, representing the state of the art of integrated optical buffers with continuous control of the delay. Advanced applications for CROW optical buffers are also demonstrated, showing the compatibility of CROWs with multilevel phase-modulated formats and their suitability to continuously control the relative delay between two orthogonally-polarized channels of polarization-division multiplexing systems. The need of optical packet switching networks for optical buffers with fast activation and large storage capacity is also addressed, by exploring the option of trap-light techniques in fast-activated silicon CROWs as a viable

strategy to overcome the speed and storage capacity limitations. Concerning optical signal reshaping and equalization applications, it is shown that ring resonators can be employed as signal reshaping building blocks in photonic integrated circuits without adding complexity to devices. The phase response of ring resonators is exploited to add novel functionalities to a ring-based wavelength router, in order to compensate for propagation impairments and restore signal integrity. It is also shown that ring resonators can be efficiently employed to remarkably extend the transmission reach of directly modulated lasers, by improving the quality of the modulated output signal. The last part of the work is

dedicated to demonstrating the enhancing capabilities of CROW structures in nonlinear optical processing applications for wavelength conversion and signal regeneration. The novel concept of traveling-wave resonance-enhanced wavelength conversion is introduced to overcome the bandwidth limitations of the resonance-enhanced four-wave mixing (FWM) process in single cavities. Efficiency and bandwidth improvements as well as greater robustness against losses with respect to single cavities and bare non-resonant waveguides are experimentally demonstrated by using silicon as a nonlinear platform. The circuit concept of traveling-wave wavelength conversion is very general and compatible with a wide range of technological platforms. It can, therefore, be applied as a complement to efficient waveguide and material designs to boost further the nonlinear processing performance of photonic devices.

## MODELING OF CONDUCTION AND RELIABILITY CHARACTERISTICS FOR PHASE CHANGE NON-VOLATILE MEMORIES

Davide Fugazza

Exploiting the ever increasing market demand for portable consumer products requiring permanent and high-density data storage, semiconductor non-volatile memories (NVMs) have gained in the last decades an explosive success. Music players, digital cameras, USB drives, cell phones and the emerging Solid State Disks (SSDs) are only a few examples of the ubiquitousness and the ever increasing part got by non-volatile memory in reshaping our lifestyle. In this play the leading role has been performed by Flash memory, that from a simple concept in the early 80's grew up and generated close to \$23 billion in worldwide revenue in 2007, undoubtedly representing the actual mainstream memory on the market. This enormous success was essentially driven by Moore's Law, that lead to dramatic reductions in unit cost and to the creation of new fruitful markets. The ever increasing demand for more memory bits largely repaid the continuous efforts devoted for manufacturing memory chips with increased performance and functionality, resulting in a sort of virtuous circle. Thus, despite their higher cost per bit with respect to magnetic hard disk drives, semiconductor memories resulted the winning solution

in all the consumer products requiring light weight, low size, low power consumption and high reliability. However, a further increase in storage capacities with a simultaneous cost per bit reduction is now mined by the physical and technological constraints of conventional memory technology and will thus require for next technology nodes something more than the scaling of feature size. As a result, to provide a better trade-off between scalability and reliability as well as the basis for the evolution of the actual storage hierarchy in its whole, new memory concepts have been recently object of intense investigation. More interestingly, most of these solutions have renewed the attention to certain classes of materials whose nature is far from the regular perfection that had permeated as a *condicio sine qua non* the outstanding evolution of microelectronics so far. Chalcogenides alloys, binary/ternary oxides, as well as other amorphous semiconductor materials exploited in these new kind of memory devices, can in fact be cited as representative examples of the materials and design revolution felt by the semiconductor industry in its whole. In this frame, among the proposed emerging concepts, chalcogenide-based

phase-change memory holds a privileged position due to the good degree of technology maturity, supported by its promising scaling potential and a broad application range. PCM working principle is essentially different from conventional floating gate devices, where the information is associated to an amount of stored charge: It relies on a change of the resistance in the active material, hence its classification as a resistive memory. Such a functionality is obtained by smartly exploiting the peculiarities of particular materials, usually chalcogenide alloys (e.g.  $\text{Ge}_2\text{Sb}_2\text{Te}_5$ ), that can be reversibly switched between crystalline and amorphous phase by electrically induced phase transitions. Given the resistive type of PCM, the transport phenomenology contributes in a fundamental way to determine the technology performance. The exploration and modeling of conduction plays therefore a central role, allowing to accurately predict the read-window budget between set and reset states and to evaluate device functionality as a function of different programming and operating conditions – such as temperature, time, voltage and current –, providing support for the technology development. The Ph.D Dissertation will

address the conduction properties and the reliability characteristics of PCM devices by both experimental and numerical investigations. New interesting phenomena for active volumes of decananometric sizes will be reported, including: (i) an anomalous dependence of the resistance on the thickness of the programmed amorphous region, (ii) current localization effects related to the non-Arrhenius behavior of the resistance in temperature, (iii) the occurrence of Lorentzian components (Random Telegraph-signal Noise - RTN) in the current spectra and (iv) experimental evidence of how drift and noise phenomena are both driven by temporal variations of the energy for hopping within the material. In order to address these phenomena, a new physics-based framework for the conduction in amorphous phase, namely the distributed Poole-Frenkel (DPF) model, will be introduced. Transport will be described in terms of Poole-Frenkel (PF) conduction through localized states with distributed energy barriers for hopping, leading to a more general form of conduction with respect to previously reported models for PCM devices. Also relying on previous works on amorphous solids and percolation in disordered

materials, the conduction process will be reformulated and numerically implemented in terms of an equivalent Random Resistance Network (RRN) with an exponentially wide spectrum of local resistances. Through the inclusion of the physical insights coming from the observed experimental dependences, the proposed approach will be shown to allow for a quantitative description of the temperature and time dependence of the programmed state, therefore contributing to a significant progress in the understanding of its properties. Finally, the model will be adopted to address some reliability scaling projections in terms of the resistance window for future technology nodes. These reliability investigations represent the key for the qualification of phase-change memory for both single bit and multilevel applications.

## DEALING WITH COMPLEXITY AND DIMENSIONALITY IN WATER RESOURCES MANAGEMENT

Stefano Galelli

Advances in scientific computation and data collection techniques have increased the level of fundamental understanding that can be built into Process-Based (PB) models, which are widely adopted to describe the dynamics of large environmental systems. These models are definitely useful to enhance our scientific knowledge of natural processes, but their structure, which becomes progressively more complicated as the complexity of the systems being studied, exposes them to a number of critiques. Different authors argue that PB models are difficult to communicate, their structure hard to judge by other researchers and their output data not trivial to validate. Over-parameterization is probably the weakest point of PB modelling, as the ratio between the data information content and the number of parameters is often such that the estimates of these latter are highly uncertain. As far as PB models practical implementation is concerned, it must be considered that their software coding is usually complex, and their computational requirements, in terms of CPU, memory and disks usage, extremely high. This is still the main limitation preventing PB models to be applied to problems requiring hundreds or thousands of model runs,

such as decision-making (e.g. planning and management), Monte-Carlo analysis, data assimilation or scenario analysis.

A common and effective solution to overcome this limitation is to perform a top-down reduction of an existing PB model, which can be replaced by a simpler, computationally efficient emulation model that mimics the dynamic behaviour of the underlying environmental system under the effect of external drivers. Literature shows a variety of dynamic emulation modelling techniques, but actually two main approaches can be discerned: a structure-driven and a data-driven approach. On its turn, the structure-driven approach comprises two methodological groups. Techniques belonging to the first group are based on a detailed analysis of the PB model structure and its subsequent heuristic simplification by removing variables or feedback mechanisms from the PB model. The usage of these techniques is limited to domain experts, as they require a certain amount of user interaction. Model reduction techniques of the second group reduce the PB model dimensionality by projecting the original systems of state transition equations into a lower-dimension sub-space, where the equations are solved for the substituted projected

states. There are two sets of projection methods, which are based on Singular Value Decomposition and moment matching. Both methods are however available only for linear and weakly non-linear systems, while theory is still under development for non-linear systems. Techniques belonging to the data-driven approach aim at identifying a low order emulation model by processing, in a data-driven fashion, a data-set appropriately generated from the original PB model. These techniques, which require an accurate generation of this identification data-set, can be easily applied to both linear and non-linear models, as they do not require any analytical assumption about the PB model structure. This is the most promising approach for the reduction of PB models in environmental modelling, where systems are typically complex and highly non-linear. However, a shared theoretical vision is still missing and different techniques were independently developed in different domain of interest.

The purpose of this work is fourfold: create a framework concerning the different techniques developed for model complexity reduction, with particular regards to environmental applications;

propose a procedural approach for the identification of emulation models in a data-driven fashion; present some novel algorithms designed for emulation modelling tasks; evaluate the effectiveness of the proposed procedure and algorithms on two real-world case studies.

The first chapter of the thesis is mainly concerned with the first purpose; in particular, it describes the families of PB models and the corresponding problems on which they are employed, it proposes a taxonomy for the different emulation modelling strategies and approaches, and it discusses, in relation to this taxonomy, the methods that have been used in the last years. Chapter 2 proposes a procedure for data-driven emulation modelling. This procedure is composed by seven major phases: *i)* Problem conceptualization - concerning the analysis of the PB model and the selection of the output variable to be reproduced; *ii)* Design of Experiments - the design of a sequence of simulation experiments to be performed with the PB model in order to obtain a data-set for the emulation model identification; *iii)* Simulation Runs - namely the generation, via simulation, of the above mentioned data-set; *iv)* Lumping - the aggregation

and projection of the PB model variables into a lower-dimension sub-space; *v)* Reduction - the selection, among the variables aggregated in the previous phase, of the most relevant and significant variables to the emulation model output; *vi)* Emulation Model Identification, the selection of a model class that is subsequently cross-validated; *vii)* Emulation model usage - the employment of the emulation model in the problem that could not be solved with the original PB model. As the selection of the most relevant variables is commonly difficult, because of their large number and non-linear relationships, Chapter 3 proposes two novel algorithms, named Recursive Feature Selection and Iterative Feature Ranking, which drive the selection of the most significant variables.

The second part of this thesis finally concerns the application of the proposed procedure to two real-world case studies. Chapter 4 describes the reduction of a 3D, physically-based model (ELCOM-CAEDYM) describing the hydrodynamic and ecological conditions of Googong reservoir (Australia). The scope of this first application is to reduce the dimensionality of ELCOM-CAEDYM, so that the emulation model can be used to design, via the resolution of a management problem, the

control policy for the mixers being employed to reduce the water quality problems affecting the reservoir. Chapter 5 describes the reduction of a large conceptual model, which provides the irrigation water demand of the Muzza-Bassa Lodigiana district (Italy), fed by the Lake Como releases. The purpose of this second application is still to reduce the PB model dimensionality, i.e. to replace it with an emulation model that can be subsequently employed to design, via stochastic dynamic programming, the Lake Como release policy.

## ANALOG REPEATING SYSTEMS FOR MULTI-USER SCENARIOS

**Jonathan Gambini**

The integration of multi-hop capability into conventional cellular networks is envisioned as one of the most promising solutions for guaranteeing almost-ubiquitous very high data rate coverage.

When one source communicates towards a destination with the help of a third party node acting as a relay, the regenerative decode-and-forward (DF) strategy is well known to be the optimal one. In this case, the coding/decoding policy is separately optimized over both the interfaces of each link. However, if multiple two-hop links coexist within the same resource, this strategy results into a passive approach of dealing with interference, as each relay attempts decoding of the intended signal by treating the contribution arising from other concurrent transmissions as noise.

On the other hand, the lesson we learn from multicell processing paradigm is that a more proactive treatment of interference arising in multi-link scenarios, which can be accomplished through some form of interference-aware multi-cell coordination at the receiving side, allows to enhance the overall system performance by mimicking the benefits of MIMO processing. In a multi-hop scenario, the employment of memoryless

amplify-and forward (AF) devices (i.e., analog repeaters) for relaying configures as the most viable option to exploit the multicell processing paradigm, as they enable transparent transferring of the interferential contribution accumulated over the different hops to destination for promoting a joint processing of all the signals. The thesis is concerned with the study of communication schemes that enable to trade the DF protocol for the benefits of a centralized processing through the adoption of AF relays that transfer to destination the perceived multi-user interferential contribution without elaborating it in any way. Main purpose of the thesis is to prove that replacing smart regenerative repeaters with simple bidirectional analog repeaters in practical two-hop communication scenarios entails large margins for maximizing the number of co-channel links that can coexist while still guaranteeing acceptable performances. Femtocells represent a new paradigm in wireless communications based on the concept of deploying low-power base stations (also known as femtocell access points – FAPs) capable of providing wireless voice and data services (such as UMTS, WiMAX and HSPA/LTE) to short-range indoor environments, and linked to the

operator's core network through a broadband connection, such as xDSL. Interference management is regarded as one of the most crucial challenges related to the femtocell paradigm, as it would be based on a distributed approach where each FAP selfishly selects its transmit parameters (in terms of radiated power and frequency allocation) on the basis of local measurements of the surrounding – highly varying – radio environment.

In counter-trend, the conventional femtocell architecture is revisited by proposing a distributed antennas system (DAS) for the benefits of multicell processing in terms of centralized interference mitigation and radio resource management. The wireless over cable (WoC) paradigm is a new technique to enable network configurations employing the DAS principle. WoC is based on the relay of radio frequency wireless signals over a cable infrastructure through a bidirectional (analog) amplify and forward device that translates the bandwidth of wireless signals to comply with the specifications of the wired links (and vice-versa). Following the WoC concept, the FAP employed in conventional femtocell systems is replaced by connecting the antenna with

the available cable infrastructure after up/down-frequency conversion provided by a totally-analog home device (namely, analog-to-analog – A/A – converter). Baseband processing and radio resource management are carried out simultaneously at a remote control unit (namely, Multi-cell Base Station – McBS) for multiple femtocells thus mimicking the same advantages as MIMO systems in terms of centralized processing of the multiple interfering signals, rather than treating each femtocell as a stand-alone entity. Moreover, macro-diversity and/or multiplexing gains are achieved without the need of any signalling.

In the uplink, the proposed Wireless over Cable architecture for femtocells (termed FemtoWoC architecture) literally exploits inter-cell interference by allowing the user data to be jointly processed by chaining wireless and cable multi-link channels, so as to mimic the benefits of a large virtual MIMO array. In the downlink, where uncoordinated and not cooperative end users prevent the receivers to perform any joint interference cancellation algorithms, the existence of a centralized agent (i.e., the McBS) that provides termination to the different A/A converters is still employed for leveraging the high degree of flexibility provided by multi-carrier transmissions by defining a power allocation strategy and a scheduling algorithm that capitalizes on the spatial diversity provided by realistic femtocell scenarios.

The FemtoWoC architecture is presented as a green solution that prevents the need of xDSL

equipment for connections over cable, and decreases the effective amount of energy expenditure per delivered information bit. Benefits of the FemtoWoC solution in terms of energy efficiency with respect to the conventional scheme are investigated through simulations by accounting for a realistic LTE cellular deployment.

A FemtoWoC environment characterized by a multi-radio roaming subscriber and two A/A converters connected to different McBS's that provide the radio connectivity over disjoint spectra is also taken into account. In this context, an optimized packet-wise vertical handover strategy for a multi-radio node that has the capability to switch between two uplink orthogonal radio interfaces for communication is investigated. The problem is formulated as the joint optimization of power allocation (at the physical layer) and probability of vertical handover (at the MAC layer), based on long-term channel state information and in the presence of time-varying fading channels, random packet arrivals and random activity of the network users.

In the context of relay-assisted wireless networks, the concept of multi-user interference propagation through analog repeating is still exploited for enhancing the system performance. Specifically, we focus on a multi-user scenario where a group of nodes are interested in mutually communicating and need to be assisted by a single amplify-and-forward relay as they are out of each other's transmission range (i.e., there is no direct-

link). Traditionally, mutual interference is avoided by time-sharing the relay among the different transmitting nodes, thus resulting in a highly spectrally inefficient scheme. In countertrend, we propose to avoid the throughput loss deriving from the interference avoidance paradigm by letting the users simultaneously employ the radio resource for concurrent transmissions towards the relay node. Under the assumption of static or slowly-varying channels, the capability of the relay node to simply analogously repeating the received interfering signals is exploited to define a communication protocol based on slotted transmissions that is tailored to let the channel estimation leverage the persistency of the training sequences across multiple frames. Benefits of the proposed protocol are stressed through comparisons with the traditional TDMA-based routing scheme where the relay is time-shared among the different users. Impacts of the hindrances to the proposed scheme (namely, hardware impairments at the nodes and finite coherence-time of the channels) are also investigated by numerical analysis. The proposed algorithm is envisioned to enable the employment of simple sense-before-talk policies at each user interested in accessing the network, thus avoiding complex forms of centralized coordination by the relay.

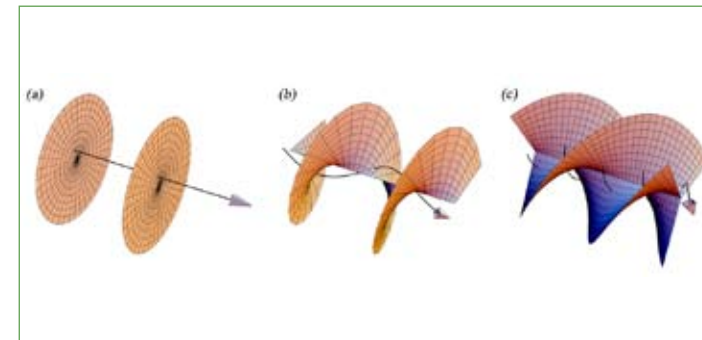
# ORBITAL ANGULAR MOMENTUM DIVISION MULTIPLEXING IN OPTICAL COMMUNICATION SYSTEMS

Alberto Gatto

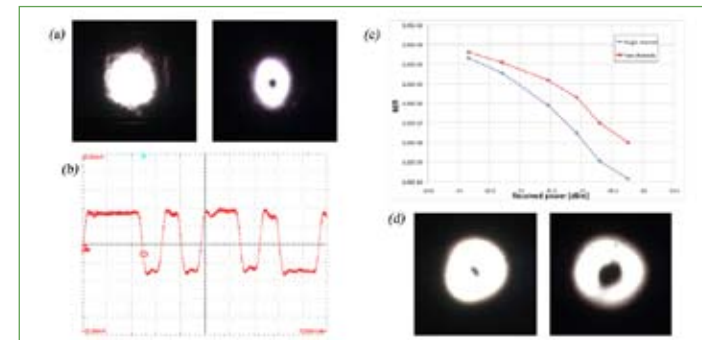
This doctoral dissertation concerns the study, design and experimental characterization of an optical system based on orbital angular momentum division multiplexing. The aim of the work is the realization of a complete innovative multiplexing system based on the phase properties of light beams carrying orbital angular momentum, usually named "optical vortices". These particular electric fields are characterized by the presence of a phase term  $\exp(il\varphi)$  that gives a helicoidal profile to the phase structure (see Fig. 1). In telecommunications, different ways of multiplexing signals together for increasing the overall capacity of the communication systems are widely used. Commonly employed methods in optical communications are the wavelength division multiplexing and the polarization division multiplexing. In the latter method two orthogonal states of spin angular momentum (i.e. polarization) of the light are multiplexed, in order to double the system capacity. In principle, also the orthogonal states of orbital angular momentum, which are identified by the integer values of  $l$  and hence are unlimited, can be multiplexed for increasing the capacity of the communication system. The orbital component of the

angular momentum of light has captured the interest of scientists only in the last two decades: this property of the electromagnetic waves is so still not completely comprehended and theoretical and experimental studies are presently performed. Its exploitation as a "label" of different propagating channels has been hypothesized for free-space systems, although none has ever realized a real operating system. Moreover, an orbital angular momentum division multiplexing in optical fibers has never been considered until now. In this work, a complete orbital angular momentum division multiplexing system is realized in free space and in fiber-optic propagation space. The exploitation of the orbital angular momentum as a "dimension" of multiplexing in advanced optical networks can lead to a breakthrough in optical communication systems. First, close attention is paid on the obtaining of the general expression of orbital angular momentum of electromagnetic waves and its relation with the classical spin angular momentum, i.e. the state of polarization of the beam in the classical domain. From this full study we will recover that helicoidal phase structure is directly related to precise values of orbital angular momentum. For this reason,

two families of light beams endowed with helicoidal phase profiles are presented and analyzed, the Laguerre-Gauss and Bessel beams. Their propagation properties and the relation to orbital angular momentum are obtained. Because this work takes into account only laser sources, the analysis has been performed exploiting the paraxial approximation. The free-space propagation doesn't introduce any impairment, thanks to the conservation law of the orbital angular momentum in a homogeneous space. Fiber-optic propagation, instead, is a more complex process and forms one of the main themes of this dissertation. The classical modal theory is reported, while the introduction of vortex modes requires the development of a new theory applied to the relation between higher-order fiber modes and guided vortices. The selective excitation of an individual fiber mode is also analyzed by numerical simulations, because this is one of the most important requirements for the development of future fiber-optic orbital angular momentum division multiplexing systems. Then, the different stages required by the new multiplexing system proposed in this work are presented. In particular,



1. Phase structure of (a) standard plane waves, (b) optical vortex with  $l = 1$  and (c) optical vortex with  $l = 3$ .



2. Experimental results on free-space and fiber-based OAM division multiplexing systems. (a) Intensity distributions detected by a CCD camera at the output ports of the modal demultiplexer when two channels characterized by an even (left) and odd (right) OAM are propagating in the system. (b) Signal trace of the detected signal at 1.25 Gb/s visible at one output port of the demultiplexer. (c) Bit error rate measurement after the orbital angular momentum demultiplexing. (d) Selective excitation of an individual fiber vortex with topological charge  $l = 1$  (left) and  $l = 2$  (right) over 1 km of standard single-mode fiber

close attention is paid on the generation of light beams characterized by a precise value of orbital angular momentum. Different generating methods are reported in terms of efficiency, reliability and stability,

in order to identify the one that permits to obtain better results. The demultiplexing stage is also found and its behavior respect to light beams with orbital angular momentum is fully analyzed. Experimental results are then

presented about free-space propagating systems (see Fig. 2). The different stages of the system are evaluating in terms of mode purity, crosstalk at the demultiplexer and robustness to environmental perturbation. First experimental results are obtained on free-space systems at a wavelength of 633 nm in Continuous Wave conditions. Then, two modulated 1.25-Gbit/s free-space channels are propagated in the system, correctly demultiplexed and received at 1550 nm (see Fig. 2(a)-(c)). This is the first realization of an orbital angular momentum division multiplexing system with high bit-rate modulated signals. Then, experimental results about fiber-based orbital angular momentum division multiplexing system are presented. System performances in terms of modal crosstalk at the receiver are obtained and the first realization of a fiber-based orbital angular momentum division multiplexing system is presented. Finally, the selective excitation of individual higher-order fiber vortex modes is considered (see Fig. 2(d)).

## HIGH-SPEED SINGLE-PHOTON CAMERA

### Fabrizio Guerrieri

Two-dimensional monolithic imagers are commonly used in many applications such as biology, astrophysics, telecommunications and security just to name few of them. As the applications involving photo-detectors become more complex and advanced, the requirements for imagers follow suit. Nowadays, the most demanding applications require high-sensitivity (possibly down to the single-photon level) together with very high acquisition speed (10 kframe/s and above).

In order to detect such low intensity signals a very low noise detector is necessary. In photo-detection systems sensitivity is mostly limited by the noise introduced by the very first electronic stage sensing the detector. If single-photon sensitivity is required, either a drastic decrease of the front-end electronic noise or an extremely long integration time is required. Since the latter is not generally compatible with most applications, extremely low noise circuits or noiseless signal amplification must be employed. For instance, in Photo-Multiplier Tube (PMT) a series of dynodes are exploited to almost noiseless amplify the photo-generated signal before it reaches the first noisy electronic component, thus allowing reaching single-

photon sensitivity. On the other hand PMTs lack spatial resolution (they are point detectors) and tend to be bulky, fragile and very sensitive to magnetic fields. A similar approach is used by special Charge-Coupled Devices (CCD) to get very high sensitivity but their readout mechanism prevents the imager from reaching high operating speeds. Other types of imager exist, but the trade-off between sensitivity and speed almost always limits their performance to levels that are well below what required by the most demanding applications.

A Single-Photon Avalanche Diode (SPAD) is solid-state device able to detect single-photons by the means of an internal amplification process. The SPAD is a "trigger" (Geiger-mode) detector, hence it differs from the most common type of solid-state imagers that work in the analog regime. The signal generated by the SPAD is in fact inherently digital and macroscopic and thus it is suitable to be used to count photons at high speed. Moreover a digital signal can be safely transferred at high speed in integrated circuits making possible to develop very high-speed photon counting imagers. State-of-the-art SPADs are fabricated in custom technologies, have spectral efficiency tailored to the

wavelength of interest, and provide the best performances in terms of detection efficiency and dark counts. Unfortunately, the process used to make custom SPADs cannot be used to integrate on-chip microelectronics too. Therefore, in SPAD arrays fabricated with custom technologies, the sensor chip must be connected to off-chip front-end electronics, thus reducing the detector performances and making custom SPAD arrays limited to tens or hundreds of pixels at most.

In recent years some research groups reported on SPADs fabricated in standard high-voltage or even low-voltage CMOS processes with sufficiently high degree of purity. Such CMOS compatibility opened the way to the development of arrays of SPADs and to the monolithic integration of SPADs with front-end and processing electronics. The SPADlab research group at Politecnico di Milano, where this Ph.D. thesis work have been developed, also had designed and fabricated some CMOS SPADs at the time this Ph.D. thesis work started and it has almost 30 years of experience in custom SPADs and the relative front-end circuits. Early CMOS SPAD arrays reported in literature were either composed by a relatively low

number of pixels operated in parallel or by a quite large number of multiplexed detectors. Both solutions are unsuitable for most applications where single-photon counting is required together with very high speed. In particular, a multiplexed access impairs the high-frame rate capability of SPAD arrays. The goal of this Ph.D. thesis work was the design and fabrication of a single-photon camera system based on a dense CMOS SPAD array able to work at very high frame-rates. To properly operate a SPAD a special front-end circuit called quenching circuit is required. The first result of this research work was the design of a novel, compact, high performing quenching circuit, the Variable-Load Quenching Circuit (VLQC). In order to achieve high frame-rates each pixel of the arrays to be developed need be able to work as independent single-photon counting channels. To achieve so, an expandable architecture for photon-counting arrays, based on compact smart pixels comprising both detectors and electronics, have been conceived. Each smart pixel contains a CMOS SPAD detector, a VLQC circuit, a counter and a memory. Based on the devised architecture both a linear 32x1 and a two-dimensional 32x32 CMOS SPAD imagers have

been designed, fabricated and characterized as a part of this thesis work. A complete camera system based on the 32x32 CMOS SPAD imager, the so-called SPAD-camera, was also the developed and fabricated. The camera was used to acquire very fast imaging events in very low illumination conditions proving its capabilities in demanding scientific and industrial applications. In addition to the complete development, fabrication and characterization of a single-photon camera system, this thesis work also comprises the use of the SPAD camera in real scientific applications that have been done in collaboration with important international research groups. The first example is a Fluorescence Correlation Spectroscopy (FCS) experiment performed at the University of California in Los Angeles (UCLA). The camera was installed in UCLA laboratories, properly modified to better fit the experiment's requirements and eventually used to do several experimental acquisitions with the goal of developing an innovative high throughput parallel FCS system. This research project is still going on at the time this thesis was written, but promising results were already obtained. A second example is a one-year-long collaboration between

Politecnico di Milano and Massachusetts Institute of Technology (MIT). The SPAD-camera developed at Politecnico was used at MIT to demonstrate, for the first time, a novel super-resolution technique allowing to acquire images with a higher spatial resolution than conventional diffraction-limited optical imaging systems. Three-dimensional imaging is another exciting applicative field where the SPAD camera has been tested. By the means of an active illumination technique it is possible, with a simple bi-dimensional imager, to extract the depth information from a scene. Even though the SPAD camera was not developed for 3D imaging applications, depth information has been extracted with promising results.

Eventually, this research work achieved several important results and the developed technology proved to be effective and easy to be used in important scientific applications. As the interest of the scientific community toward low-light high-speed imaging increases, the opportunities for CMOS SPAD arrays and the SPAD-camera become bigger. The development of this technology can result in further important achievements that can have an impact on important issues of biology, medical science, imaging and security.

## IMPROVING SINGLE PHOTON AVALANCHE DIODES FOR TIME CORRELATE SINGLE PHOTON COUNTING

Antonino Ingargiola

Silicon Single Photon Avalanche Diodes (SPADs) have nowadays gained a wide acceptance for a broad range of applications in fields like 3-D imaging, or single molecule investigations. In comparison to other singlephoton detectors, such as PMTs, SPAD devices present multiple advantages: higher photon detection efficiency (PDE), higher timing resolution, lower dark count rates (DCR), low-voltage operation, ruggedness and lower cost. In recent years we observed a strong impulse toward parallelization of the measurements systems which employs SPAD detectors. For this reason there is a great interest in developing monolithic SPAD arrays in order to enable the creation of compact multi-channel systems. In order to avoid electrical crosstalk between pixels in SPAD arrays, the avalanche sensing discriminator must employ an high threshold voltage (more than 100 mV). However, if not properly engineered, SPADs exhibit seriously impaired photon-timing at high threshold values. Understanding the physical mechanisms that determine the SPAD photon-timing performances it is of paramount importance, not only to be able to develop SPAD devices with lower photon-timing jitter, but also to enable

the developing of SPAD arrays in which each pixel has nearly the same performances of the standalone detector.

In the past, the physics of the avalanche current growth in SPAD devices has been studied and the main physical mechanism taking place during the avalanche multiplication have been outlined [24, 25, 23, 26, 46, 45]. Moreover, an avalanche growth model that fits the photon-timing jitter of devices available at the time has been proposed [45]. The main mechanism responsible of photon-timing jitter has been found to be the statistics of photons injection position on the device active area.

In recent devices, the field profile has been engineered in order to reduce the tunneling contribution to dark count rate and to increase photon detection efficiency [12]. Furthermore, while at low thresholds the new-generation devices have a photon-timing jitter well below 30 ps, at high thresholds the jitter increases considerably. The previously proposed model fails to explain this trend of the photon-timing jitter. Therefore, the need arises to investigate to a greater extent the avalanche physics in order, not only to explain the timing performances of current devices,

but also to provide tools that can assist the development of new generation devices.

In this dissertation the physics of avalanche growth has been thoroughly investigated from ground up both from a theoretical and an experimental point of view.

The avalanche growth can be roughly divided into two phases: the buildup phase and the lateral avalanche propagation. Tan *et al.* [48] investigated SPAD buildup statistics within the assumption of constant electric field, that is a reasonable assumption for III-V SPADs. Since in our silicon devices the electric field profile is far from constant this model is not accurate enough to provide a good estimate of the buildup contribution to the total photon-timing jitter. The buildup study is also important to accurately describe the subsequent avalanche lateral spread. In fact, the lateral propagation of the avalanche current is a multiplication-assisted diffusion phenomenon [25, 45] and is thus deeply tied to the buildup growing rate and jitter. Therefore, a Monte Carlo buildup simulator including the real (non-constant) electric field profiles has been developed both to investigate the buildup contribution

to the photon-timing jitter and to extract physical parameters required by the lateral avalanche propagation modeling. In order to achieve a good accuracy different models for the field-dependent electron and hole ionization coefficients have been included and calibrated. The buildup simulations were compared to lowthreshold photon timing jitter measurements and has been found that the buildup contribution is appreciable compared to the total photon-timing jitter.

A second photon-timing jitter contribution is related to the lateral avalanche propagation phase. When increasing the threshold in the avalanche sensing circuit, new-generation devices experience an increase in the photon-timing jitter which can not be ascribed to the buildup phase. During the present research the photon-timing jitter as a function of the injection position has been characterized, by mean of a custom TCSPC setup employing a laser focused on the SPAD active area. For the first time, an interesting dependence of the timing performances on the avalanche triggering position is reported. Furthermore, the fact that the low-threshold photon-timing jitter gets better moving the injection position from the

center towards the border of the active area, is an evidence that resistive effects play a significant role in determining the SPAD timing performances. Resistive effects are ascribed to two different physical phenomena: the space charge effects and the neutral region (bulk) resistance. The full 3-D SPAD structure has been simulated in order to obtain the bulk resistance experienced by the avalanche current as a function of the injection position and the activated area. The simulations have been validated by electro-luminescence measurements which allowed to obtain the steady state current distribution on the device active area. Finally, in order to provide a practical tool to help driving the engineering process of new devices a figure of merit (PTFoM) as been defined and validated. This tool has proven to be useful in comparing different designs for what concerns the photon-timing jitter performances.

Finally, a 2-D simulator of the lateral avalanche spread has been developed. The simulator includes a novel SPAD model including: the buildup current growth time constants as obtained from buildup simulations and depending on the electric field profile; non-local space charge effects

obtained by 3-D electrostatic simulations; distributed bulk resistance calculated from 3-D resistance simulations; an accurate lumped-component description of the external circuits and parasitic. The avalanche spread is simulated solving the multiplication-assisted diffusion equation.

A statistical model that accounts for the photon re-injection effect is included. The simulated current for different injection positions are in good agreement with the current acquired by means of time-resolved electro-luminescence measurements that employs a focused laser to fix the injection position. In conclusion, including a 3-D statistical model for secondary photons avalanche re-triggering, the simulator has been able to produce a photontiming jitter estimate that shows a vastly improved match with experimental jitter compared to the previous state-of-art models.



# SAFE MOTION PLANNING AND CONTROL FOR ROBOTIC MANIPULATORS

Bakir Lacevic

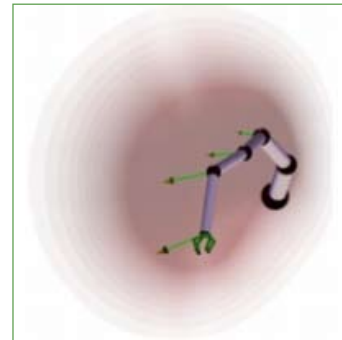
In recent years, robots have been increasingly moving to everyday human environments, such as the home or office. The barriers between humans and robots in the industrial environment also seem to vanish. The growing demands for establishing the interaction between human and the robot raise several important issues. A weak interaction concept considers humans and robots sharing the same workspace, while working on different tasks (coexistence), while strong interaction implies humans and the robots sharing the same task (cooperation). The highest priority requirement in both cases is certainly the human safety.

The concept of safety should include both physical integrity of the human (robot must not physically injure the human), and psychological aspect, since the motion of the robot must not cause any unease or discomfort (like fear or shock) to the human. Planning and control capabilities should ensure safe coexistence and cooperation and compensate the unavoidable insufficiency of intrinsic safety as an objective of mechanical design.

This thesis represents an attempt to establish a framework for planning and control for the safe behavior of the robotic manipulators. The framework exploits the

knowledge of the system that is built upon the perception skills. This knowledge about the relations between robot, human and the environment includes proximity/velocity measures and a purposefully designed danger/safety assessment called *danger field* (see Figure 1).

This quantity captures the state of the robot as a whole and indicates how dangerous the current posture and velocity of the robot are to the objects in the environment. The field itself is invariant with respect to objects around the robot and can be computed in any given point of the robot's workspace using measurements from the proprioceptive sensors. Furthermore, the danger field is closed form computable, which allows for real time applications. Apart from being a pure safety assessment, the danger field provides a natural prelude to a purposeful design of a safety-oriented control strategy. Namely, the information about the danger field can easily be fed back to shape standard control schemes in order to make the motion of the robot safer to the environment. According to the degree of danger, the robot reacts promptly, always maximizing the human safety. The essence of the approach is based on the generalization of the impedance

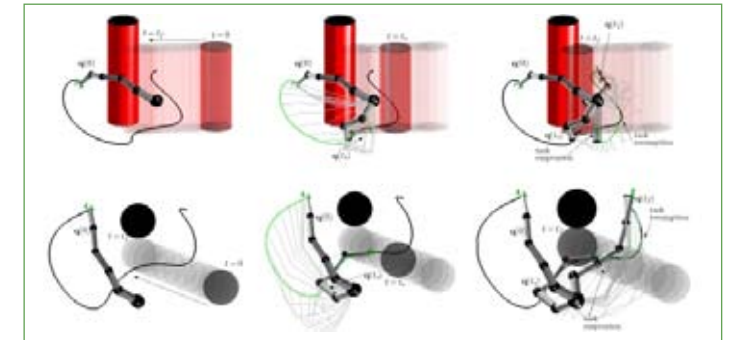


**1. Isosurfaces of the danger field around a 6 DOF manipulator. The velocities of the link endpoints are indicated**

control problem (for instance an impedance controller based on the virtual force - a quantity, calculated usually from the visual data, and fed back to the control system) and a potential field method. Further, the possibilities of exploiting the redundancy are investigated, e.g., shaping the internal motion of the robot in order to minimize an overall danger without compromising the end-effector motion (see Figure 2).

Above the reactive, danger field based control, there is a global path planner that does not only provide the collision-free paths but strives for safer ones in order to reduce the overall danger in the planning phase. This is achieved by embedding a suitable safety-oriented

heuristic function into some well known planning paradigms. Four different path planning algorithms have been proposed. The first is based on so-called bubbles of free configuration space. The algorithm founded on the bidirectional A\*-search that grows two trees in the configuration space via bubbles. The algorithm is proven complete. The second one path planner is based on deterministic sampling of the configuration space using low-discrepancy sequence and then performing a bidirectional search over the set of samples. The algorithm uses a novel collision checking sub-routine, which is based on bubbles of free configuration space. Moreover, two safe motion planning algorithms that use rapidly exploring random trees paradigm are presented. Both algorithms are the safety-oriented modifications of existing planners. The first uses the information about the workspace-defined goal to guide the sampling in the configuration space, whereas the second is a classical bidirectional algorithm whose goal is to merge two trees in the configuration space. Both safety-oriented algorithms turned out to provide substantially safer paths than their classical counterparts.



**2. A robot's motion in the non-stationary environment. One of the obstacles moves and intersects the desired path of the manipulator. Initial configuration (left), the moment when the danger field reaches the critical value and the task becomes suspended (middle), the deviation from the desired path and eventual task resumption (right). Images in the second row show the scene from the bird's eye view.**

Finally, a special attention is devoted to the guarantees for gentle and compliant behaviour of the robot when the intended contact between human and the robot has been established. The proposed method is based on deriving the controller in order to achieve prescribed endpoint admittance of the manipulator. Joining complementary stability approach with a simple target admittance enabled very simple techniques for the closed-form controller design.

## A METHODOLOGY FOR PERFORMANCE ESTIMATION OF HETEROGENEOUS MULTIPROCESSORS EMBEDDED SYSTEMS

Marco Lattuada

In the last years there has been a considerable growing in the performance request in the embedded systems. To satisfy this growth, multiprocessor heterogeneous systems have become the de-facto standard in the embedded systems. The use of this type of architecture introduces new problems in the design flow. Indeed, to fully exploit the computational power provided by this type of architecture, we need to partition and map applications onto the target architecture. These complex activities can be performed by exploiting automatic design space exploration techniques, by hand or by combining these two ways. There are some critical analyses which in any case can not be too approximated. One of them is the analysis of the application performance since meeting the performance constraints is a mandatory objective for the design flow and timings constraints are among the most critical aspects of the embedded systems design flow. Verifying these constraints only in the last phases of the design process is not possible: fixing the system in the last design process stages can require significant changes of the system itself delaying the final system production of an unsustainable amount of time. For this reason, system

performance has to be evaluated from the early design steps. Despite of the complexity of this process, the time-to-market of embedded systems has not been extended to allow to treat exhaustively these new problems. On the contrary, a faster design flow is required by manufacturers to make products available on the market as soon as possible. This acceleration of the design process potentially impoverishes the quality of the design results producing not optimal solutions. Nevertheless, we need to guarantee that these solutions meet the performance constraints. There are some specific problems in estimating the performance of a complex system such as an MPSoC. Before estimating the performance of the overall application, it is necessary to estimate the contribution of the single tasks on the different processing elements. Since the hardware of an embedded system is not fixed, often the system designer has to select which components will be included into the final system and has to guarantee that these components will satisfy timing, power and economic constraints. Judging which are the more appropriate components is not a trivial task: in principle the designer has to evaluate the performance of each available candidate on the

different portions of the application which are supposed to be mapped on it. Finally, assumed that the processing elements have been chosen and correctly estimated, the estimation of the overall performance can be itself a problem since interactions among different tasks on different processing elements have to be taken into account. Both for single processing elements and complex architectures, the methods to early evaluate the performance can be mainly divided into three categories: direct measures, estimations by simulation, estimations by use of mathematical models. Most of the times the first solution is not affordable since the real components are not available during these phases of the design and integrating the measurement on the real component in a Design Space Exploration framework can be a difficult task. For these reasons, techniques based on estimations have to be preferred. In the simulation based ones estimation is achieved by simulating the behavior of a component or of a system on a particular code by running a Cycle Accurate Simulator. Finally, in the last category techniques, the estimation is obtained by exploiting mathematical models which correlate some numerical

features of an application with its performance. In general, they are less accurate than the simulator based, but they are much faster so they have to be preferred in the early phases of the design flow. In this thesis we propose a new methodology flow for fast evaluation of design solutions composed of several techniques and based on the static analysis of the application, expressed in form of Hierarchical Task Graphs, combined with profiling information retrieved by executing the application itself on the host machine. In all the design flow we exploit the intermediate representations of the GNU GCC compiler to represent, analyze and decompose the different design solutions. This methodology, differently from most existing ones, does not require the designer to have any knowledge of the characteristics of the components of the target architecture to guarantee accurate estimations. In this way considered target architectures can be easily extended by adding new components. The proposed methodology is mainly composed of three parts: profiling, single processing element estimation and task graph estimation. The first part of the flow is composed of two different profiling techniques. In particular, we propose a target profiling technique based on target-independent automatic instrumentation of source code. This technique is used to collect information about the characteristics of the components of the architecture (e.g., synchronization costs). The second proposed technique

is instead a host profiling technique that combines source code instrumentation with static analysis to collect information about the path (i.e., sequences of branches) executed in the control flow graph of the application. The second part of the methodology flow consists of a set of techniques used to estimate the performance of the single tasks: starting from profiling information collected in the previous phase both on target and on host machines and exploiting linear regression technique, the methodology builds linear performance models of the single processing elements which compose the architecture. The performance models are built using as input sequences of operations extracted from the GNU GCC internal intermediate representations. Use of this type of representations allows to extract and exploit the knowledge of the target processing elements embedded in the GCC itself; use of sequences instead of single operations allows to model the performance effects of interaction among operations (e.g., pipeline interaction). On the contrary, the last part of the methodology flow consists of two techniques for the actual performance evaluation of the whole partitioned application on an heterogeneous multiprocessors embedded system. Proposed techniques can be used in different scenarios since they have different requirements. The first technique requires the availability of some prototypes of the final architecture, but can be used even if part of the application is provided as library code. It combines information coming from static analysis, host

and target profiling to compute the estimation of the whole application. Finally, the second technique does not require any components nor prototype of the final architecture but only that the source code of the whole application is provided. Estimation of the execution time of the application is produced combining static analysis results, host profiling information and performance models built in the previous phase of the flow. We validate the proposed flow by applying it to a set of standard embedded system benchmarks targeting multiprocessors architectures based on ARM processor, LEON3 processor and MAGIC Digital Signal Processor. The results show that the proposed flow is able to estimate the performance of the single tasks (e.g., 7.38% estimation error on LEON3) and of the whole task graph with good accuracy (4.1%).

## SYSTEMS OF OPTICAL INTERCONNECTIONS

### Diego Lucerna

The tremendous growth of the Internet and the introduction of new communication services and applications have changed the landscape of the telecommunications industry. Today, telecommunications play a vital role in our world and this role will continue for as long as we need to communicate and deal with each other. Innovations in the field of communication technologies and networking continue to unfold and will keep changing the way we live our lives, perform our work, and interact with each other. The emergence of broadband multimedia services and distributed applications has resulted in a great amount of bandwidth requirements. Optical Wavelength-Division Multiplexed (WDM) networking technology offers a promising solution to this huge bandwidth requirement foreseen for the next generation Internet. The WDM innovation represents itself a revolution inside the larger optical communications revolution, as WDM sustains and fuels the latter exponential growth. The success of optical fiber communications is based on the invention of the laser, particularly the semiconductor junction laser, the development of low-loss optical fiber, of the Erbium-Doped Fiber Amplifier (EDFA), and of other components

such as integrated optical devices. We should never forget that it took more than 25 years from the early pioneering ideas to the first large-scale commercial deployment of optical communications: the system linking Washington - New York (1983) and New York - Boston (1984). This is when the revolution got started in the marketplace, and when optical fiber communications began to seriously impact the way information is transmitted. The market demand for higher capacity transmission was helped by the fact that computers continued to become more powerful and needed to be interconnected. This is one of the key reasons why the explosive growth of optical fiber transmission technology parallels that of computer processing and other key information technologies. These technologies have combined to meet the tremendous global demand for new information services, based on a huge amount of exchanged data (i.e., the broadband services). Clearly, their rapid advance has helped to fuel this demand. WDM technology has an interesting parallel in computer architecture. Computers have a similar problem as lightwave systems: both systems trends, pulled by demand and pushed by technology advances, show their

key technological figure of merit (computer processing power in one case, and fiber transmission capacity in the other) increasing by a factor 100 or more every ten years. However, the raw speed of the IC (Information and Communications) technologies, which computer and fiber transmission rely on increases by about a factor of 10 only in the same time frame. The answer of computer designers is the use of parallel architectures. The answer of the designers of advanced lightwave system is similar: the use of many parallel high-speed channels carried by different wavelengths (WDM). In early 1996, several research laboratories reported prototype transmission systems breaking through the Terabit/second barrier, i.e., the currently information capacity carried by a single fiber. This breakthrough launched lightwave transmission technology into the "tera-era". Five years later in 2001, a WDM research transmission experiment demonstrated a capacity of 10Tb/s per fiber. This is an incredible capacity: recall that, at the Tb/s rate, the fiber can support a staggering 40 million 28kB/s connections, transmit 20 million digital voice telephony channels, or a half million compressed digital television channels. Even more importantly, we should recall that the dramatic increase in

lightwave systems capacity has a very strong impact on lowering the cost of long-distance transmission. Nowadays, thanks to the utilization of Raman fiber amplifiers that are being employed in addition to the early erbium-doped fiber amplifiers, of new fibers, of new techniques for broadband dispersion compensation and broadband dispersion management, the potential bit-rate per WDM channel has increased to 40Gb/s and higher. Unfortunately, the capability of switching and routing will never been able to support the perspective spectacular improvement on transmission capability if it will be still implemented using only the conventional hard-wired architecture. As a matter of fact conventional hard-wired technology has fundamental drawbacks in wide bandwidth transmission: the high frequency components of a signal are attenuated due to increased conductor resistance in the high frequency region, and also cable material degrade the signal intensity as frequency increases. The transmission distance has to be shortened for higher transmission rates to occur. This limitation has a crucial effect on the system design. Possible approaches for higher transmission capacity using conventional hard-wired

technology include multiplication of channels for more efficient throughput. An increase in the number of parallel channels, however, makes the so called skew problem (misalignment of data arrival time) more pronounced. Many techniques have been devised to address these problems: addition of skew-adjustment circuitry, a pre-emphasis method for enhancing high frequency signal components, and an equalizing method to adjust the frequency component strength when a signal arrives. All these techniques have a serious side effect in that they require a larger power consumption. Based on these considerations, it is generally accepted that the useful range of bandwidth for hard-wired circuitry is limited to below 10 Gb/s per-channel. To address the imminent hard-wired bottleneck problem, the development of an alternative data transmission technology that can replace the conventional method is urgently needed. Short-haul optical interconnection is the most promising candidate for providing a solution for hard-wired deadlock. In terms of linkages using optical means, optical fiber communication has been in practical use for more than 20/30 years, mainly for long distance links between cities and continents. However,

as the problem associated with hard-wired links become more apparent, this approach is gathering new attention. Therefore, this dissertation deals with the topic of optical interconnections with aim of seeking new architectures or subsystems to be adopted at multiple hierarchic levels of a WDM network: in *part I* we focus on long-haul or local-area optical scenarios characterized by kilometeric distances between devices; in *part II* we study the innovative optical interconnection scenarios on uncommon optical distances "inside the box" (in the order of meters/centimetres), i.e., to implement rack-to-rack and backplane connections.

## IMPLEMENTATION OF A CELL REDUNDANT BASED BMS FOR LI-ION BATTERY PACK USED IN LOW AND MEDIUM POWER ELECTRIC VEHICLE APPLICATIONS

**Antonio Manenti**

Nowadays, due to the growing cost of fossil fuel and the issues related to atmospheric pollution, and thanks to a higher awareness from manufacturers and consumers about the opportunity to use alternative energy fonts for mobility, more and more vehicles, electric or hybrid, are being produced.

This pushed the batteries market and, in turn, those related to the battery management systems, to develop more performing systems able to increase the operative efficiency and the lifetime. The main application areas are extremely various ranging from chemistry, to modeling, through system control, up to power electronics. All this fields play key roles for the reliability and effectiveness of electric vehicles, also hybrid or plug-in.

In particular, during the major research I worked on systems for the control of the power flow from and to the accumulator, especially in systems based on multi-accumulator configuration (battery packs). Moreover, I worked on mathematical battery models able to closely mimic the behavior of the accumulator under different operating conditions.

In the first part has been tackled the problem of voltage unbalance (and hence the State of Charge imbalance) between cells belonging to a battery pack.

This phenomenon end up to be one of the main causes that limit the lifetime of battery packs made by multiple cells serially connected. Moreover, for some chemistries, and especially for battery pack based on Lithium cells, a prolonged voltage imbalance can cause potentially dangerous situations since the cell can even explode or set in fire. In order to maintain a correct state of charge along the battery pack and guarantee uniform cell degradation is necessary to properly manage the charge and discharge current profiles of every single cell. In this context, after a careful analysis of the solutions already proposed in literature, several solutions have been evaluated and a new BMS (Battery Management System) architecture for low-medium power vehicles has been designed, prototyped and experimentally tested. Such system can be classified as a non-dissipative, bidirectional equalizer. First of all, this means that the proposed BMS does not use extra-power to prevent imbalance, thereby increasing the overall efficiency. In addition, it continues to work during both charge and discharge operations, improving balancing effectiveness. This system is especially suited for linear packs made up of maximum 10 Li-ion cells for low-medium power applications (in the range from

100W to 500W). This solution can be oriented to light electric vehicles (i.e. electric bicycles), which usually need low-medium voltage (24V-36V) and current (10A) associated to a simple, inexpensive and space saving layout. In fact, no extra components such as DC/DC converters, power inductors, and transformers are needed by the BMS. The proposed BMS strategy is based on the availability of an additional cell that can be continuously disconnected from the load. In this way, the bypassed cell is isolated to be optimally balanced while the remainder of the pack continues to provide energy to the load. At any time, while the pack works at least one cell is disconnected from the string. There are three main features of the proposed architecture. First, balancing is active in both the charge and discharge phase. Second, the BMS layout is considerably space saving, which is a primary task for two-wheel vehicle applications that are often subjected to hard space constraints. Third, the solution offers an intrinsic hardness to a cell fault. In fact, if the pack should contain a cell with a severe fault, the BMS can identify and bypass it permanently. Moreover, the proposed architecture is very flexible, economical and stable. Compared with other balancing

systems it does not require any specific driver or PWM controller (that can strongly affect the overall cost), neither high power inductors nor diodes (responsible of space consumption). The MOSFETs used to implement the switches do not need to meet any special requirement except the low on-resistance. So, in future, the development of a custom IC integrating the switch transistors (MOSFETs and BJTs) could be considered. A drawback is that a substantial effectiveness of the technique limits the dimension of the pack to about 10 cells, i.e. up to 500W. This means that the proposed BMS architecture is suited for low-power automotive applications, such as electric scooters and bikes. For higher power demanding loads, a modular combination of several cell-redundant BMSs can be conceived. Of course, in this case the complexity of the layout is also sensibly increased.

The second part of the research focuses on modeling of Lead acid (PbA) batteries for deep cycle automotive applications. The complexity of designing battery models comes from the intrinsic nature of these systems that involve both electrical and electrochemical aspects. On the other hand, a suitable battery model representative of the behavior under certain operating conditions, allows to predict the system output, which in turns is important when a model-based diagnosis/prognosis framework is to be built. This is motivated by safety and reliability issues. Depending on the particular application, each model has to meet different requirements. Often, especially in vehicle

applications, it is desirable to deal with low order models which do not require high computational effort, hence excluding first principle, electrochemical models. The goal of this type of models might be limited to track specific variables, for instance state of charge (SoC), terminal voltage, body temperature etc. Over the years, the input-output behavior of batteries has been modeled mainly through electrochemical and equivalent electrical circuit models. The former are usually adopted by battery manufacturers and researchers to understand the influence on battery performance of the use of new materials, change in physical dimensions and tolerance on abuses. Complexities of electrochemical models generally prohibit them from being used effectively in embedded systems where computational power is usually limited. The latter are more intuitive and easy to handle for all that people that usually deal with battery-powered systems even without a deep knowledge of cell chemistry principles. Equivalent circuit models represent a simplification of electrochemical models by using electrical circuit elements to describe the battery behavior. These are more frequently adopted in on-vehicle applications on which the battery type is well defined and a simple model is used by the BMS to prevent dangerous situations or, for example, to predict the SoC and state of health (SoH) of battery pack. Despite the fact that new battery chemistries are available nowadays for use in EVs, HEVs and PHEVs, PbA batteries still play an important role in automotive applications thank

to their robustness and low cost. They are often used as main storage system in the most of all-electric medium-power vehicles (golf carts, electric forklifts, electric wheelchairs, indoor passenger vehicles) as well as military vehicles. Moreover, this chemistry is extensively adopted in all steady applications where the battery pack does not have to be moved, such as uninterruptible power systems and telecommunication back-up systems as well as storage unit for solar panel applications. Although this battery chemistry has been used for more than a century, some interesting phenomena are still not completely understood and correctly managed by most of the battery models. Among them, there is the so called 'Peukert effect'. Basically, it phenomenologically describes the apparent capacity reduction at high current discharge rates. This part of the research aimed to design, calibrate and validate a new PbA battery model based on equivalent electrical circuit able to rationally capture the aforementioned 'Peukert phenomenon' under all operating conditions.

# VEHICLE-TO-DRIVER, VEHICLE-TO-VEHICLE AND VEHICLE-TO-ENVIRONMENT INTERACTIONS: MODELING, CONTROL AND DATA-ANALYSIS

Vincenzo Manzoni

Transportation systems are an important tool for applying the latest information technology results. In particular, two areas have lately been attracting more interest. On one hand, data analysis techniques can be applied to better understand the environmental impact of vehicles and can act to reduce their carbon footprint. On the other hand, the pervasiveness of sophisticated small devices, such as mobile phones, can dramatically improve the interaction between the vehicle and the user, other vehicles and the infrastructure. The goal of reducing CO<sub>2</sub> emissions has increased demand for vehicle engines that emit less and less carbon dioxide per kilometer. It is well known that driving style has a strong impact on fuel consumption, which directly corresponds to CO<sub>2</sub> and NO<sub>x</sub> emissions. Previous research demonstrates that by improving the driving style, drivers can reduce their fuel consumption. However, it is not easy to quantify the driving style of a driver without impacting on the vehicle system architecture. CO<sub>2</sub> emissions can also be reduced by leveraging the rising interest in electric vehicles. An accurate estimation of the power demand according to different deployment scenarios is needed to motivate people to invest in green power plants. Moreover,

such knowledge can also drive long-term investments in new generations. Finally, nowadays mobile phones are not only used as devices for making calls in mobility. They are provided by large touch screens and are usually always connected to the Internet. Moreover, with the increasing computational capability of their processors and the integration of new sensors, such as accelerometers, magnetometers and GPS receivers, they can be exploited as *mobile laboratories* for applications requiring logging and data analysis of inertial data. This Thesis has been developed within this interesting and evolving context, with the aim of proposing innovative solutions which are both effective from the application viewpoint and theoretically sound from a methodological perspective.

## Thesis aims and organization

This Thesis regards the modeling, the control and the data analysis of the vehicle dynamics and of the interaction between vehicles and drivers, other vehicles and the infrastructure. The Thesis is divided in two parts. The first part discusses the estimation of the vehicle dynamics using data gathered from vehicle-independent systems. Three vehicle-independent systems are presented, two based on custom hardware and one based

on modern mobile phones. This methodology together with data analysis techniques has been applied and experimentally validated in three different contexts. The estimation of the driving-style economy and safety using inertial information. The analysis of the grid impact of electric vehicles using inertial. The automatic identification of transportation modes and a real-time estimation of CO<sub>2</sub> emissions using mobile phones. In the second part, modern mobile phones are exploited as middleware for the interactions between the vehicles and drivers, other vehicles and the infrastructure. First, using a smartphone and a wireless Bluetooth medium, an original interaction system for motorcycles based on vocal synthesis and speech recognition has been implemented. Secondly, this system has been used as an interface between the driver and the hardware and software implementation of the architecture of the European Project SAFESPOT. SAFESPOT aims to assess and demonstrate how the cooperation among vehicles can improve the road safety. In this context, one of the key problems is providing an accurate absolute position estimation and it is interesting to understand how low-end receivers perform in this scenario. Therefore,

an accuracy assessment of widely deployed GPS receivers and an algorithm to improve the position estimation based on satellite constellation is presented.

## Driving Style Estimation Algorithm

A correct driving style can strongly affect the fuel consumption and therefore the CO<sub>2</sub> emissions. This Chapter proposes a method to quantify the driving style fuel economy and safety via inertial measurements. Firstly, it presents a low-cost and vehicle-independent system for the acquisition of the variables related to the dynamics of a vehicle. Then, it describes the signal processing and the algorithm which calculates the over-consumption of energy with respect to different reference profiles. Finally, experimental results show the algorithm applied to real data gathered from public transportation vehicles.

## Analysis of the Grid Impact of Electric Vehicles

This Chapter proposes a methodology to estimate the mechanical energy at the wheel of electric vehicles through inertial measurements. Simulations based on experimental data have been carried out to analyze the grid impact of electric vehicles and the size of their battery pack according to different recharge policies and the presence of a system to accumulate energy from a regenerative brake. The experimental campaign is carried out with different types of cars, traveling in urban and out-of-town roads.

## Transportation Mode Identification for CO<sub>2</sub> emission estimation

This Chapter proposes a novel method to estimate in real-time the CO<sub>2</sub> emissions using inertial information gathered from mobile phone sensors. In particular, an algorithm automatically classifies the user's transportation mode among eight classes by using a Decision Tree classification algorithm based on features computed from the total acceleration. A second algorithm computes the traveled distance in energy efficient manner, through an optimized mixed use of GPS and Internet map services. Finally, the system makes use of the results of the two algorithms for estimating traveler's CO<sub>2</sub> emissions in real time.

## Smartphone-based Driver-to-Infrastructure Interaction System

This Chapter concerns the definition and implementation of an interaction system for motorcycles. The system consists of a vehicle-to-driver and a vehicle-to-environment communication mechanism based on a smartphone core and a wireless Bluetooth medium. The system is constituted by a vehicle with a CAN bus, an embedded electronic which uses a CAN-to-Bluetooth gateway, a smartphone and a Bluetooth helmet. The driver-to-vehicle system is based on an audio interaction. The vocal synthesis, the speech recognition, and the web gateway are smartphone applications. The Chapter describes the hardware and software architecture and some case-study specific implementation issues. An evaluation of the critical aspects of the system is also provided.

## Implementation of the SAFESPOT Architecture on a Motorcycle

The SAFESPOT Integrate Project is a European co-founded project which aims to understand and assess the potential of a cooperative approach in terms of road transportation safety improvement. This Chapter focuses on implementation of the SAFESPOT hardware architecture designed to collect, fuse, and store data from the vehicle sensors and from the external positioning subsystem of a Powered Two-Wheeler vehicle, the Piaggio MP3 250 cm<sup>3</sup>.

## GPS Optimization for Road Safety Applications

Vehicle localization is an important component of Intelligent Transportation Systems. Localization typically relies on a *Global Positioning System* (GPS). However, for many applications the accuracy and the reliability guaranteed by commercial, low-end GPS devices is not sufficient. In this Chapter a novel method to estimate an index which summarizes the state of the satellite constellation is proposed. Experimental results show that it is correlated with the localization error. An algorithm for the estimation and the correction of the GPS offset is therefore proposed. The improvement of the localization accuracy is demonstrated using an automotive, high-end GPS.

## ENHANCING USER INTERACTIONS WITH HAPTICS: FROM SIMULATIONS TO CONCEPTUAL MODELLING

Davide Mazza

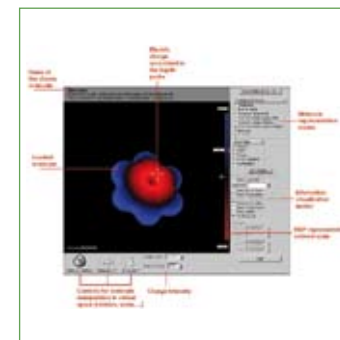
Haptic devices are becoming a common way of user interaction in several fields of applications, from gaming, to mobile, automotive, etc.. Such kind of devices are able to provide force or tactile feedbacks to the user. Figure 1 shows the Sensable PHANTOM®, a widely used haptic device: the user can move taking the stylus in its hand, and can feel a force that opposes or facilitates the hand's movement. While applications adopting haptic technologies are increasing, the common practice for the development of haptic-enhanced applications still consists in writing the application code manually, with the help of proprietary software libraries for interfacing with each used device. In this work, starting from two case studies taken from real-life haptic-enhanced applications, the process of haptic applications' development is abstracted by the definition of a conceptual language that could allow the specification of haptic feedbacks in applications. The language covers those aspects of modelling of that force response to render. The first case study is taken from the chemistry field, and consists in the development of a framework simulating the interaction between an electric charge and a molecule, where the user can feel the interacting force with the haptic device.

This tool can be used as a *technology-enhanced learning* framework for educational purposes in secondary school- or university-level chemistry courses. Figure 2 shows a screenshot from the running framework. The user is shown the 3D geometrical model of the molecule, which is loaded from a file in one of the formats widely used in the chemical and biological field, to rely on the real information available to scientists and researchers in the field. The involved forces to simulate the interaction is computed using the data obtained as the outputs of a computational chemistry tools to make the simulation as more realistic as possible, and adherent to what is the state of the art of the modelling of such phenomena in the research of the field. Molecules can be chosen and loaded from a set of molecules of different nature and complexity stored in a repository. The second case study is taken from nano-manipulation field, and deals with the simulation of a magnetic tweezer, a tool used in research for the manipulation of nano-dimensional objects exploiting the generation of magnetic fields. This simulation is coupled with a real magnetic tweezer and with a haptic device, through which the user can feel the resistance force of moving the nano-scale

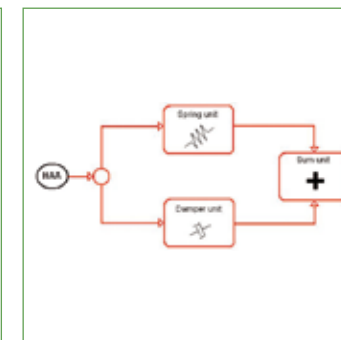
objects into their real physical environments, and exerting a control on the motion of such objects, thus making the tweezer generating the proper magnetic fields to obtain the desired movement. During the development of such cases, a set of studies on force perception by the user has been conducted. In particular, the interest was on determining to what extent the user was able to correctly perceive the force feedback provided, and consequently at which degree the information associated with that feedbacks were correctly understood by the users. A usability study was also conducted on the chemical framework, in order to see how the implemented haptic elements can affect the usability and the way in which it can be improved by the addition of multi-modal elements, e.g., visual / audio effects coupled with the haptic feedback to integrate the force information felt by the user. From the two case studies, a high-level graphical formalism, named *HaptML (Haptic Modeling Language)*, has been defined in order to allow the abstract definition of force feedbacks for haptic-enhanced applications. Exploiting a *model-driven approach*, HaptML allows to specify haptic properties of *haptically-active areas (HAAs)*



1. The Sensable PHANTOM® haptic device



2. A screenshot of the chemical framework



3. An example of force model specification with HaptML

in a 3D virtual environment, and to automatically generate the code for their rendering. In particular, the language identifies the basic components that allow the definition of force models, to be used to define the haptic feedbacks. Such components are called *units*; 4 classes of units have been identified:

- *Force units*, which represent the most used force models;
- *Modifier units*, which allow to modify the force values computed by force units, e.g. by amplifying the force to be rendered, or by changing its duration in time;
- *Composition units*, which allow to build up the final force feedback by composing different force effects;
- *Status units*, which allow to set or retrieve the status variables (position, speed, input force, etc.) of the used haptic device.

Units can be composed in *chains*, in order to obtain the desired haptic effect. Figure 3 shows an example of a chain of HaptML unit. Two force models are here employed: a *spring* force, which determines a position-based force according to the position of the stylus of the device with respect to a given point, and a *damper* force, which a velocity-based force used to model a viscosity effect. These two forces are *summed up* to determine which is the final force felt by the user. HaptML is thus a force modelling language, that allows to specify the *function model* used to define the force to render. Code generation for a specific haptic platform can be obtained with the use of a set of *code template files*, which are parametric template files that define how each single unit of HaptML can be

implemented on a selected target platform. Finally, haptic devices have been applied to improve current Web page navigation and contents exploration. Due to expected spread of haptic devices in next years, they can be used to substitute current devices, e.g. mice, for browsing the Web and other similar tasks. The coupling of the developed formalism with an already existing domain-specific language for Web application model-driven development is proposed, in order to allow the automation of the development process of haptic-enhanced Web applications.

## EVOLUTIONARY TESTING OF STATEFUL SYSTEMS: A HOLISTIC APPROACH

**Matteo Miraz**

Testing should be one of the key activities of every software development process. However it requires up to half of the software development effort when it is properly done. One of the main problems is the generation of “smart” tests to probe the system, which is both difficult and time-consuming. The research community has been proposing several ways to automate the generation of these tests; among them, the search-based techniques recently achieved significant results. This doctoral dissertation presents TestFul, our evolutionary testing approach for stateful systems; it is tailored to work on object-oriented systems. It uses a holistic approach to make the state of object evolve, to enable all the features the class provides, and to generate the shortest test with the utmost coverage for the class under test. We employ several complementary coverage criteria to drive the evolutionary search. We aim to generate tests with high fault detection effectiveness. To this end, we consider the system from complementary perspectives and we combine white-box analysis techniques with black-box ones. The evolutionary search is completed with a local one, and we establish a synergic cooperation between them. The evolutionary search concentrates

on evolving the state of objects, while the local search detects the functionality not yet exercised, and directly targets them. All the proposals were subject to an extensive empirical validation. We devised a benchmark composed of independent benchmarks for tests, public libraries, and third party studies. As comparison, we consider both search-based, symbolic, and traditional (i.e., manually generated by human being) approaches. The achieved results were encouraging: TestFul efficiently generate tests for complex classes and outperforms the other approaches. The proposals presented in this dissertation open new interesting research directions. On one side, one can continue refining the search strategy, by considering more advanced search techniques and by leveraging more advanced coverage criteria. On the other side, one can adapt the approach to work either at a coarse-grained level –and focus on the integration testing– or on other kind of stateful systems (e.g., components or services).

# BEHAVIORAL MODELING, INFERENCE AND VALIDATION FOR STATEFUL COMPONENT SPECIFICATIONS

Andrea Mocchi

This thesis illustrates and consolidates our research effort in the field of component specification modeling, inference and validation. In this thesis, we focused on stateful components, and in particular in data abstractions, which are fundamental entities used to build more complex components.

Historically, formal modeling and specification of data abstractions have been investigated for a long time, but despite of this effort, no solution is universally accepted. Moreover, formal specification writing is a difficult task which requires solid mathematical skills and it is usually (at least) as expensive as writing code. Because of this issue about formal specifications, in practical software development processes they are usually absent, or kept inconsistent during the evolution of software artifacts. However, formal specifications are essential for many software development activities like verification and validation; thus, their absence can be an issue that may impact on software quality.

To address the issue of specification absence, recent research activity has been focusing on specification recovery, that is, on providing automatic or semi-automatic techniques that, through software analysis,

are able to retrieve formal software specifications. The proposed methodologies differ on the specific kind of software artifact to be analyzed and thus on the its formal model to be extracted.

This thesis focuses on specification languages and methodologies for object-oriented software, and in particular specification techniques for data abstraction, which is the most important abstraction mechanism in object-oriented programming. To provide extraction or inference mechanisms that support data abstractions, a proper specification mechanism must be adopted. Most popular ones, like JML, which rely on method pre and postconditions, cannot easily handle data abstractions. In fact, such methodologies use some kind of abstract model to specify the behavior of the data abstraction, such as mathematical sets, arrays, lists, etc. Such abstract models play the role of an abstract implementation. Although abstract models differ from the implementation, they may introduce an implementation bias in the specifications, and they somewhat lower the level of abstraction. In fact, on the one side, we would like specifications to abstract away from the implementation details of data objects and just express their “externally observable” behavior.

On the other, such behavior depends on the internal state of the object.

This issue also emerges when addressing the specification inference problem casted to data abstraction as source artifacts. In fact, adopting a specification language that uses abstract implementations might introduce further effort by considering all the possible combinations of models that could be used to explain the behavior of the data abstraction.

This thesis mainly faces the problem of specification inference for data abstractions by providing three main contributions:

- A Behavioral Specification Language for data abstractions, called Intensional Behavior Models, based on a fully black-box model which leverages a trace representation of possible object instances; the specification method also includes an incremental approach in writing specifications starting from partial models to full-fledged specifications;
- a Synthesis Algorithm for the proposed specification language, taking two possible artifacts as source: either a black-box implementation of a data abstraction or a specification

based on contracts à la JML; a validation approach, which leverages an adaptation of the small scope hypothesis to verify that intensional behavior models are coherent with properties expressed as algebraic specifications.

The proposed specification method is based on an incremental approach which first outlines a partial behavior model of the artifact to be specified and then proceeds to generalize it to a full-fledged specification. The partial models considered are called Behavioral Equivalence Models (BEMs). A BEM provides a way to describe the behavior of classes of behaviorally equivalent instances for a given data abstraction. Essentially, a BEM is a finite state machine in which each transition represents a method invocation and each state is labeled with observer return values; moreover, each state represents a set of behaviorally equivalent instances.

The main motivation of outlining a partial model like a BEM is a variation of the small scope hypothesis. In this sense, we consider the example behaviors as described by a BEM as relevant, that is, the intended overall specification of the data abstraction is simply a generalization of the BEM. Both normal and exceptional behaviors should be described by example in the BEM. A natural generalization of a BEM, which is a labeled finite state machine, is a formalism which could intensionally generate all the possible valid BEMs for the data abstraction to be specified. The Intensional Behavior Model

is exactly such kind of formalism. Because an automaton can be effectively described as a graph, an intensional behavior model is a graph grammar that can produce all the possible BEMs.

Another contribution of this thesis is the SPY approach for specification recovery, which addresses the problem in the context of components that behave as data abstractions. The SPY approach considers components as black boxes and uses dynamic analysis to extract a precise description of their behavior. First, the approach uses automated testing to build a precise finite-state description of the component behavior within a limited and small scope, in terms of method parameters and length of test cases. This intermediate description is a BEM, and it is then generalized to an intensional behavior model which describes every possible behavior of the component even outside the observed scope. This artifact can be viewed as a likely specification of the data abstraction. The generalization is essentially obtained by identifying canonical method sequences describing every possible state of the component together with well known invariant detection techniques.

Later, we analyzed the possibility of deriving behavioral equivalence models from contracts. Contracts are usually expressed in first-order logic and use some abstract implementation to describe class invariants and pre-post condition of methods. This formulation usually hardens the possibility of verifying externally observable properties, such as the ones which can be expressed with algebraic

specifications. Thus, behavioral equivalence models derived from contracts may help this kind of validation analysis. We use the JForge Specification Language (JFSL) as a source language for contracts, and we first derive a BEM for the specification within a small scope, by essentially grouping behaviorally equivalent representations and deriving abstraction functions.

The absence of a universally accepted specification methodology determines also the relevant problem of specification consistency. Whenever different formal descriptions are given for the same software artifact, or whenever they are extracted with specification recovery tools, it can be relevant to determine their consistency. We addressed this issue in a particular instantiation, whose motivation derived from the problem of comparing specifications recovered with SPY with competing approaches. In practice, we propose a model checking based technique which determines the consistency of intensional behavior models with algebraic specifications. However, the approach has relevant application even outside the context of specification recovery. In fact, axioms of an algebraic specification can be easily considered as behavioral properties - that is, properties expressed only in terms of externally observable behavior - and can be used to validate BEMs alone or BEMs derived from intensional behavior models.



## INTERFERENCE MANAGEMENT IN MULTI-USER MULTI-CELL SYSTEMS

**Daniele Molteni**

In this Ph.D. dissertation we present the study of interference management techniques and resource allocation strategies in multi-user multi-cell wireless systems. Performance of current multi-cell systems is heavily limited by the available bandwidth, complexity constraints of commercial devices and impairments affecting the radio environment. Increasing interference level is one of the bottleneck for the expansion of wireless networks and a severe limitation for throughput improvements. Interference problem is foreseen to become more significant year by year both in new 4G technologies such as in IEEE WiMAX and 3GPP LTE systems as in old infrastructures such as GSM networks.

In this dissertation we focus on two transmission paradigms: Time Division Multiple Access (TDMA) and Orthogonal Frequency Division Multiple Access (OFDMA), respectively in Part I and Part II of this work. The thesis provides contributions focusing on two aspects of the considered technologies: physical (PHY) layer and radio resource management (RRM) algorithms design. For the OFDMA system, different multi-user access policies are studied and their performance are evaluated through a novel analytical framework. The RRM design

problem is then addressed by proposing a channel-aware scheduling algorithm developed to reduce the interference generated by the multi-cell system. Considering TDMA, a multi-user access approach is accounted following the idea provided by the Orthogonal Sub Channel (OSC) included in the latest GSM/EDGE standard release.

An optimized uplink receiver is proposed for the interference limited scenario and its performance are then employed to define an optimized scheduling algorithm developed adopting the guidelines also used in the OFDMA case. In particular, considering a general MIMO-OFDMA system, we propose an analytical framework where the main resource allocation policies, namely randomization and coordination approach, are evaluated. Performance assessment is necessary to provide to the RRM reliable models for the system level optimization. The wide range of possible propagation/interference scenarios and cellular deployments makes this evaluation a very complex task. Analytical models for physical-layer performance (even if for simplified settings) are relevant to reduce the computational burden of extensive simulations when defining simplified

strategies. We propose a methodology that provides the PHY layer performance in terms of bit error rate (BER) at the output of the FEC decoder. In order to be useful and practical, the analytical framework must account for the main systems settings such as multi-cell scenarios with realistic channel/interference models subject to space-time dispersive fading. Specifically, the proposed bound is derived accounting for bit interleaved convolutionally coded transmissions and different multi-antenna techniques ranging from beamforming to OSTBC coded systems. Furthermore, the analytical approach is designed to take into account non-stationary intercell interference generated by the fluctuations of the traffic load over the multicell system. Finally the described bound is cast to assess the performance of realistic 4G systems, in particular for WiMAX (IEEE 802.16d-e) and 3GPP LTE.

We deal with the RRM protocol design of the considered MIMO-OFDMA system providing an optimized scheduling algorithm. Here the problem of resource allocation of radio resources is discussed. The bandwidth is divided into logical subchannels (i.e., group of subcarriers over a frame) and the resource allocation aims at assigning the subchannels to the users

following an optimality criterion. The optimum is defined as the minimum transmitted sum power used by the BS to serve all the users in the cell. The algorithm is centralized and carried out at a cell-level, each BS is supposed to perform the scheduling locally without any kind of coordination among neighboring BSs. On the other hand all the MSs are supposed to communicate through a feedback channel the state of the DL connection with their own BS. The information are related to the SINRs of all (or a subset) the logical subchannels. We account for a multi-cell environment where the allocation strategies adopted in each cell directly influence the inter-cell interference. The resource allocation methodology provide the best association between the subchannels and the MSs and the relative transmitting powers constrained to the service requirements characterizing the link to each user. The main idea is to choose the best combinations of users and radio resources as to minimize the transmitted power, to do so, we design the algorithm around a combinatorial optimization between users and available resources. Here a framework based on game theory has been modeled to evaluate the effectiveness of the proposed resource allocation approach, studying the conditions and the probability of a Nash equilibrium in a simplified two-cell environment. Numerical results show a significant gain with respect to common multi-user access randomization strategy. Accounting for the TDMA systems, we address the analysis

of a novel multi-user access strategy: the OSC multiplexing technique. The OSC has been designed with the aim of doubling the GSM cell capacity multiplexing two users on the same radio resource. In the uplink (UL), two MSs transmit over the same subcarrier fully interfering each other. In the downlink (DL), the base station (BS) transmits two overlapped GMSK-modulated streams with a 90deg phase offset exploiting a QPSK-like modulator. For both UL and DL, the detection of the desired stream must rely on receivers with interference management capabilities. The Successive Interference Cancellation (SIC) scheme is the decoding strategy suggested by 3GPP, while a family of interference rejection receivers such as, for instance, the Single/Double Antenna Interference Cancellation (SAIC/DAIC) algorithm can be employed in DL. We design a new two-stage UL receiver: the Joint OSC Receiver (JOR) which combines a front-end prefiltering stage for the mitigation of the out-of-cell interference, with a MUD that optimally handles the mutual interference between the users that access to the same resource. Differently from conventional SAIC/DAIC, here prefilters are optimized jointly for the two OSC users, so as to minimize the interference from neighboring cells while preserving the two multiplexed signals. The second stage is a MUD that jointly estimates the two transmitted sequences: joint maximum likelihood sequence estimation (JMLSE) is considered as optimal non-linear solution, while linear detection (L-MUD) is proposed as a good trade-off

between performance and computational complexity. The JOR is here employed for the definition of a RRM scheduling algorithm which provides optimal pairing and power allocation in a GSM-OSC cellular system. The guidelines used for the OFDMA RRM algorithm are here employed and adapted for the new scenario. The algorithm is centralized and deals with the optimization of the pairing of the users and logical OSC channels. The optimization criterion is the minimization of the transmitting powers constrained to service requirements defined for each user as maximum BER for a voice quality target. Fairness is required by the OSC approach to guarantee reliable voice transmissions to both users sharing the radio channel. The proposed algorithm is designed for both UL and DL exploiting for each user some channel state information such as path-loss and interference level. Moreover, for the DL, the latest Adaptive QPSK (AQPSK) feature is adopted in the optimization to further reduce the overall transmitting power. The reduction of the required transmitting power directly affects the lifetime of the devices and the overall interference generated on neighboring cells, thus providing benefits to the whole multi-cell environment.

# CONTEXT BASED QUERYING OF DYNAMIC AND HETEROGENEOUS INFORMATION SOURCES

Giorgio Orsi

The proliferation of freely-accessible data-intensive websites, the growing availability of pervasive and mobile applications, as well as initiatives for open-accessible linked data in the Web, provided the users with many sources of potential information. This information is “potential” because it is hidden in the data stored into data sources typically distributed, hybrid and structured in accordance to heterogeneous data models. As a consequence, the added value we can get from such data highly depends on our capability to access and interpret them. In addition, data require reshaping and integration in order to be adapted to other contexts of use than the ones they were envisioned for. From a more practical point of view, these data represent potential new business opportunities for industry, while their processing and management is a rich research field for academics.

Moreover, in the recent years the pervasive and mobile settings have grown in importance, adding to data-management requirements a degree of dynamicity that urges a re-thinking of the current techniques, to comply with this new open-world and open-technology setting. Dynamic, on-the-fly, pay-as-you-go

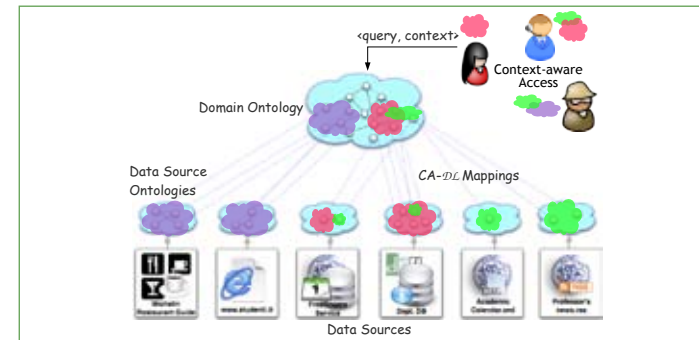
integration of heterogeneous data sources are especially useful in a mobile setting. However, in these scenarios the reduced amount of power and available memory re-proposes urgent scalability requirements even in those situations that are no longer considered problematic for enterprise information systems.

The ability to structure and integrate data does not necessarily correspond to the ability to interpret them correctly to obtain valuable information and even less it implies the ability to do something useful with it, in fact:

- information is often noisy, and should be appropriately focused to the current application situation.
  - information is, in general, too much to be examined manually, and should be appropriately reduced and shaped before presenting it to the user.
  - the mobile setting poses peculiar constraints that must be dealt with on a case by case basis.
- One of the possible means to make the answers precisely fit the current user's needs is represented by context-aware information reduction, which uses context meta-data to reduce and personalize the information space in order to improve their processing and, eventually, their interpretation.

This thesis presents the concept of Nomadic Data Integration Systems (N-DIS) and discusses context-aware tailoring and integration techniques for querying heterogeneous and dynamic data sources, using ontologies as a means to represent the available data and the contexts in which the user and the system might be required to operate. Figure 1 shows an abstract architecture of an N-DIS. In such a scenario, Semantic Web languages such as RDF(S) and OWL, which might fail on large scale data integration tasks, can play an important role as tools for on-the-fly integration of small pieces of information belonging to different, independent and possibly heterogeneous data sources.

In particular, we are concerned with the problem of query answering in ontology-based context-aware and dynamic data integration systems, where queries are specified over an application-domain ontology which acts as a global representation of the information space, and then contextualized and distributed towards a set of data-source ontologies automatically extracted from the data sources, following automatically-created mappings between the domain and the data-source ontologies. The problem of query answering



1. Architecture of a N-DIS

in data-integration systems has already been investigated for years under the names of Query rewriting/reformulation. Similar techniques have been developed for consistent query answering using data-intensive ontologies as well as in for highly expressive combinations of description logics and rules. In this work we apply some of the results coming from these theories to an evolving scenario, composed by transient (e.g., mobile) heterogeneous data sources whose logical structure may change over time with the additional constraint of the contextualization of queries and, as a consequence, of their answers.

In our setting the interaction between the data sources and the corresponding ontologies is taken care of by domain aware wrappers that produce an ontological representation of a data source taking into account the domain ontology. The query

issued by the user is transformed into a query on contextual views over the domain ontology, and then rewritten in terms of the data-source ontologies.

The general effect of this approach is a reduction of the amount of data that will be retrieved from the data sources, increasing the focus of the results and possibly reducing the number of involved data sources. We consider ontologies expressed using CA-DL, a DL-based language extended with constraints that allows us to address the problem of contextual query answering by reducing it to the problem of query answering under constraints.

The possible contexts envisaged in the application, and used to reduce the information space, are specified at design time by means of a general context model that extends the one of [2] and is also represented in CA-DL.

## SECURITY AND REPUTATION-BASED SYSTEMS FOR WIRELESS MESH NETWORKS

Stefano Paris

Wireless Mesh Networks (WMNs) have recently been accepted as an effective means to provide broadband wireless connectivity without the need of a costly wired network infrastructure. The network nodes in WMNs, named mesh routers, provide access to mobile users, like access points in wireless local area networks, and they relay information hop by hop, like routers, using the wireless medium. Therefore, mesh routers form a self-organized and self-configured backbone network by collaborating in the execution of management and control operations. The nodes in the backbone network use the IEEE 802.11 standard as wireless technology to establish the radio links and maintain the mesh connectivity among themselves via routing protocols borrowed from the Mobile Ad-Hoc Networks (MANET) paradigm, such as DSR, AODV, and OLSR. Unlike MANETs, mesh routers are usually fixed and do not have energy constraints. Moreover, the gateway functionalities performed by a subset of mesh routers enable the integration of WMNs with several existing technologies, like cellular systems, Wireless Sensor Networks and WiMAX. The flexibility provided by the wireless technology has fostered the development of new communication paradigms,

like Wireless Mesh Community Networks (WMCNs), where devices owned and managed by different service providers or individuals collaborate to extend the network coverage. Mesh router owners can connect to the backbone network with their wireless devices, whereas customers can only access the WMCN services through mesh routers. Community users and customers may be charged different fees to access WMCN services. As a consequence, these services must satisfy Quality of Service requirements, and penalties can be envisaged if QoS requirements are violated. Recent research literature has focused on designing new and more efficient communication protocols to satisfy the increasing bandwidth demands and tighter quality of service requirements of user applications. Although security represents a primary concern for any customer that wants to subscribe to reliable services, little attention has been devoted so far to security problems introduced by these new communication paradigms. Due to the open and shared nature of the wireless technology and the multi-hop communication paradigm, WMNs are more vulnerable to attacks. In this thesis, we systematically analyze the vulnerabilities of WMNs and WMCNs, and propose solutions to cope with

the security threats we have identified. Outsider adversaries can easily perform a wide variety of attacks, such as eavesdropping, jamming, man-in-the-middle and spoofing. On the other hand, the lack of an authentication and authorization system that limits the access only to authorized mesh nodes may result in internal attacks through which adversary nodes can steal sensible information of users or seriously affect the network operation. Furthermore, in the design of WMNs, special attention must be devoted to the protection of the integrity and authenticity of the control information exchanged by mesh routers, like management and routing messages. This thesis addresses the aforementioned security issues by designing two security architectures tailored for WMNs. The former architecture, called MobiSEC, assigns to a mesh router the role of authentication and key management server. All nodes periodically receive from this node the information necessary to generate the sequence of cryptographic keys used to protect the data and control messages transmitted over the backbone. While representing an effective security solution for WMNs, MobiSEC, like all centralized solutions, is characterized by a single point of failure that can be exploited by

adversaries to attack the network. We therefore design a distributed architecture, named DSA-Mesh, that improves the robustness of the entire system by increasing the number of mesh routers that are liable for the authentication and key management services. Regrettably, in some network scenarios the protocols of DSA-Mesh exhibit a greater latency than those of MobiSEC, due to their distributed nature that requires the collaboration of several network nodes to complete the authentication and key distribution tasks. To increase the responsiveness of DSA-Mesh, we propose an Integer Linear Programming (ILP) model to select the mesh routers that minimize the overall latency of the distributed protocols. We show that a careful planning of the roles assigned to mesh routers can increase the responsiveness of DSA-Mesh, and distributed architectures that adopt similar collaborative approaches. Due to their collaborative nature, WMCNs are prone to insider attacks performed by adversaries whose behavior is mainly driven by selfish interests. In fact, all community users have two opposed interests: on the one hand, they compete against the customer devices which they serve for the available bandwidth provided by mesh routers, since they have to share the capacity of their outgoing wireless links

established with other mesh routers of the WMCN. On the other hand, mesh routers have an incentive to serve fairly a large number of users, since it can be realistically assumed that they are rewarded by the mesh community network considering both the number of served customers and their satisfaction. A selfish user that provides connectivity through his own mesh routers to other network nodes might try to greedily consume the available bandwidth by favoring his own traffic while selectively dropping others, setting firewall rules on their devices to drop almost all packets sent by other participants or customer stations, or limit the maximum transmission rates available to the served devices. Such misbehavior can lead to severe unfairness and severe performance degradation, since periodic packet dropping at relaying nodes can decrease the throughput of closed loop connections (such as TCP) established by other nodes, even when the fraction of dropped packets is small. We investigate systematically the impact of the uncooperative behavior of selfish mesh routers in the routing process. Specifically, we evaluate the performance of OLSR, the most widespread routing protocol used in WMCNs, when the path selection is driven by data-link layer metrics

that capture only the quality of wireless links, like ETX (Expected Transmission Counter). Our results show that in the presence of even a low percentage of adversary nodes data connections experience severe throughput degradation and unfairness. To address the the problem of reliable routing in the presence of selfish participants, we propose a novel cross-layer metric, EFW (Expected Forwarding Counter), and two further refinements that combine information across the data-link and network layers to select the most reliable and highperformance network paths. The results obtained through simulations and experiments on real-life testbeds show that our cross-layer metric and its refinements accurately capture the path reliability, considerably increasing the WMN performance. Finally, we present a detection framework based on a novel and effective reputation model to identify the insider adversaries that disseminate false information about the relaying behavior of other mesh routers to affect the routing decisions and favor their colluding nodes. Numerical results show that the proposed mechanism provides high detection accuracy, even when a high percentage of mesh devices misbehave.

## A GOAL-ORIENTED METHODOLOGY FOR SELF-SUPERVISED SERVICE COMPOSITIONS

Liliana Pasquale

A key requisite for modern enterprises is offering business processes able to accommodate the changing requirements of its stakeholders dynamically.

In this context, the adoption of flexible and easy-to-integrate technologies, like Service Oriented Architecture (SOA), plays a fundamental role in the provisioning, maintenance and evolution of complex business processes.

SOA is an architectural paradigm that fosters the provision of complex functionality by assembling disparate services whose ownership and evolution is often distributed. The architecture of these applications, in general, expressed as a BPEL process, does not provide a single integrated entity, but it is a composition of third party services (partner services) that interact in a loosely coupled way. This way of working fosters reusability, since applications are built by gluing together existing components (i.e., services), and reduces the development costs since applications do not need to be built from scratch. SOA also offers flexible and agile solutions since new/changing requirements can be simply handled by adding, removing or substituting the partner services in order to obtain completely different solutions. Service compositions can provide and maintain business processes

by representing and executing business tasks through a set of BPEL activities that can be modified or restructured to accommodate new/changing business needs.

Despite their wide adoption, services compositions are intrinsically unreliable, as loose coupling and distributed ownership allow their service components to evolve separately. Unexpected failures can take place at runtime due to physical faults (e.g., malfunctions in the network medium, or failures in the middleware of the hosting servers), development faults (e.g., parameters incompatibility, or unexpected changes of the interface exposed by service components), or interaction faults (e.g., partner services can be down or unavailable, or they violate some SLA terms). For this reason, service compositions continuously need to self-adapt to satisfy the stakeholder's requirements, while overcoming failures that can take place due to the nature of their technology.

So far, service compositions have been mainly addressed from a technological perspective: several proposals support the design and the reliable execution of these applications. BPEL is the de-facto standard for the definition of service compositions, and several BPEL engines are available for

the execution of BPEL processes (e.g., ActiveBPEL, ApacheODE, or Oracle BPEL Process Management). Considerable research effort has also focused on making service compositions more reliable: from dynamic verification and validation of service-based systems, to feedback-control loops that integrate analysis and adaptation activities, to conceptual models for service-centric systems, just to mention a few. Besides such technological achievements, further steps are still necessary to ensure that service compositions enact business processes in a coherent and effective way (i.e., according to their requirements), and support their evolution.

This objective poses challenges at different levels of abstraction. At requirements level, we must provide suitable models to represent requirements, along with their potential changes. Available models and technologies are not able to completely address the issues discussed above. Goal models and scenario-based solutions do not provide explicit support for modeling uncertainty, nor for adaptation. Moreover, common business process notations, like BPMN (Business Process Modeling Notation), are not expressive enough to embed requirements into the business process, trace them, and account

for user-oriented reliability. The business dimension of processes imposes frequently changing requirements, while their distributed nature makes them intrinsically unreliable: this means that requirements must state the functionality of the system to be, its quality of service, and also self-adaptation capabilities (i.e., how to cope with changes in both the requirements and the execution environment).

At application level, we need methods and tools to precisely relate requirements to services and service compositions. This allows us to assess if requirements continue to be satisfied at runtime and coherently support system evolution. Unfortunately, the BPEL language lacks specific constructs for relating higher-level requirements to process activities, making it difficult to assess if a composition effectively achieves or maintains underlying business needs. It is also hard to accommodate new/changing business needs with the running process, because this requires identification of process activities that must be modified. Moreover, existing methodologies to provide self-adaptive services, like service supervision (that includes data collection, monitoring and recovery activities), do not provide any guarantee that they are

effective and they really avoid requirements violations. To this aim, adaptation must be properly configured, that is, we must identify what, when, and how it must be performed depending on the business needs and the supervision features available in the execution infrastructure.

Finally, at the infrastructure level, requirements must be constantly assessed and self-adaptation capabilities be activated when needed to keep the execution on track and give the user the perception of a reliable and trustworthy solution.

This thesis addresses the problems described above at the methodological and technological level. First, we propose a goal model to represent requirements of the system together with self-adaptation capabilities. We conceive requirements as live entities that can evolve at runtime to coherently trace the actual business needs. We add more flexibility to the model by introducing fuzzy goals that trace the satisfaction level of non-functional requirements and guide the selection of the adaptation actions taken at runtime according to their satisfaction degree.

Second, we support requirements traceability by linking the goal model to both a functional and

a supervision model of the service composition. The former represents the set of compositions able to satisfy stated requirements. The latter defines how to assess the requirements of interest and keep the application on track (i.e., how to add reliability to service compositions).

Finally, we provide a flexible runtime infrastructure that supports the controlled execution of service compositions. The infrastructure adopts suitable engines, along with data collectors, monitors and adaptors to enact the directives defined in the supervision model.

## DESIGN METHODOLOGIES FOR IMPROVING EMBEDDED SYSTEMS WITH HARDWARE ACCELERATORS

**Christian Pilato**

Heterogeneous multiprocessor architectures are the de-facto standard for embedded system design. Today, to accelerate the different parts of the applications, they are usually composed of several general purpose, digital signal and hardware accelerators (e.g., Field Programmable Gate Arrays - FPGAs or Application Specific Integrated Circuits), interconnected through various communication mechanisms. When developing an application on such embedded systems, the designer has to determine when (*scheduling*) and where (*mapping*) the groups of operations (i.e., the tasks) and the data transfers (i.e., the communications) should be executed, depending on a set of constraints and dependences, in order to optimize some design metrics, e.g., the program execution time.

Moreover, custom hardware accelerators are known to outperform the corresponding software solutions by different orders of magnitude, but the resources for their implementation are usually limited in the final architecture. Unfortunately, even if the synthesis of behavioral specifications has been a hot topic for research for the last two decades and significant achievements

have been obtained, existing methodologies are usually able to generate only one-shot high-quality solutions. For this reason, the design space exploration has to be usually manually performed, for example, by constraining the number of resources. Indeed, when implementing a single task, the designer cannot know in advance which is the best hardware solution, but he or she should be able to explore them at different level of abstractions: the proper RTL architectures have to be identified and optimized at logical level and, then, the different high-quality Pareto-optimal solutions should be properly selected and combined at system-level, based on the requirements of the application and the constraints of the target architecture.

Given these motivations, this thesis aims at developing a methodology extending PandA, an existing framework for hardware/software co-design of embedded systems, to generate and optimize multiple hardware implementations, and combine them with the existing software ones in order to improve the overall performance of the application.

The proposed methodology starts from a partitioned application (i.e., a task graph) and a model of the

heterogeneous target architecture. Additional information about the software implementations, such as the execution time of each task on each processing element, can be also provided. Then, it generates multiple hardware implementations for each task. In particular, starting from a multi-objective genetic algorithm, a framework for design space exploration for high-level synthesis has been developed for generating different Pareto-optimal solutions. An additional logical optimization can be also performed on these resulting descriptions to allow the designer to further enhance the performance of the applications. Indeed, this enables to fit faster solutions into the available area since they now require fewer resources for their implementation.

Finally, the initial description of the partitioned application is analyzed with a novel algorithm, based on Ant Colony Optimization, for mapping and scheduling on the given target platform. In particular, it optimizes the execution time of the entire application by identifying the proper combination of hardware and software implementations and the proper order of execution of the different tasks and communications, respecting

all the constraints imposed by the target architecture (e.g., the limited area for hardware accelerators). All proposed solutions have been extensively evaluated against state-of-the-art solutions and significant improvements have been obtained with respect to them. In particular, the logical optimization is able to reduce both area occupation and timing after the place and route by more than 10% in average. The design space exploration performed at architectural and system levels is then able to obtain application implementations 16% better in average compared with the ones obtained with traditional methods.

## SCHEDULING OF APPLICATIONS ON DYNAMICALLY RECONFIGURABLE ARCHITECTURES

**Francesco Redaelli**

The goal of this thesis is to provide a complete study of the problem of scheduling multithreaded applications onto partially dynamically reconfigurable architectures. The work has been divided in two steps. The first part considers the offline problem: given the target architecture, and given the applications that need to be executed on it represented as task graphs, find the shortest possible schedule. We modeled the problem as a ILP formulation to show the need of a reconfiguration-aware scheduler. Given these optimal results, and the knowledge gained while developing this model, a heuristic scheduler has been proposed. It exploits in a very effective way all the features that reconfiguration provides: module reuse, configuration prefetching, and anti-fragmentation techniques. A model and a heuristic scheduler have been developed for a 1D reconfiguration scenario, for a 2D reconfiguration scenario, and for a 2D HW/SW reconfiguration scenario. The results obtained show how reconfiguration aware scheduling policies can overwhelm classical state of the art algorithms, gaining about 18% in terms of schedule length. The second step of this work has been to develop a simulation engine able to perform design

space exploration of the scheduling policy. Given an architecture and given some online scheduling policies, our framework is able to show which one best fits the target architecture. On the other hand given a scheduling policy, and a set of architectures to be chosen, our framework shows which architecture best fits that particular policy. We proposed also three different architectures and provided the three corresponding online scheduling algorithms that achieve the best performance results. Compared with state of the art approaches, our algorithms achieve better performance in terms of schedule length, when running on the appropriate architecture. Finally, we developed the first prototype of a reconfiguration aware user space scheduler that it is able to exploit all the features offered by a reconfigurable architecture. This scheduler works on a linux operating system. It is different from the scheduling approaches presented in the literature because it allows an out of order execution and reconfiguration for those tasks that are selected to be placed as hardware functionalities.

### Innovative Contributions

The main contributions of this thesis are described in the following. A ILP model of a static scheduler over a 2D partially

dynamically reconfigurable HW/SW architecture has been developed. A heuristic static scheduler that provides good scheduling results for large applications. Its results can be used as baseline schedule by the online one. This heuristic scheduler has been realized starting from the knowledge acquired during the development of the ILP model. An online simulation engine that implements different scheduling policies tailored to the underlying architecture below has been developed. This scheduler is able to perform design space exploration in order to decide which policy is better for a particular system architecture. Also, its results can be used as baseline schedule by the online one. A user space online scheduler has been developed. It has been built on a linux operating system and it is able to schedule multithreaded applications that ask for hardware functionalities.

## HIERARCHICAL AND DISTRIBUTED CONTROL METHODS FOR LARGE SCALE SYSTEMS

**Carlo Romani**

This work deals with large scale systems that can be viewed as composed by several subsystems, such as manufacturing systems, process plants and networked systems. The control design of these large scale systems with a centralized scheme can be difficult due to the computational complexity, to robustness and reliability problems and to communication issues. This facts motivate the development of distributed and multi-level control structures and the increasing interest in this field.

The thesis aims at providing a contribution to the development of design tools of distributed and multi-level control structures for large scale system, focusing the attention on the formulation of the control problem and on the design of a controller based on the Model Predictive Control (MPC) approach, that provides stability and robustness. This work also presents a distributed estimation algorithm for nonlinear systems, which can be used to estimate the states of the system, usually assumed to be known in the design of the MPC algorithm. The novel distributed Moving Horizon Estimation (MHE) scheme can be applied to large scale systems characterized by a non linear model and composed by several subsystems with non overlapping states.

Finally an example is described, reporting the modellization of a part of a Hydro Power Valley (HPV), the design of a controller and the application of the MHE algorithm.

# MODELING AND MONITORING OIL AND GAS RESERVOIRS WITH SAR INTERFEROMETRY

Alessio Rucci

Surface deformation monitoring provides unique data for observing and measuring the performance of hydrocarbon reservoirs, for Enhanced Oil Recovery (EOR) and for Carbon Dioxide Capture and Sequestration (CCS). To this aim, radar interferometry (InSAR) and, in particular, multi-interferogram Permanent Scatterer (PS) techniques have already proven to be valuable and cost-effective tools. For oil/gas reservoir management, PS measurements allow one to examine the temporal and spatial pattern of long-term reservoir response to pumping and extraction highlighting possible compaction affecting the area surrounding the reservoir and generating potential damages to local infrastructures. Apart from the environmental impact of subsidence and uplift phenomena induced by reservoir exploitation, recent optimisation techniques ask for timely information about many geophysical parameters, both downhole and on the surface. In fact, depending on reservoir characteristics and depth, water, oil or gas production can induce surface subsidence or, in the case of CCS, ground heave, potentially triggering fault reactivation and in some cases threatening well integrity. Mapping the surface effects of pressure variations at the reservoir layer, due to either fluid extraction

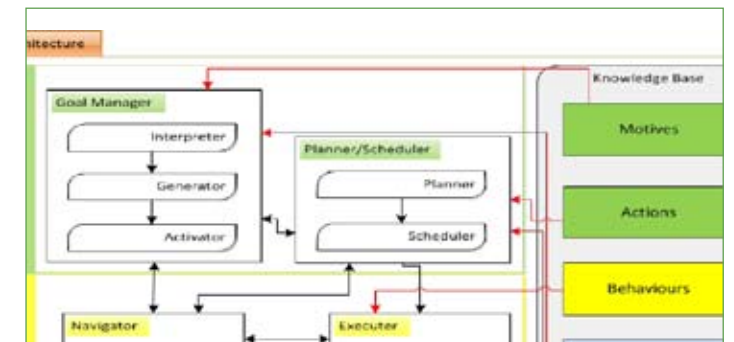
or injection, usually requires the availability of hundreds of measurement points per square km with millimeter-level precision, which is time consuming and expensive to obtain using traditional monitoring techniques, but can be readily obtained with InSAR data. Moreover, more advanced InSAR techniques (PSInSAR) developed in the last decade are capable of providing millimeter precision, comparable to optical leveling, and a high spatial density of displacement measurements, over long periods of time without need of installing equipment or otherwise accessing the study area. In this work I present examples of applications for reservoir monitoring and modeling with the aim of highlighting the importance of providing more and more accurate displacement measurements and the necessity to increase the density of measurement points in order to better constrain the reservoir parameters inversion. In chapter 1 I propose an asymptotic formulation to use surface deformation for a reservoir characterization in terms of volume/pressure changes and permeability. In this chapter is explained how accurate and spatially dense surface deformation measurements are necessary to better constrain the inversion for geophysical parameters.

To ensure a physical meaning for the estimated volume/pressure changes, sign constraints are introduced according as we are considering a production or injection area. Due to the no more linear behavior, traditional technique for the assessment of model parameter estimates are not necessarily accurate. Thus, in chapter 2 I propose three different approaches to infer a posteriori variance of the estimates when inequality constraints are included in a linear problem, which is the case of the problem to be solved in chapter 1. In chapter 3 I describe the basic principles of the standard PSInSAR algorithm explaining its advantages and the main drawback. The main limit is that the technique struggles to extract information outside urban areas which is the case for many oil/gas fields. That's why a new algorithm is introduced in chapters 4-6 to optimally process non urban areas, characterized by the presence of Distributed Scatterers. As DS are characterized by lower SNR with respect to Permanent Scatterers I chapter 4 I introduce , first of all, an adaptive filter in the space-time domain (Phase Triangulation Algorithm - PTA). The aim of the filter is to preserve the point-wise radar response of PS while filtering the DS interferometric phases.

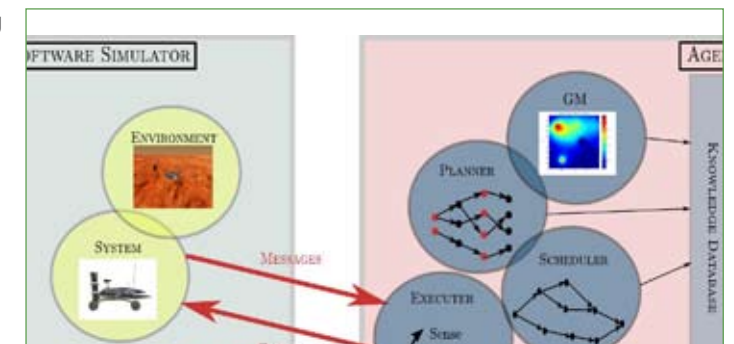
If are interested in the average displacement rate, instead of a displacement time-series, is possible to increase moreover the density of measurement points using a Maximum Likelihood estimator of DEM and velocity for DS. This estimator is described in detail in chapter 5 with a criteria to select reliable estimates. Once DS have been properly processed In chapter 6 I provide a rigorous approach to jointly analyze PS and DS, with the aim of increasing the spatial density of points in correspondence of which I can provide an accurate displacement time-series.

In chapter 7, the new algorithm for InSAR data processing is applied to measure surface displacements over the Krechba field in Algeria, involved in one of the largest project in the world of Carbon Capture Sequestration (CCS).

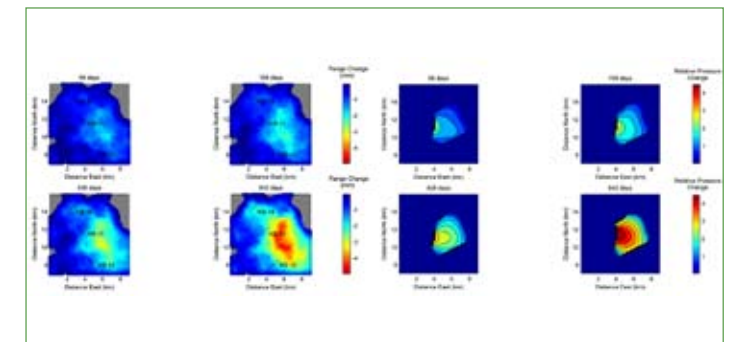
It is shown how surface data can be used to monitor CO2 sequestration activity, both to track the injected CO2 and to monitor possible fault reactivation due to induced high pressure variations. Moreover surface displacement measurements can be used to estimate the pressure variations and permeability of a producing gas reservoir.



1. Comparison between an interferogram filtered only spatially and filtered applying the PTA algorithm



2. Comparison between the average displacement map obtained applying the standard PSInSAR and the new algorithm proposed



3. On the left panel four snap-shots of surface displacement, occurred over an oil producing field, used to estimate the pressure changes at the reservoir layer.



# PHASE PRESERVING SYNTHETIC APERTURE SONAR FOCUSING AND MULTI TEMPORAL INTERFEROMETRY

**Silvia Scirpoli**

In the framework of sustainable development activities carried out by Eni Exploration & Production division, the Eni Geological Service in association with Eni R&D department has promoted and financed the development and the experimentation of an innovative technology for environmental monitoring. Experimental activities were conducted in the context of a two-year project with the aim to test innovative approaches to better assess the altimetric accuracy and variations of the seabed. The selected methodology is based on very high resolution acoustic imaging by repeated surveys with a sophisticated high-frequency sonar mounted on an underwater autonomous vehicle (AUV).

A Synthetic Aperture Sonar (SAS) is an acoustic imaging system that provides high resolution images combining the data collected along a virtual array of receivers. The virtual array is synthesized by the platform motion (an Underwater Autonomous Vehicle like the one shown in Fig. 1). Repeat-track interferometry is a well known technique used in Synthetic Aperture Radar (SAR) to obtain precise measurements of altimetric variations. In principle, the same technique can be exploited with Synthetic Aperture Sonar for seabed deformations. However

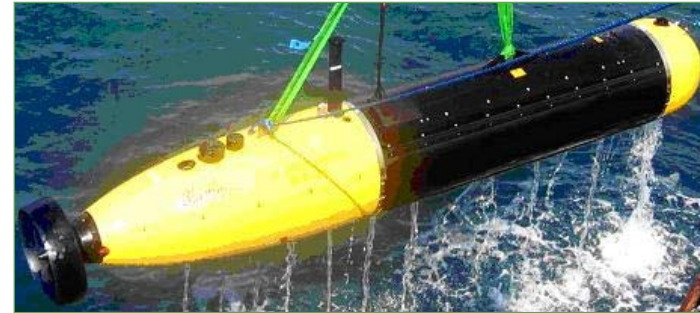
exploiting sonar data to obtain measurements of altimetric variations through interferometry, poses many more problems with respect to radar. In fact acoustic waves speed shows stronger and faster variations than electromagnetic waves; the actual platform trajectory is not known as precisely and the statistics of natural targets coherence and its behaviour with time do not exist.

This thesis presents the processing developed for the elaboration of the data and the results obtained applying this processing on the real SAS dataset collected during this 2 year study led by Eni. The sonar echoes were gathered in several sea campaigns in the Tyrrhenian Sea near La Spezia (Italy).

The processing developed mainly consists in a phase preserving, full-bandwidth SAS focusing technique (including motion compensation and auto-focusing) developed for the 300 KHz SAS system operated by NURC (Nato Undersea Research Center). The general approach used to focus SAS data consists mainly in two steps: in the first step the echoes of each ping are focused independently, obtaining low resolution focused images of the area; in the second step these focused images are coherently summed together, obtaining the full resolution reflectivity map

of the observed area. Single ping data are focused in the wavenumber domain by means of a modification of the standard Omega-K technique in order to cope with the multi-static acquisition system. On the other hand, in order to sum coherently the data collected during different pings, the actual position and orientation of the platform over time should be compensated: at this purpose the shifts and phase errors due to the irregular motion of the platform are estimated through the interferometric analysis between the focused images of consecutive pings. Results are shown, obtained both on natural and artificial distributed or point targets that are used both as ground references for focusing and for interferometry. The quality of the achieved focused images can be analyzed taking advantage of particular features of the artificial reflectors specifically designed for this experiment.

After the focusing operation the coregistration of the images is discussed. The co-registration is necessary in order to obtain interferograms as it compensates for possible crossings of the trajectories, different orientations on the AUV, variations of the platform velocity, variations of the propagation medium...etc... The images are divided in small blocks and the range and azimuth



1. A photo of the Muscle Autonomous Underwater Vehicle operated by NURC and used to collect the SAS data

offsets of each block, with respect to the master image, are retrieved by means of coherence maximization.

Once the co-registration is done, it is finally possible to obtain the repeat-pass SAS interferograms that are shown throughout the thesis and analyzed showing potentials and limitations of this technique. For this sake a comparison with independent measurements carried out at the same location with the presence of an artificial reflector is illustrated. Interferograms have been illustrated with temporal baselines between 20 minutes and one day.

The statistics of natural target coherence and its behaviour with time were unknown before the experiment described in the thesis. From the achieved interferograms a high coherence of the seabed can be observed,

even with one day of time interval between the acquisitions. This is surely a necessary condition to measure altimetric variations. However, the fringes visible in the interferograms are mainly due to the relative motion of the platform during the two acquisitions. This suggests that this technique is particularly suitable for measuring altimetric variations of the seafloor that are high pass in space. On the other hand measurements of low pass spatial altimetric variations are more problematic since they require a very precise knowledge of the platform trajectory.

The present work has been carried out in the framework of a project leaded and supported by Eni S.p.A. I wish to thank Eni management for allowing the related publications. I wish also to thank the staff of the Nato Undersea Research Center of La Spezia for their precious collaboration.

# DESIGN METHODOLOGIES FOR HYBRID WIRED-WIRELESS MONITORING SYSTEMS

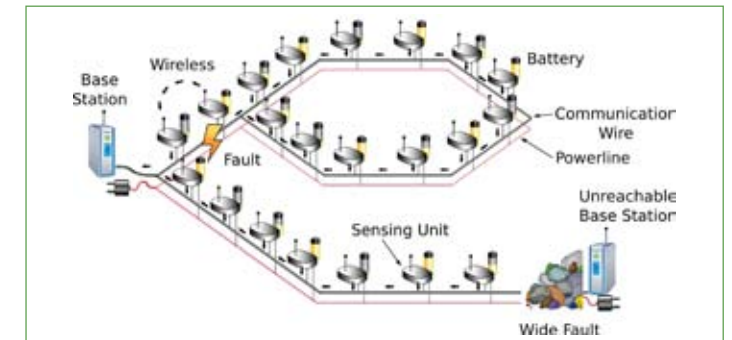
Luigi Sportiello

There are different scenarios where it is necessary to monitor environments, structures, machineries, people or animals through the acquisition of measures that represent their state. The motivations for such an operation can be different, as for instance the identification of faults, incidents or dangerous situations, the detection of subjects or objects in a determinate space, the study of a phenomenon for research purposes. Since in monitoring applications it is typically necessary to gather data from different points of a remote location, e.g., identify fires in a wood, check the status of machineries in a factory, detect the presence of people in a building, systems that are able to carry out a distributed monitoring in the area of interest are necessary. That being so a typical monitoring system is usually composed of different units that are equipped with sensors and are deployed in the monitored area, where they gather data and transmit them towards a base station, which typically stores, shows to users and eventually processes the received information. Depending on the specific nature of the application and its constraints, the communication among the units and from the units to the base station may rely on wired, wireless or hybrid wired-wireless technologies. In

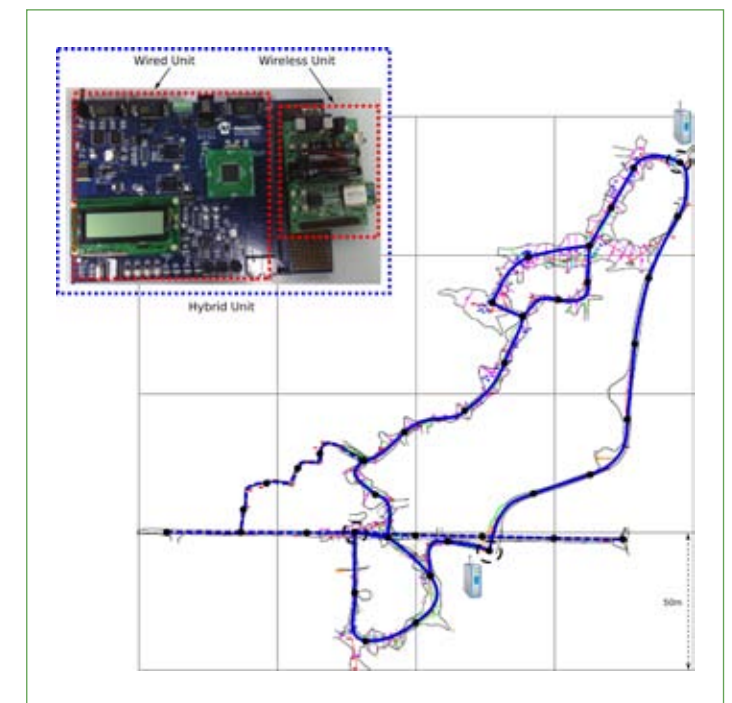
particular we focus on those monitoring scenarios where a data communication based on a wired technology represents the traditional and preferable solution. This is for instance the case of status monitoring for some civil infrastructures like tunnels, mines and pipelines, where the presence of other facilities like aeration and powering systems makes easy and natural the deployment of cables for data transmission. In case of monitoring relative to closed environments, like tunnels and mines, it could be hard or even impossible to rely on efficient energy harvesting solutions, e.g., solar panels, to power the different sensing units, so making mandatory the deployment of powerlines whose presence favors the adoption of communication wires. Anyway for these monitoring scenarios different incidents like fires, explosions and structural collapses have to be taken into consideration, as they may happen and induce faults to the deployed monitoring system. Faults on the communication cables, or even the failure of whole portions of the system, could partially or totally compromise the system activity during incidents, just when the sensed data would be fundamental to manage potential emergency situations. For this reason, the necessity of monitoring systems that are able to deal with faults and

that grant, at least a reduced, continuity of service appears evident. Faults may obviously also affect powerlines, but it can be supposed that mechanisms to detect the power supply decay and make the units able to switch onto local energy accumulators, e.g., batteries, are adopted. Anyway in such a case the energy becomes a scarce resource and so it is necessary to limit the system power consumption. In this work we present the study of a new methodology for the development of robust and efficient monitoring systems that are suitable for such a kind of scenarios. The units are equipped with both a wired and a wireless transmission interface and by default they rely on a wired communication to deliver their data to the base station, using the wireless as backup transmission in case of faults on the communication cables (Fig. 1). In particular we propose an architecture that allows to confine fault effects reducing the switching to wireless only to those portions of the system affected by a failure, thus saving most of the wired communication, which is typically preferable in terms of bandwidth, reliability and energy consumption. The system data delivering relies on a distributed and dynamic routing solution, which even in case of wide failures to whole portions of the system, allows to re-

organize the communication in the remaining working sections saving most of the monitoring service. The developed routing solution is also designed to adaptively identify and select at any time those data delivery paths in the system that minimize the energy consumption, taking into consideration that the routing of data through areas of the system switched to wireless for faults tends to require more energy for the communication. The defined architecture with its routing solution favors scalability and does not imply restrictions neither on the number of base stations gathering the unit data, nor on the system topology that can reflect that of the monitored structure. All the mechanisms to coordinate the hybrid communication among the units and set their tasks in the system have been defined, embedding also energy saving policies. To complete the study an extensive evaluation of the developed solution, through system models, simulations and implementations (Fig. 2), is presented. The results show that the proposed methodology can be effective to support the development of hybrid systems that provide robustness to faults and an efficient management of the available energy



1. Hybrid Wired-Wireless Monitoring System



2. Hybrid Monitoring System Model and Implementation

# THE DEVELOPMENT OF THE VIRTUAL RIDER FOR ATV BASED MOBILE ROBOTS IN ROUGH TERRAINS

Adnan Tahirovic

An autonomous mobile vehicle capable to traverse a wide range of poorly natural terrains is an indubitable useful concept that can be utilized in different kinds of fields. A variety of needs of planetary explorations, rescue missions in hazard areas, humanitarian de-mining as well as agriculture applications have recently triggered a lot of research works aiming at developing sufficiently reliable motion and navigation planning approaches in such environments. This work is an attempt to propose a new such approach to guide the vehicle (Fig. 1) to reach a goal position on irregular terrains.

The proposed navigation planning approach is based on the model predictive control paradigm (MPC) that allows a vehicle to select less difficult terrain regions to approach the goal position taking into account a variety of different constraints, such as preventing the vehicle from sideslip and rollover. In addition, an MPC path planner continuously repeats the optimization during the task execution allowing for new local sensor measurements to be taken into account. Besides the safety issue regarding collision free paths, such planning policy gives a certain level of robustness to an MPC generated path comparing to other approaches. Finally, if there is an algorithm to compute a sufficiently good approximation



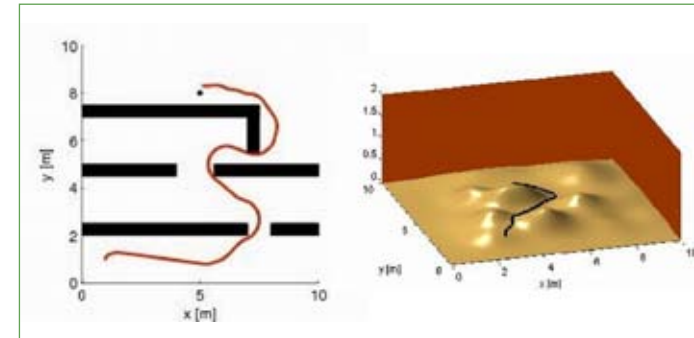
1. An ATV based vehicle (left) and NASA/JPL Mars exploration rover

of the remaining traversability measure from each terrain location toward the goal position, then an MPC based planning approach can be considered as a near optimal planner.

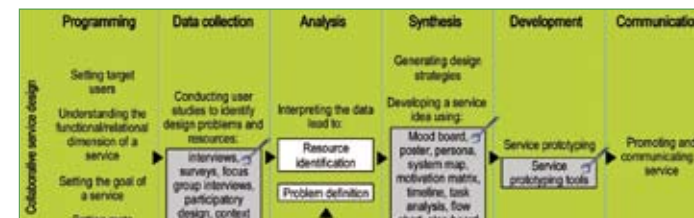
As the first step of the work an MPC optimization setup combined with the control Lyapunov function (MPC/CLF) is adapted from flat to rough terrains. The optimization setup uses the CLF to stabilize the goal position, implying the task completion. Since the proposed framework uses the unicycle mobile vehicle to generate the reference trajectories in order to predict future vehicle states within an MPC optimization horizon, a new MPC navigation approach has been proposed based on the passivity control concept (PB/MPC). The PB/MPC navigation approach stabilizes the goal position using both energy shaping technique by the use of navigation function and the passivity control concept.

Unlike the MPC/CLF approach where the control action should be found in advance in order to satisfy the MPC constraint on the CLF, the PB/MPC control action is a direct consequence of the passivity based approach. Such a property makes the PB/MPC navigation approach general and suitable to be used for a wide range of vehicles. For this reason, the PB/MPC can also be easily adapted to rough terrains where a truly complex vehicle that comprises the terrain model is used to predict trajectories within the MPC optimization horizon. Simulation examples are shown in Fig. 2.

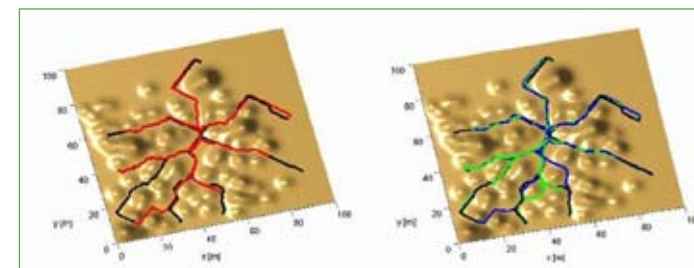
In order to improve an MPC path planner, a roughness based navigation function (RbNF) is proposed to approximate an optimal cost-to-go map providing information on the cost-to-go term required by the MPC optimization setup. The RbNF represents a numerical map that



2. Paths generated by the PB/MPC optimization framework for a flat terrain (left) and a rough terrain (right)



3. A Mars terrain region (left), a roughness map for a portion of the terrain region (middle) and the computed RbNF (right)



4. Paths generated by the optimal planner (black), gradient-based RbNF planner (red) and different MPC path planners. In all simulation examples the goal position is set in the terrain centre.

contains an estimated cost-to-go value for each terrain location. Fig. 3 shows a Mars terrain region and its corresponding RbNF. It was shown that the algorithm to

compute the RbNF is much faster than the one used to obtain an optimal cost-to-go map. Such an advantage makes an MPC path planner guided by the RbNF to

be a good solution for large-scale maps, where finding an optimal path may be computationally too expensive. The work demonstrates that an MPC path planner combined with the RbNF is a near optimal navigation approach and is a good strategy to deal with terrain uncertainties. Simulation examples are shown in Fig. 4.

The work also illustrates the possibilities to use the RbNF as a useful tool for solving other rough terrain related problems. Finding the most appropriate landing region for a rover in a Mars mission is one of such open research problems. The execution time and simplicity of the RbNF algorithm make the RbNF map a superior tool to be used in large-scale terrains compared with the algorithm that computes the optimal cost-to-go map, where the map is used to compute the information on the mobility of a rover from each candidate landing region toward a priori defined goal positions. In addition, the work shows how the RbNF can be used as a tool to enhance a Rapidly Exploring Random Tree (RRT) path planner. The RbNF is used to expand the exploring tree providing an RRT-like planner to be considered as a near optimal navigation approach.

## SPECIFICATION AND VERIFICATION OF QUALITY OF SERVICE IN OPEN-WORLD SYSTEMS

**Giordano Tamburrelli**

Modern software systems are increasingly built out of components that are developed, deployed, and operated by independent organizations, which expose them for use by potential clients. We refer to systems designed in this scenario as Open World Systems [1]. Web services and, in particular, Web service compositions represent a concrete example of such systems.

Continuous change is typical of this world and may occur in critical components of the system, clients' operational profiles, requirements, or deployment environments. Unpredictable changes continuously affect software systems and have a severe impact on the quality of service potentially jeopardizing the system's ability to meet desired requirements.

The thesis precisely addresses this issue and focuses on models for quality of service at run-time. Indeed, the adoption at run-time of software models and model checking techniques allow software engineers to automatically reason about changes, detect harmful configurations and—potentially—react appropriately.

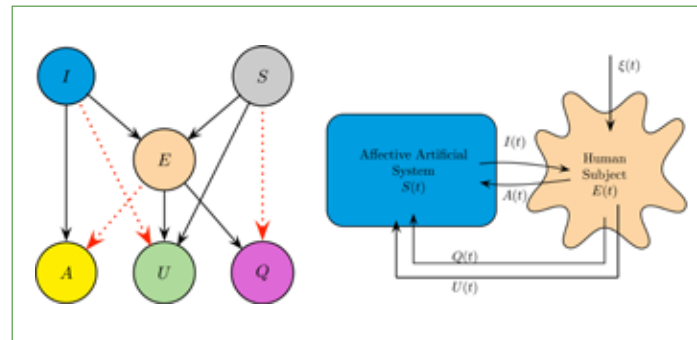
We propose an approach for run-time modelling of systems and a framework (i.e., KAMI) which implements the approach. Moreover, we defined a run-time modelling technique for reliability of system and change-point detection technique based on Discrete Time Markov Chains. Finally, the contributions of this thesis has been implemented and validated through a case study and numerical simulations.

## A METHODOLOGICAL FRAMEWORK FOR PHYSIOLOGY BASED AFFECTIVE COMPUTING: DEFINITION AND EVALUATION

Simone Tognetti

In this work, we have faced the problem of estimating emotions of people that interact with an artificial system. An emotion, in this work, is considered to be the internal state of a person, which cannot be directly observed, and which we want to estimate from a set of measures gathered during the interaction with the machine. This problem has been addressed mainly by two disciplines: Psychophysiology and Affective computing.

Psychophysiology is a branch of Psychology in which researchers aim at mapping psychological states to physiological reactions by designing models that generalize over many people and contexts. Psychophysiology has developed models and methodologies to manage this complex problem. Psychophysiology has a strong grounding on methodological aspects that support the identification of valid, reliable and general descriptive models of the phenomena; however, it suffers from applicability mostly limited to lab settings. Being a science, Psychophysiology aims to describe phenomena, namely affective phenomena. Affective Computing is a recent discipline that differentiated from Psychophysiology with the purpose to overcome some of its limitations. First, it uses different sources of information to evaluate emotions. The use of



1. General model of affective human-computer interaction, **Q** is the questionnaire answer, **U** is the unintentional form of interaction, **A** is the action, **I** is the stimulus and **S** is the external noise

physiological signals, extensively used in Psychophysiology, was not able to produce the expected results all the times. The introduction of new signals (e.g., speech, facial expressions, posture, etc.) aims at overcoming the limiting results of Psychophysiology. Moreover, Affective Computing leaked the experimentation outside the labs. Affective Computing has thus a strong grounding on techniques and methods to achieve the goal of facing the problem in an application scenario; however, it suffers from poor attention to methodological aspects so that, often, it is difficult to replicate the obtained results. The two disciplines have followed parallel paths in which few attempts have been made in order to give a general overview of the problem. By taking into account the interesting aspects

and overcoming the limitations of both disciplines, the main contribution of our work is the proposal of a general methodological framework that provides: (1) a general model of the affective phenomena that takes place in real life Human-Computer Interaction (HCI) scenarios; (2) a methodology that guides the definition and realization of experiments to estimate valid and reliable models in these scenarios.

The model we are proposing is grounded on the findings of both disciplines, and it aims to face the emotion estimation problem within a unifying view that is useful for two main reasons: firstly, it helps to formalize variables, relationship among variables and the hypotheses that are relevant for the problem, then, it helps the validation



2. Wearable device developed during our research. It acquires BVP, GSR and a 3-axes accelerometer with bluetooth data transmission

process, thus improving the model toward a valid and reliable formal definition. The methodology we are proposing, on the other side, is useful to provide "guidelines" to use and to estimate the model. An agreed methodology is useful for the following reasons: it supports in the decisions about critical aspects, reducing variability on results, and it promotes the replication and sharing of results. An agreed model and methodology are still not available in literature but are two key elements that people facing this problem should implement: they both enable the validation and the replication of results, the inner fundamental aspects of scientific method. The second contribution of our work is the application of our methodological framework to three different scenarios. The first application we considered is a Rehabilitation

Robotics task. Human Robotic Interaction is an interesting application for affective computing, since robots need to perceive people in order to have proper reactions. In a rehabilitation robotics task it is important to perceive patient's stress in order to adapt the therapy and allow for a better recovery. We involved impaired patients that have been asked to perform a rehabilitation protocol characterized by a set of sessions with a rehabilitation robot. The second application scenario is a multimedia context where it is relevant to perceive the subjects' interest to customize the multimedia content they receive. 15 subjects have been exposed to different multimedia contents in order to recognize their arousal. Finally, a third context is given by a video game scenario. The videogames are very popular in the Affective Computing community since they provide a good testing

environment that is able to effectively produce a wide range of emotions. We involved 75 subjects, we asked them to play a racing game and their preference among different game settings has been modeled. The third contribution of our work is related to the problem of the suitability of devices to measure physiological signals. Most of the current research on emotion recognition is based on laboratory instruments that become less suitable when the research is brought outside laboratory contexts. These instruments have been designed to work in a lab, they affect the movements of the subjects and they require a relevant amount of time to be placed on the person. We designed and implemented a wireless headset that is easy-to-wear, that does not compromise the interaction, and that can reliably measure the physiological signals in many application contexts.

# NO-REFERENCE AND REDUCED-REFERENCE METHODS FOR MULTIMEDIA FORENSICS AND QUALITY ASSESSMENT

Giuseppe Valenzise

In the last fifteen years, the production and distribution of digital media content, such as images, audio, video, etc., have grown at a formidable speed. On one hand, this content can be easily and cheaply acquired through off-the-shelf devices, even by people without specific training in multimedia technology. Moreover, editing these types of content in order to enhance them, save them to another format, or even tamper with them to alter their semantic content is a matter of few clicks thanks to popular commercial software. On the other hand, users are nowadays able to share, advertise and disseminate both self-generated and third-party content over on-line archives (e.g., YouTube or Flickr), social networks (e.g., Facebook or MySpace), blog spaces and peer-to-peer networks, thus rendering each single copy virtually everlasting. This has been enabled, among other things, by efficient video and audio coding tools developed in the last two decades, which have dramatically reduced the bandwidth required to transmit and store multimedia content. Examples are the DivX (MPEG-4) and the state-of-the-art H.264/AVC codecs for video, or the MPEG2-Layer 3 (MP3) or the Advanced Audio Codec (AAC) in the case of audio. In turn, also the traditional distribution schemes adopted by

the music and film production companies have to undergo a radical transformation to contrast the advancing of competing production (user-generated content) and sharing (peer-to-peer clients, web-based streaming) paradigms. Therefore, we have been gradually observing a shift of the market from the retail sale of physical media (CD, DVD, etc.) and television broadcasting, to web stores (e.g. the iTunes Store by Apple or the online streaming system offered by Netflix), Internet Protocol Television (IPTV), podcasting, etc. These new production, editing and dissemination paradigms raise important issues related to the integrity and the quality of the downloaded content. In fact, each individual copy downloaded by a user is potentially unique, in that it could differ from the original due to several factors:

- a distinct acquisition process: e.g., the same piece of speech might have been acquired with different recording devices; or a video might have been "ripped" from a DVD or grabbed in a theater;
- coding: the same content may have been coded (or transcoded) with different coding standards, with different results in terms of quality. For instance, the H.264/AVC standard is able to compress video at half or less the bit rate of MPEG-2, while

preserving approximatively the same quality;

- editing: e.g., photos are typically enhanced by means of commercial software tools, and retouched (this could be either a form of enhancement, or be aimed at doctoring the semantic content of the image);
- transmission: audio-visual content transmitted over error-prone networks may suffer from transmission errors, packet drops, delay or jitter, with a consequent loss of visual quality for the final user.

In the thesis, we describe methods for evaluating the last two sources of disparity between a copy of a digital content and its original version. The focus is on the user side, where either the authenticity, the integrity or the quality of the received content are to be assessed. Such a viewpoint is particularly useful, as it allows the evaluation of these criteria from the same perspective experienced by the final content consumer. At the same time, focusing on the receiver side is extremely challenging, since the end-user may receive the digital content after a chain of processing operations including the four categories (acquisition, editing, etc.) described above.

One of the major limitations when working at the receiver

side is that the original version of the content, which would be the natural term of comparison for the received copy, is not available. There are two main approaches to judge the received content in this case, which correspond to different architectural choices or constraints:

- The *reduced-reference* (RR) approach, where the receiver has access to a compact representation of the original signal by means of a signature or a hash. The RR information is typically transmitted to the user through an auxiliary channel, which is supposed to be unsusceptible to losses or malicious intruders that might alter the signature. In some cases, it is possible to insert the RR information directly into the original digital content (e.g., using watermarks).
- The *no-reference* (NR) approach, where no information about the original content is available at the receiver side. In this case, the receiver can only rely on statistical characteristics of the digital signal to assess the impact and the extent of possible modifications.

Generally speaking, RR methods are more precise than NR, but integrating them seamlessly into the network architecture requires the deployment of a dedicated RR channel, which is not always

the case. On the other hand, NR methods entail the maximum flexibility from the network point of view; however, they are more challenging, and typically require an accurate modeling of the signal and of the degradation in order to provide accurate results. The discussion in the thesis follows these two approaches, for the two sources of disparity discussed above (editing and transmission). The impact of these tools principally involves, but is not limited to, the fields of images/video forensics and video quality assessment. With regard to the integrity of the received content, we analyze the problem of tampering in the RR setting for image forensics, designing a multimedia hash that enables to verify the authenticity of the multimedia content and, in case of forgery, to identify the kind of attack. The hashing system leverages recent findings in the field of Compressive Sensing, and exploits distributed source coding tools to reduce the overhead cost of transmitting the RR information. We then focus on the case of video, and describe preliminary results on the no-reference identification of encoding parameters (e.g., quantization parameters) which could be a useful tool in forensic analysis. As for the quality of the received content, we concentrate on quantifying the degradation of

a video due to channel errors. We propose a RR approach based on distributed source coding tools. We also discuss a NR video quality assessment technique which is able to infer from the decoded pixels which are the portions of the video that have been lost, without accessing the bitstream. This information is fundamental to produce accurate estimates of channel-induced distortion.

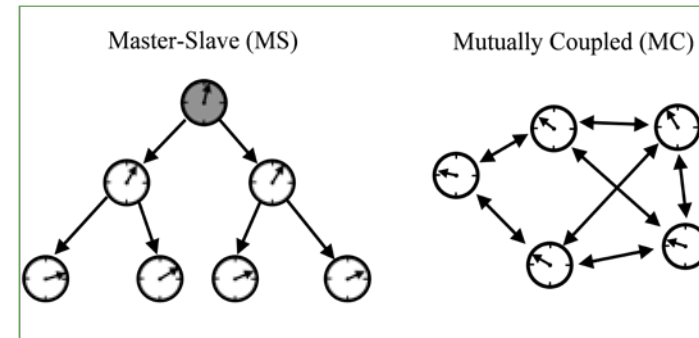
# DISTRIBUTED SYNCHRONIZATION ALGORITHMS FOR WIRELESS SENSOR NETWORKS

Nicola Varanese

The ability to distribute time and frequency among a large population of interacting agents is of interest for diverse disciplines, inasmuch as it enables to carry out complex cooperative tasks. In a wireless sensor network (WSN), time/frequency synchronization allows the implementation of distributed signal processing and coding techniques, and the coordination of the access to the shared wireless medium. Large multi-hop WSN's constitute a new regime for network synchronization, as they call for the development of scalable, fully distributed synchronization algorithms. In this thesis, the problem of realizing network-wide synchronization is approached by employing distributed clock control algorithms based on the classical concept of coupled phase and frequency locked loops (PLL and FLL). By observing that WSN's allow for greater flexibility in the design of the synchronization network architecture, this work considers both peer-to-peer (mutually coupled - MC) and hierarchical (master-slave - MS) architectures (see Figure). In general, MS topologies guarantee faster synchronization, but they are hindered by higher noise accumulation, while MC topologies allow for an almost uniform error distribution at the price of much slower convergence. Also, while most

of previous research focused on synchronization at the application layer, this thesis considers synchronization at the lowest layers of the communication protocol stack of a WSN, namely the physical and the medium access control (MAC) layer. At the physical layer, carrier frequency synchronization issues arise when employing distributed space-time coding (DSTC) or other cooperative communication techniques. In this case, multiple nodes transmit at the same time, and classical synchronization techniques cannot be employed. We develop a synchronization algorithm based on the concept of distributed frequency locked loops (D-FLL) in order to compensate different carrier frequency offsets (CFO) within an ensemble of nodes. A novel frequency difference detector (FDD) is designed to estimate the synchronization error between each node and its neighbours from the preamble of physical-layer frames. The computed error is employed by a proportional controller to adjust the frequency of the local oscillator. We derive stability conditions and evaluate the steady-state synchronization accuracy of a network of D-FLL's with both MC and MS architectures. The deleterious effects of CFO and the benefits of frequency synchronization based on D-FLL are assessed for a multi-hop relay network employing a

distributed differential space-time code. At the MAC layer, time synchronization issues arise when employing slotted medium access protocols, e.g., TDMA. We first focus on synchronization acquisition in peer-to-peer MC architectures, where convergence of distributed algorithms can be especially problematic. The considered MAC protocol prescribes nodes to exchange time information by broadcasting signalling frames (beacons) including either frame sync sequences or time-stamps. Depending on the signalling format, beacon frames can be transmitted either at random (contention-based) or in a deterministic (reservation-based) fashion. For all signalling strategies, the analysis provides sufficient conditions for almost sure stability of distributed synchronization algorithms based on PLL's. Simulation results show that the most efficient signalling strategy is the random transmission of frame sync sequences as it allows faster convergence. The accuracy of network synchronization is analyzed in depth for the case of reservation-based signalling. First, we consider the case of clocks whose frequency is stable, i.e., it does not change between subsequent synchronization updates. Assuming a linear clock



1. Master-slave (MS) and mutually coupled (MC) synchronization architectures

model, whereby each clock is uniquely defined by its initial phase and (constant) frequency, we develop a novel algorithm for the distributed estimation of clock parameters. Closed-form expressions are provided for the estimation error over both MC and MS architectures. Distributed estimation is compared with the accuracy achieved by employing PLL algorithms based on a proportional integral (PI) clock controller. In particular, the time error at steady-state is characterized for a network of coupled PLL's with regular MC topology. Simulation results show that coupled PLL's achieve better accuracy with respect to distributed estimation over both MC and MS architectures. Also, the Cramer-Rao lower bound (CRLB) is computed for the observation model at hand. Comparison with the CRLB shows that both coupled PLL's and distributed estimation algorithms

are quite far from the bound when employing MC topologies. The assumption of clock stability does not hold when the network operates with very low duty cycles. Low duty-cycles imply that nodes leave their radio turned off for most of the time, thereby limiting the frequency of synchronization updates. As the time between subsequent updates grows larger, the frequency of local clocks does change due to, e.g., varying environmental temperature. In this case, the clock controller needs to be optimized in order to effectively track the changing frequency. With the intent of simplifying controller optimization, we present a hybrid PLL/FLL algorithm that adjusts phase and frequency by employing two control loops in parallel. The tracking accuracy of both PLL and hybrid PLL/FLL algorithms is derived in closed-form for the case of regular

MC architectures. Finally, the performance of clock control algorithms is compared with conventional techniques based on raw temperature compensation (temperature compensated clock - TCC). Simulation results show that PLL and hybrid PLL/FLL algorithms are both superior to TCC on any synchronization architecture. In conclusion, synchronization algorithms based on phase and frequency locked loops prove to enable accurate and scalable synchronization with both peer-to-peer (MC) and hierarchical (MS) architectures, thereby providing absolute freedom in synchronization network design. Also, clock control algorithms show to be robust against packet collision events, infrequent synchronization updates, and errors introduced by different noise sources, such as transmission delays and clock frequency instabilities.

## STABILITY ANALYSIS AND STABILIZATION OF SWITCHED POSITIVE LINEAR SYSTEMS

**Annalisa Zappavigna**

Since the systems and control theory exists, stability properties of dynamic systems have been one of the most outstanding problems.

Still now there are some classes of systems that require a deeper study about it. Switched positive systems, that are extremely common in current technology, are one of them.

Stability of switching systems is generally not a trivial issue, since unconstrained switching may destabilize a switched system even if all individual subsystems are stable. At the same time, however, it may be possible to stabilize a switched system by means of suitably constrained switching even if all individual subsystems are unstable.

Therefore two main problems arise: the first one is to find conditions that guarantee asymptotic stability of a switched system for arbitrary switching signals; the second one occurs if a switched system is not asymptotically stable for arbitrary switching, since in this case it is interesting to identify those switching signals for which it is asymptotically stable.

Positive systems, instead, have the peculiar property that any nonnegative input and nonnegative initial state generates a nonnegative state trajectory and output for all

times. Positivity of the variables often emerges as the immediate consequence of the nature of the phenomenon itself.

Examples of switched positive systems are very common: networks of tanks regulated by valves opening, chemical plants where reagents concentration can be varied by mean of additional inputs, air conditioning systems. Also many applications in Communications networks involve algorithms that lead to extremely complex positive systems, typically involving parameter switching, such as networks employing TCP and other congestion control applications, synchronization problems and wireless power control applications.

There are many topics that can be studied in this field. This thesis has been focused on four main issues, that we can classify by mean of the switching signal, with the exception of the last one, that is about some numerical considerations:

1. Arbitrary switching: we study stability analysis of switching positive systems when the switching signal is arbitrary. Then, we prove that the existence of a common co-positive linear Lyapunov function is sufficient to guarantee the stability

under arbitrary switching of a switched linear positive system even in the presence of uncertain but bounded time-varying delays. Finally, we present a computational test, in terms of the original matrices, to determine whether a common co-positive linear Lyapunov function exists, this in turn implies exponential stability in the presence of time-delays.

2. Constrained switching: in this section the switching signal cannot be arbitrary, but it is constrained to be slow enough to guarantee the system stability. Hence, in this section, we establish dwell time results for a class of switched linear positive systems and an upper bound to the  $L_1$  induced norm of switched linear positive systems is computed.

3. Controlled switching: we provide a result on state-feedback stabilization of autonomous linear positive switched systems through piecewise linear co-positive Lyapunov functions accompanied by a side result on the existence of a switching law guaranteeing an upper bound to the optimal  $L_1$  cost. Then, the induced  $L_1$  guaranteed cost is tackled, through constrained piecewise linear co-positive Lyapunov

functions. The optimal  $L_1$  cost control is finally studied via Hamiltonian function analysis.

4. Switching positive systems discretization: we find some sufficient conditions on the sampling time in order to preserve positivity through diagonal Padé discretization. Moreover we prove that this kind of transformation preserve common Lyapunov functions and, hence, the stability for switching positive systems when the sampling time  $h$  is suitable. We propose also a different approximation to the exponential matrix and we show that, in some cases, it can be better than the Padé.