Scientific Activity of Digiteo

Version 3.2, 18/9/2008

Contents

This document has been elaborated and written by the following working group:

AIGUIER Marc, ECP
BAPTISTE Philippe, LIX
BONASSIEUX Yvan, LPICM
DAGUE Philippe, LRI
EISENBEIS Christine INRIA/Saclay Ile de France
FLEURY Gilles, Supélec
FOURNEAU Jean-Michel, PRISM, UVSQ
GAUDEL Marie Claude, Digiteo, Comité des Programmes
JUVIN Didier, CEA/LIST
PLATTER Denis, CEA/LIST
PLUQUET Alain, CEA/LIST
REY Christian, Institut Farman, ENS Cachan
REYNAUD Roger, IEF
ROBIN Maurice, Digiteo, Director
TARROUX Philippe, LIMSI
TOUBOUL Françoise, CEA/DEN
WALTER Eric, LSS

Thanks are due to Alain Denise and Jean-Marc Martinez

The work was coordinated by Marie-Claude Gaudel

1 – Introduction

At the end of 2004, six organisms located in the vicinity of "Plateau de Saclay", namely: CEA, CNRS, École Polytechnique (X), École Supérieure d'Électricité (Supélec), INRIA and Université Paris-Sud XI, signed a collaboration agreement to express their ambition to build the first French research cluster in the domain of Science and Technology of Informatics and Communication. They chose a common name: Digiteo labs, and started to establish common scientific activities, based on an analysis of their strengths and competences ([1]).
Since then, a major event took place: in April 2007, the French government created several national RTRA (Réseau Thématique de Recherche Avancée), and among them, Digiteo, which kept the partners and the themes of Digiteo-labs with, in addition, the association of a few laboratories and organisms chosen because of their proximity: Ecole Centrale de Paris (actually located in Chatenay-Malabry), Ecole Normale Supérieure de Cachan, INRIA-Rocquencourt, Université de Versailles Saint-Quentin.

This document is an update of ([1]) after two years and the arrival of the new partners. Its intended audience is the research community of Digiteo (there are more than 1500 researchers) and the recently created scientific board. It will be used as a basis for the call for projects and other activities within Digiteo.

The scientific domain of Digiteo can be characterised as follows: design and development of systems with strong software implication; the scope of the considered systems goes from systems on chips to high performance computations and large software infrastructures, including embedded systems and robots. All these complex systems raise the following problems:

- They must provide advanced interaction with the external world (intelligent sensors and actuators) and with their users (vocal activation, haptic interfaces, visualisation, interactive environments);
- Their development requires some modelling and/or simulation, in order to assist their design, anticipate their performances, their dependability, commandability and controllabilility;
- They require a deep knowledge of software and high level aspects of hardware, both the theoretical foundations and the methodological aspects that make it possible to take into account criteria such as correctness, efficiency, reliability, robustness;
- Last but not least, they often raise the issue of managing the huge and disseminated mass of data and computation power that are now available thanks to computer networks.

Most these software systems will be embedded and/or distributed, sometimes on a large number of sites.

Analysing and organising the scientific activity of such a large consortium is far from being obvious. The establishment of the consortium cannot aim, reasonably, at reorienting every local team toward new research objectives. Diversity can bring much and is necessary to successful innovation. The idea is to take advantage of the diverse competences and pieces of knowledge that are present in the consortium, to consolidate them, to exploit their complementarities and possible synergies, and to prepare the future.
Thus, this document is organised as follows.

---

[1] See « La recherche dans Digiteo-labs », version 5.1, 20/6/2005, in French, available on request at mcg@lri.fr

First, six main research themes are presented that have been identified as recognised fields of competence of Digiteo, essential to the progress of its domain. They must be consolidated and encouraged so that they better support transversal topics.
Then four transversal topics are described. They are transversal in the sense that aspects of several of the six main themes above are required to lead successful research in this area. Some works are already done in Digiteo, but it seems likely that they could benefit of more interaction, both inside the topic itself and outside it, with some colleagues of the themes above. Such interactions must be looked for, suggested and encouraged so that they take shape.
Last but not least, three exploratory subjects are mentioned. They are of common interest for several of the previous themes. Their list must not be seen as limiting. The goal is not to mount big projects in these areas in the near future, but rather to ensure that there is some competence on these topics in Digiteo and that the Digiteo community is aware of that.
The six so-called main research themes in Digiteo are presented in Section 2. They are titled:

- Software
- Modelling and Simulation
- Computing Systems
- Sensing Systems
- Decision and control systems
- Interactions, visualisation and Virtual Reality

The four transversal topics are described in Section 3; they are titled:

- Robotics
- Bioinformatics
- Hybrid systems
- Future Networks

Section 4 is devoted to the three exploratory subjects that are titled:

- Uncertainty
- Probability in modelling, algorithms, and verification
- Optimisation

2 – The main research themes in Digiteo

In this section, six main research themes are presented that have been identified as recognised fields of competence of Digiteo, essential to the progress of its domain. They must be consolidated and encouraged.

2.1 - Software

Digiteo addresses a variety of problems raised by software development, validation and verification. The common target is to make these activities safer, resulting in better quality and even certification of the resulting software, and easier, resulting in a more efficient production process.
Thanks to the wide range of competences of the involved groups, the researches on software led within Digiteo seamlessly spread from theoretical foundations (semantics, type systems,) to practical methods and tools. In all the cases the main challenge is how to scale up, both in size and quality.

Software specifications and models

It is now widely recognised that software development, verification and validation need to be based on models. Various kinds of specifications and models are studied in Digiteo, mainly oriented to the description of embedded real-time systems and distributed systems. They

include both pragmatic approaches (UML), and formal ones such as JML, or timed automata, or sophisticated process algebra (to cite just a few). The current orientation of Digiteo is more toward the use of these models for validation and verification (see next section), where a considerable amount of expertise has been developed and exploited, than in their use for development, even if some successful specialised projects are led in this area.
An emerging and promising research theme in Digiteo is to take advantage of the variety of studied specifications and models to combine them in order to increase the expression power while managing their heterogeneity in a formal and effective way.

Safety and security

The production of safe and secure software-based systems is a challenge where much effort is devoted in Digiteo. Researches on validation and verification are led both at model level and at program level.
There is a kernel of works on powerful generic tools in the area of theorem provers, constraint solvers, model-checkers, cryptography, combinatorics that requires fundamental research in computer science and mathematics. More applied projects are devoted to static analysis and proof of programs, model-based or program-based testing, model-checking of probabilistic, timed, or symbolic models. As an example, among many, in this area, one can mention the static analysis of rounding errors propagation in programs using floating numbers. All these approaches are validated on significant case studies and industrial projects.
Scaling-up is a major concern. Several works are exploring approximate verification methods, or statistical and probabilistic approaches with promising preliminary results.
Last but not least, fault-tolerant architectures and on-line diagnosis are the subject of some fruitful collaboration.

Boundaries with hardware

Within Digiteo, a long lasting collaboration has taken place on compiler optimisations for high-performance general-purpose and embedded processors. Moreover, there is a strong body of knowledge on software tools for fast processor simulation.
There is also some work on the development and certification of safe real-time operating systems.

Data and knowledge engineering, concepts and tools for data mining

Works led within Digiteo in this area cover research on data bases, data warehouses, knowledge modelling and the use of domain models, or ontologies, in information retrieval systems. Applications to web services, and algorithms and tools for data mining are also studied in conjunction with these topics.
Information integration systems have to provide easy and relevant accesses to sources. When these sources are numerous and heterogeneous, as Web sources, such systems rely on automatic tools to annotate and query documents, statistical analysis, or learning methods for data mining. Digiteo teams have developed strong competences in these complementary areas, and will take advantage of this.

2.2 - Modelling and Simulation

Modelling (based on a precise understanding of physical mechanisms) and numerical simulation are bound to play a strategic role in the design and evaluation of systems. Increasingly often, simulation indeed replaces actual testing on real systems or plays a key role in the optimisation of qualification plans, based on "smart" experiments.

Numerical simulation applies to virtually all industries where economic competitiveness is closely linked to the duration of the design/development cycles, to risk management and to the reduction of environmental impact. It can drastically reduce design and maintenance costs, time to market and the impact of regulatory changes.
Numerical simulation programs become more complex as they must handle multiple physical phenomena (to take into account the various components of the systems) and multiscale aspects (to manage the components hierarchically). They benefit from ever-increasing computing power, call upon intensive high performance or grid computing and often have to handle very large volumes of data. The schedules of development of software components and hardware components must be consistent. The codes must also meet internationally recognised standards of verification, validation and qualification.

A non-exhaustive list of topics of interest to Digiteo is:

Numerical methods: model reduction and optimisation, mesh adaptation, domain decomposition, quantification of numerical errors, specific finite element or finite volume methods (discrete element method, SPH (Smooth Particle Hydrodynamics), discontinuous FEM...). Particularly challenging is the search for generic aspects in methods initially developed for specific applications.

Multi-physics numerical modelling (neutronics, thermodynamics, hydraulics, mechanics, chemistry, metallurgy…). One of the challenges is to interface software components dealing with different physics while maintaining their best performance and without creating defects in the convergence of methods or in the continuity of results.

Multi-scale numerical modelling in time (*e.g.*, storage of nuclear waste, from the nanosecond to the century) **and in space** (*e.g.*, modelling materials from the dislocation level to the structure level): scale space methods, homogenisation/localisation...

Adaptation of software to hardware:
- Use of clusters: load balancing
- Dedicated algorithms for computation on grids/High performance computing (access to TER@TEC, CNRS / Blue gene)

Code supervision:
- Pre- and post-treatment adapted to models with very large meshes (millions of degrees of freedom)
- Strategies to couple codes and to achieve multi-scale computations
- Optimisation (cf. 4.3)
- Learning: construction of simplified models, response surface method for non-linear models, genetic algorithms,
- Treatment of uncertainties (modelling and propagation, cf. 4.1),
- Probabilistic methods (cf. 4.2).

2.3 - Computing systems

Computing systems involve multiple computing paradigms, from very fine grain components in processor architecture design or embedded co design up to massively parallel frameworks underlying grid computing or more futuristic bio-inspired or spatial computing models. They encompass from embedded processor up to high performance supercomputers, including general purpose processors. Europe is well positioned on embedded computing systems but a

lot of technical progresses have to be addressed in order to maintain and even improve this situation.
In fact, the exploitation of instruction level parallelism and also the increase of frequency are no more valid ways to improve computing performance. For some years now, the quest for the ultimate performance on a single chip uni-processor has come to a stop. Therefore, a massive shift towards multi-cores architectures is on the move. The increasing number of transistors lets think that in a near future, integration of tens or even hundreds of cores on a same die is a realistic issue. But this new paradigm has a deep impact on all technical aspects ranging from the architecture design up to the programming languages.

Besides the shift from performance-aware to silicon/energy-aware approaches to design programmable or reconfigurable cores, the main challenges are linked to the management of the complexity brought by lots of homogeneous or heterogeneous resources. In this context, it is also crucial to propose innovative execution models and solutions to overcome bottlenecks such as the latency of memory access and interconnection networks that do not scale with the number of processors. Power management techniques and reliability are also key issues regarding the actual integration capacities. On the other side, the need for high-level abstraction language and programming framework is becoming crucial for tackling the complexity issue and to fully exploit parallelism, keeping the programmer as far as possible from the architecture concepts.

These are main key research topics in Digiteo in the domain of computing systems.

- high performance computers and compilers,
- multicores on chips: design and programming, interconnection Networks on Chip (NoC) and memory management,
- distributed systems, multithreads and grid computing,
- power aware compilation and run-time power management,
- Programming models and languages, Algorithm-Architecture Adequacy (cf. 2.1)
- Future computing models (spatial, bio-inspired, quantum computing, ...)

Beside these topics some of the Digiteo teams could be involved in the design of a multicore architecture for high performance computing together with its programming environment (for instance using streaming concepts). This could bring a broad visibility to Digiteo as well as offer a joint framework on the top of which other research projects can be conducted.
DIGITEO partners represent an impressive research group on both hardware and software topics. Thanks to this potential, Digiteo promotes the ambition to become a leader on embedded computing systems.

2.4 - Sensing systems

A specific approach

The traditional approach in instrumentation follows a linear scheme going from some physical quantity to be measured to signal via transduction and then from this signal to a numerical estimate of the quantity of interest via signal processing.
A multitude of new schemes are now possible, thanks to the available computing power and the massive use of *modelling, simulation, prior knowledge, data bases and networks*.
The traditional linear scheme is therefore no longer the only option available and new research avenues open. The key concept of *information*, and information theory, should be at the core of sensing systems, which are at the boundary between the physical world (where the sciences are physics, chemistry, biology, *etc*. and the technology is that of sensors) and the

digital world (where the sciences are statistics, numerical analysis, *etc*. and the technology is mainly software).

Research topics

The scientific and technological aspects on the physical and digital sides are more and more interconnected, mixing fundamentals in physics, chemistry, biology, sensor technologies, and mathematics and software technologies. This leads to the multidisciplinary concept of *sensing system*, where crossover and synergy between instrumentation, information and computer sciences and technologies should be the source of new functionalities and performances.
A generic definition of the scientific program would be:

- Any approach based on information processing that transforms data produced by sensors (actual or simulated) into estimates of quantities of interest,
- Any approach where sensor design is governed by information-processing issues.

A non-exhaustive list of topics of interest to Digiteo partners is:

- Inverse problems, coupling physical sensors, physical models and simulations.
- Software sensors to estimate quantities that cannot be measured directly from available data and prior knowledge.
- Experiment design by statistical approaches to minimise uncertainty (cf. 4.1).
- Control theory for linear and non linear parameter and state estimation (see the *Decision and Control Systems* section).
- Coupling data bases and prior knowledge with physical measurements.
- Sensors networks.

Digiteo partners offer a global expertise in sensing systems, from physics, biology and chemistry to knowledge management and information processing and from generic design to final integration in industrial prototypes.

2.5 - Decision and control systems

This is about taking decisions based on an automated processing of the available information. Examples of such decisions are *classification* (e.g., *medical diagnosis* or *fault detection and isolation*) and *control* (i.e., the computation of input signals to enforce some desirable behaviour on a system). Such decisions usually rely on mathematical models that may have to be built, either from basic principles and laws of physics, chemistry, biology, etc. (*white-box* or *knowledge-based models*) or purely from input-output behaviour (*black-box models*), with all the possible *grey-box models* in-between.
Thus *modelling*, and *parameter estimation*, often called *identification*, are important ingredients. Also important is the ability to infer, from the available data and using the model, the values taken by variables that are not directly accessible to measurement but play a critical role in the decision process. Depending on the community, this is known as *state estimation*, or *observation* or *software sensing*. Because of the imperfect nature of the model and of the unavoidable perturbations, control usually involves some *feedback*, whereby output information collected via sensors, and suitably processed, is used to modify the input signals as implemented by actuators, in order to improve performance of the actual system to be controlled. Critical then is the ability to ensure the *stability* of the controlled system.
When accurate models are available and the objectives to be pursued easily translate into an optimality criterion, as is often the case in spatial control problems, *optimal control* is a natural approach. In many cases, however, the situation is far less clear-cut. Robustness of performance in presence of discrepancy between the actual system and its mathematical model, and *adaptability* of the control system to changing conditions, may then become more important than optimality.

The techniques and tools to be mobilised in the design of decision and control systems pertain to artificial intelligence, control theory, optimisation, signal processing and statistics.
Digiteo is interested in all aspects of decision and control systems, with a strong emphasis on black-box modelling, parameter and state estimation, nonlinear dynamical systems, the robust control of uncertain systems, constrained control and trajectory planning. Much is expected from the collaboration of Digiteo groups that attack the same problems with different tools and from different angles. Here is a non-exhaustive list of examples:

- In nonlinear state estimation, specialists of control, signal processing and statistics need to work together to deal with nonlinear, uncertain dynamical models and produce estimators that take uncertainty explicitly into account, as well as the intended use of the estimate.

- Black-box modelling, artificial intelligence, control theory and statistics each contribute their set of tools. Attempts to bridge gaps should be encouraged.

- In control, it is often assumed that the model and state of the system are available at the outset. The interactions between modelling, state estimation and control in a context of a very partial knowledge on the system and of unavoidable perturbations need to be investigated.

- Some teams that may not consider themselves as working on decision and control systems do deal with actual uncertain dynamical systems that may need to be modelled and on which decisions have to be taken. Other teams that develop control and decision methodologies do not have the equipment, skills and manpower required to deal with actual applications, e.g., in biology, communication networks, grid computing, physics, robotics. Both types of teams should benefit from common application projects.

- Control devices may be embedded in the system to be controlled, distributed on a network, which may itself be used to control distributed objects. Interactions between control theory, telecommunications and computer science should therefore play an important role. See also the section about hybrid systems.

2.6 - Interactions, visualisation and virtual reality

Interaction between humans and machines is one of the areas in computer science that has evolved a lot these last years. Progresses and innovations are mainly due to increases in computer power and technology of sensor devices and interactive software. It is also the result of new ways of considering the interaction with computers and the role of computers in everyday life.
Another field renewed by technological advances in display systems but also by the increasing power of computer systems and the growing up of massive forms of computation is physical systems modelling and data visualisation. Moreover, visualisation of large amount of data must be combined with new ways of interacting with a machine in virtual reality systems.
It is worth to mention that these researches imply the study of the close relationship between human and machines during the interaction, the consequence being the development of researches relating sophisticated software realisations, psychological and psychophysical studies in an interdisciplinary framework. It is indeed increasingly mandatory to understand how humans perceive these technological environments and interact with them but it is also necessary to validate the technological solutions mediating these interactions with end users.

Interaction and visualisation

Interaction paradigms

Digiteo partners develop interaction techniques as well as new tools for developing these techniques, and new methods for controlling the design process of interactive systems.
Some of them focus on the invention and development of new devices and interaction paradigms between users and computer-based systems. Some of them are focused on developing interfaces with natural communication with the user, employing all the user's senses and adapting to any disabilities, to increase the uptake of the systems and reduce their cost. It is the case, for instance, of interfaces that are adapted (or adaptable) to their contexts of use by taking advantage of the complementary aspects of humans and computers. In the long run, these researches seek to create a new generation of interactive environments as an alternative to the current generation of desktop environments. One use of these new interfaces is in the context of ambient computation and intelligent environments in which human will be able to interact with technological objects in a direct, situated and more flexible way.
Other new forms of interfaces studied by Digiteo partners are based on virtual agents (Animated conversational agents, talking heads) able to mediate the interaction with computational systems through language and emotional cues.

Information search and dialog systems

This topic mainly focuses on knowledge engineering, multimedia and multilingual information research, multimedia document indexing and semantic analysis of multimedia documents (image and video). One of the main concerns of this topic is written and spoken language analysis (and synthesis) for which Digiteo has a well-established expertise. These researches aim at providing tools for multimedia document automatic indexing through semantic and structural elements extracted from text, speech or images and videos, and more generally from unstructured documents (e.g. from the internet). They also focus on the development of question and answer systems and on their coupling with information research, allowing them to integrate extra linguistic elements such as prosody or emotional content.

Visualisation

Another important topic concerns the development of multimodal and multi-scale interfaces for interactive visualisation of large data sets. It leads to the study of the visualisation process and to the development of new software tools for the display and visual exploration of such data.

Virtual and augmented reality

Digiteo partners conduct researches on different aspects of virtual and augmented reality, from the development of algorithms dedicated to these technologies (geometric real-time algorithms, collision detection, rays-tracing), to the design of methodologies for interaction, with a strong emphasis on the question of multimodality and trans-modality. One particular topic well developed even at the hardware level is haptic interaction and more generally the question of the interaction with systems with high degrees of freedom.
Most of these research subjects are motivated by the involvement of some Digiteo partners in the development of tools for the numerical plant in which a large part of production studies will be conducted through numerical simulations. We can mention in this field the questions raised by the computer assisted design in virtual systems (modalities, object and action coding, object manipulation…). Another related question is the design and use of virtual humans with the questions raised by the poly-articulated systems, the grasping and the object manipulation by these avatars. In the field of the numerical design one major topic is

collaborative work in distant immersive environments for which there are active collaborations between Digiteo partners.

3 – Transversal topics

In this section, four transversal topics are described. They are transversal in the sense that aspects of several of the six main themes above are required to lead successful research in these areas. Some works are already done in Digiteo, but it seems likely that they could benefit of more interactions, both inside the topic itself and outside, with some colleagues of the previous themes.
Such interactions are likely to take place via the choice of a few challenges, guided by some thorough analysis of the specific competences and resources in Digiteo. Such challenges should be the basis of original and highly advanced research projects leading to significant scientific breakthrough.

3.1 – Robotics

Robotics brings together several very different engineering areas and skills and is a quite transversal domain going from mechanical design, automatically controlled, reprogrammable, multipurpose, manipulator programmable in several axes, to multi-modal emotional man machine interface. So, most Digiteo's partners contribute, directly or indirectly, to research in robotics. In order to organise the presentation of expertises available in the consortium, we will first present the research topics carried out by the partners, then the applications and we will conclude by the safety issue.

1/Research topics

The characteristic of a robotic system is to be able to autonomously adapt its behaviour to its environment. The robot needs to perceive its environment, to compute the most relevant behaviour according to its perception, its memory and its goals and then to act. So the development of robotics systems requires developing researches along these three stages.
1-a: Perception: Perception is needed to localise a mobile robot in its environment and to identify objects within this environment.
Digiteo's actors are involved in localisation based on inertial sensors, odometry, and satellites information but also on bio inspired methods for localisation. They are also involved in the study of perception-action relationship and active perception (e.g. attentional mechanism) in which action is required for perception.
1-b: Behaviour control: Work in this area goes from low level control of actuators to high level path planning and man machine interface.
Digiteo's partners are working on:

- low level control loop of classical electrical actuators or new technology based actuators (shape memory alloy, piezo-electric, magneto-rheological fluids …) using conventional approaches, or learning and evolutionary methods,
- adaptive control architectures and models of the perception-action loop based on connectionist, bio inspired and Bayesian paradigms
- development of vision based control methods to control either mobile robots or manipulators in order to servo robot motion on vision feedback,
- path planning with different methods,
- development of man machine interface based on haptic devices and graphical interfaces, or of affective avatars as interaction support between man and robot.

1-c: Action: this issue requires efforts on the design of the physical architecture of the robot. Digiteo's partners are involved in modelling and identification of complex mechanical

structures in order to optimise their design, in flexibility and friction modelling for back-drivable mechanisms or long-range carriers.

2/ Applications

Based on the technologies presented above, the main applications developed by Digiteo partners are flying robot, mobile robot, automatic car control, tele-robotics, adaptive robotics, assistance to disabled people, haptic interfaces, humanoid robot, mini-invasive robotic devices…

3/ Safety

One of the main issues for robotics is safety. Humans and robots will have to work together closer and closer. It is necessary for robotic industry to ensure that robot will not become a threat for human beings. Ability to prove that software is doing what it is designed for is critical. The software engineering expertise developed by Digiteo's partners can bring much here.

Digiteo's partners offer a global expertise in robotics and related problems, from the fundamental mechanics to the emotional modelling and from the generic design to the final integration of industrial prototypes.

3.2 - Bioinformatics

From genomics to metabolomics, from molecular structure modelling to regulatory pathway extraction, from medical image compression to electromagnetic human exposure analyses, all these biological, and health-related, challenges are parts of research activities within Digiteo.
Bioinformatics turns out to be a natural transversal research theme since it is rooted in computer science, mathematics, signal processing and control expertises. Moreover, strong interactions with biologists' research groups have been developed, thank to the location of Digiteo. Objectives of Digiteo, such as dealing with massive distributed data and modelling for inference and prediction, are essential issues in this kind of research. The very goal is finally to integrate a very large amount of knowledge (heterogeneous qualitative and quantitative data, living mostly in huge dimension spaces), on biological systems in a multidisciplinary approach.

These researches can be structured in three axes:

1. System biology
    - Graphical models of interactions (as protein-protein interactions, regulatory and metabolic networks…)
    - Dynamical interactions model identification (population evolution processes, cellular processes…)
    - Macromolecular structures (RNA and protein structure prediction, comparison, design)
2 Dedicated sensors and related signal processing
    - Medical image processing (3D video reconstruction for computer-aided radiology and surgery, very efficient lossless compression, features extraction for oncology diagnosis)
    - Exposure statistical analyses (electromagnetic, chemical, radioprotection)
    - Micro and Nano biocompatible technologies (diamond technologies, biodegradable polymer nanostructures)
    - Human-machine interfaces (brain computer interface for disabled persons)
3 Specific knowledge interpretation and extraction methods, learning

- Huge dimension statistical data analyses, with very few replicates and a poor signal to noise ratio (gene discovery, functional genome annotation, biomarker selection, regulatory pathways extraction…)
- Design of heterogeneous database architectures, querying and integrating data, guiding the analysis process, mining dedicated data (signalling pathways extraction)
- Directed information estimators for casual relationship extraction (regulatory pathways extraction)

This constitutes a wide spectrum of research driven by biological, health-related and even environmental purposes, and based on fundamental researches in domains such as knowledge integration, interpretation and extraction, combinatorial and multi-criteria algorithms design, uncertainties modelling and propagation. This leads to developing generic tools that will give access to innovative solutions of the specific problems proposed by the biology community.
Two senior invited chairs have been funded by Digiteo for the next four years (2008-2011). Professor Alfred Hero (Department of Electrical Engineering and Computer Science and, by courtesy, Department of Biomedical Engineering, and Department of Statistics – University of Michigan) will work on "Distributed active network sensing and estimation (DANSE)". Professor Peter Clote (Department of Computer Science and, by courtesy, Department of Biology – Boston College) will work on "RNA structural and systems biology". Both professors will have certainly a strong impact on the development of the current works in this area.

3.3 - Hybrid systems

This is about designing and validating systems that combine continuous-time parts (associated with, e.g., the physical world) and event-driven parts (associated, e.g., with embedded computers). Combustion engines, for instance, are hybrid systems *per se*, as the nature of the continuous-time physical processes taking place in the cylinders depends on the stroke (intake, compression, combustion, exhaust), which itself depends on conditions on the engine state. Most often, however, hybrid systems involve computers that may be embedded in the part of the physical world with which they interact and may also be connected to it and together through networks.

The number of such hybrid systems explodes as computers interconnect at a staggering rate with physical systems that they contribute to make more adaptable and more efficient. Design is necessarily multidisciplinary and multi-trade. It relies on a variety of mathematical models (PDEs, ODEs, discrete-events, hybrid), where control theory, computer science, signal processing and statistics play fundamental roles. The main challenge is to develop and disseminate methodologies to increase robustness and performance while reducing time and cost of design, development, validation, production and maintenance. Since many hybrid systems are safety critical (transportation or life-care systems, for instance), reliability is a major issue. Cooperation between Digiteo groups should bear fruit in many directions. Here is a non-exhaustive list of examples:

1. Foundational multi-disciplinary research (Control, Computer and Communication) and proofs of concept addressing the whole chain from observers, sensing, monitoring and actuation to adaptive and cooperative monitoring and control and decision making need to be developed for hybrid, possibly networked, systems. Parameter uncertainty, delays, disturbances, limited communications bandwidth, actuation constraints and node availability must be addressed.

2. Global approaches that address the design of the system as a whole, with its many physical, functional and logical aspects, taking into account the constraints inherent to embedded systems, should be encouraged.

3. New possibilities offered by computer sciences may impact the very nature of control algorithms. The safe implementation of control laws requires dedicated operating systems, dynamical reconfiguration of architectures and scalable algorithms for the control of evolvable, distributed and adaptable systems.

4. Formal tools that are being developed in computer science for analysis and proof of software bring new possibilities to deal with models of physical systems where time plays a critical role. To extend static analysis to hybrid systems, for instance, it is necessary to assess the range of the values taken by variables of uncertain dynamical models of physical systems, e.g., via interval analysis and guaranteed integration.

5. Besides the current application domains (including transportation, energy management, environment monitoring, factory automation, personal communication, process industry), new application domains should be explored.

3.4 – Future Networks

In the networking field, the main purposes of current researches and developments are to increase capacity in order notably to handle new traffics and new services, to improve quality of service in terms of both performance and reliability and to support the ubiquity of a global and generalised mobility (network mobility and multi-homing).
The number of available access technologies is growing fast (2G, 3G, 3G LTE, 4G, Wi-Fi hotspots, Wi-Max, xDSL, FTTx, cable, or satellite).
A key concept is the generalisation of self-organisation and autonomic networking, as a mean to create new service paradigms with a user-centric approach or to reduce the operational cost of existing services. Peer-to-peer applications and ad-hoc networks are typical examples of self-organisations at the application or network level. Various Digiteo groups are currently working on several of these topics within several collaborations with telecommunication companies and operators.

Mobile networks

One of the major contributions in this area has been OLSR (Optimised Link State Routing Protocol), a table driven proactive protocol with efficient flooding of control messages in the network. OLSR has been developed by some groups of Digiteo. It is now widely used and contributes to the development of ad hoc networks.
The objectives in the next step has been first to enhance the protocol OLSR and ad hoc networks with features such as quality of service (Bandwidth, real time service, multi-metric), multicast, security, etc. , and then to work on the future Internet (super nova) including thing2thing communication, connectivity, auto-configuration and delay tolerant networks.

Network Coding

Network coding is a new communication concept that permits the packet transmission from one or several sources to one or several destinations, departing from the traditional concept of routing. With network coding, nodes in a network require less topology information (such as routing tables) to deliver information, because the packets may be interweaved in several directions at the same time. The idea is to perform coding inside the network: typically the forwarder nodes will make linear combinations of the data packets they receive. A destination

node will be able to decode as soon as it has a sufficient amount of linear combinations of packets. Several teams in Digiteo already work together on network coding through a common industrial contract.

Network Information Theory

Network Theory is devoted to the investigation of theoretical foundations aiming at the exploration of performance limits of wireless networks. These involve thee characterization of the theoretical limits in end-to-end traffic transport capacity of wireless networks, relationships with Shannon capacity and multi-terminal information theory, scaling laws characterization as well as the potential impact of autonomic behaviour.

Performance evaluation

It remains one of the key issues for future networks design, architecture and management in particular in the domain of flow and congestion control as well as in traffic analysis. Several Digiteo teams develop new methods (analytical, numerical, approximations or bounds, stochastic network dynamics by algebraic methods) and use them to analyse networks, control algorithms and traffic and to optimise the network design. Typical problems concern the analysis of the interaction of a large number of TCP connections in networks composed of many routers with various characteristics WFQ, FIFO, priorities, RED or the analysis of flow and congestion control.

All Optical Networks:

Optical communication has evolved from high-capacity, point-to-point links toward full optical networking. Wavelength division multiplexing, originally developed to increase the capacity of the installed fibre base, now forms the basis of flexible optical networking for wide area, metropolitan and access networks. All optical networks remove Optic/Electronic/Optic conversion to design very high speed switches operating from Gigabits to Terabits. But the network control plan has to be redesigned for all types of switching (fast circuit, optical burst switching, optical packet switching), as it is not possible to use electronic memory anymore. Typically the routing of packets or bursts is quite different from nowadays routing protocols, which are all based on store and forward strategies. Access methods also need to be modified to be consistent with routing mechanism. Digiteo teams explore this issue via collaborative projects with industry at the medium access level or at the application level with the aim to provide a high speed network to interconnect data and computing grids and large visualisation devices.

Algorithmic aspect for cooperation between selfish operators

The strong competition between operators has raised many interesting problems about cooperation of selfish operators and ISP. They must cooperate to provide a full connectivity but one must add some politics to insure cooperation. In particular pricing, billing and routing strategies have to be designed within this trade-off between competition for customers and cooperation to insure quality of service, mobility and connectivity. This leads to the equilibrium property (Nash) and optimality (Pareto). Game theory, as mentioned in 4.2, is likely to be of major interest on these issues, as well as some optimisation methods mentioned in 4.3.

Radio

Wireless communication systems use dedicated frequency bands. Terminals are multiband (GSM, UMTS, Wifi, Wimax) and multimode (TDMA, CDMA, FDMA). Next generation

terminals will benefit from software radio to dynamically reconfigure radio interfaces at the physical and medium access levels. Teams in Digiteo will address the optimisation problems, the mobility management and the signal and radio measurements methods as well.

All these competences form a unique framework for crossing scientific and technical results, which should favour the development of original and efficient innovations in this area.

4 – Exploratory topics

This section gives some hints on three exploratory subjects. They are of common interest for several of the previous themes. Their list must not be seen as limiting. The goal is not to mount big projects in these areas in the near future, but rather to ensure that there is some competence on these topics in Digiteo and that the Digiteo community is aware of that. It is very likely that during the next years, new subjects will arise.

4.1 Uncertainty

Computerised simulations are used for designing complex systems and for optimising their performances, from a technical, an economical, or even an environmental point of view.
They are now widely used in automotive industry, aeronautics, or energy production. For instance nowadays, simulation tools give access to the design and the production of business aircrafts, without any prototype development. In the energy field, simulations tend to replace expensive thermo-hydraulic tests in order to optimise the reactor use, or even to evaluate more precisely their safety factor by taking into account the ageing structures process…
These computerised simulations are generally based on the evaluation of extremely complex models.
Various sources of uncertainty must be taken into account if these simulations are to be realistic: uncertainty may be due to the model, to inaccurate or partially unknown inputs or outputs, to external perturbations, aging, etc.
Therefore, it is important to develop techniques to characterise and if possible minimise the effect of these sources of uncertainty on design objectives.

In this area, Digiteo partners are very active in addressing various and complementary topics, among which one can cite:

- Error representation, propagation and assessment, for instance via models such as Polynomial chaos, improved Monte-Carlo methods, interval calculus
- Approximate modelling, for instance using surrogate modelling methodology, and characterisation of prediction inaccuracy, for instance via kernel based representation (e.g. Kriging)
- Sensitivity analysis and analysis of variance, using for instance Sobol's indices
- Experiment design to select numerical simulations to reduce the effect of uncertainty
- Estimation of extreme values and percentiles
- Robust optimisation
- Protection against outliers

Two extreme types of models are being considered: black-box models, which can be tackled with non-intrusive methods only, and white or grey box models for which intrusive methodologies such as Galerkin methods and generalised spectral decompositions are promising.
Many classes of problems remain virtually unexplored although with a clear practical interest (among others, hyperbolic problems).

All the problems above are studied in various contexts, with in mind diverse application areas. Digiteo partners have started the organisation of a series of meetings for mutual presentations of problems and solutions in order to encourage as many cross fertilisations as possible.

Complementarities between deterministic and statistical approaches for the characterisation of uncertainty in major calculation codes are also a research area for Digiteo. Partners within Digiteo have the, not so widely held, ability jointly to use numerical and symbolic aspects for model development.

4.2 Probability in modelling, algorithms, and verification

Logical approaches, which have occupied a major place in computer science until the last ten years, have turned out not to scale up well in many cases. Probability, randomness and statistics, which used to be confined to performance evaluations, reliability estimations or machine learning, are now developed in many areas under various forms: given the computation (and storage) power available nowadays, and the number of accessible processors, it is often more efficient to make some well suited random choices than to perform some systematic explorations.
For similar reasons, it is often necessary to observe, model, analyse very complex computerised systems by probabilistic or statistic approaches.
These ideas are currently the bases of many significant advances in areas such as randomised algorithms, communication protocols, web searching, cryptography, validation and verification… These approaches raise difficult theoretical issues but are opening extremely promising perspectives.
In Digiteo, there is a variety of works relying on probability, randomness and statistics for attacking diverse problems such as, to cite just a few:

- randomised algorithms, approximations and complexity
- quantum computing
- formal languages (mainly based on process calculi) for modelling stochastic systems
- probabilistic model checking
- model-based random testing
- network modelling and analysis based on game theory
- statistical pattern recognition, statistical learning
- modelling faults as rare events

It is planned to take advantage of this diversity and to explore possible cross-fertilisations and complementarities between the groups involved in the use of probabilistic and statistic approaches.

Although they make intensive use of similar mathematical concepts, the two topics above are of different nature: the first one, titled “uncertainty”, aims at dealing in a controlled way with those inaccuracies that come from external sources such as measurement errors or approximation in modelling (due to high intricacy of the considered phenomena). The second one, titled “probability in modelling, algorithms, and verification” aims at exploiting approximation and probability in computing methods in order to deal with problems that are known as “hard” by computer scientists.

4.3 - Optimisation

Optimisation problems can be stated as follows: given a set of variables taking values in some given domains (either discrete or finite) and a set of constraints, find an assignment of values to variables that minimises some objective functions.

Optimisation is tremendously useful in real life applications such as routing, scheduling, computational biology, etc. People working in the area of optimisation either focus on building efficient models of real life problems or on specific techniques that improve the performance of solvers. Optimisation is a multidisciplinary research area that uses techniques from algorithmic, applied mathematics, artificial intelligence.

- Algorithmic questions are mostly focused on complexity issues (is the problem NP-Hard?) and on approximability issues (how good are solutions that can be found in polynomial time?).
    - o Hot topic: Efficient algorithms for computing Nash equilibriums
- Applied mathematicians have heavily contributed to the efficiency of mixed integer programming solvers. Over the last 15 years the introduction of efficient cuts, that remove regions of the search space that do not contain optimal solutions, has made it possible to solve large problems.
    - o Hot topic: Semi-definite Programming
- One of the key ideas of constraint programming is that constraints can be used "actively" to reduce the computational effort required to solve combinatorial problems. Constraints are thus not only used to test the validity of a solution, as in conventional programming languages, but also in an active mode to remove values from the domains, deduce new constraints, and detect inconsistencies. Constraint programming solvers have shown to be extremely efficient on some combinatorial problems such as scheduling.
    - o Hot topic: Hybridisation and integration with other optimisation technologies
- Evolutionary Computation and stochastic optimisation are adapted to global optimisation, to multi-criteria optimisation, and to ill-posed optimisation problems, such as those involved in machine learning, data mining, identification, optimal policies, and inverse problems.
    - o Hot topic: Relationship between machine learning and evolutionary computation.

APPENDIX: Digiteo laboratories

Originally, Digiteo was founded by six research and higher order education institutions: CEA, CNRS, École Polytechnique (X), École Supérieure d'Électricité (Supélec), INRIA and Université Paris-Sud XI (UPS). Later on, three new associate members joined Digiteo: Ecole Centrale de Paris, Ecole Normale Supérieure de Cachan (ENS Cachan), and Université de Versailles Saint-Quentin en Yvelines.
The laboratories listed below are involved in totality or partially (when indicated) in Digiteo. They are listed local institution by local institution, but several of them are attached to two, or even in one case, three, institutions. Moreover, some research teams are common with INRIA. These are good signs of the current synergy between institutions, but in some cases, it makes tricky the statement of global figures.
This appendix aims at providing a general idea of the partners of Digiteo. It does not aim at presenting in details the laboratories: for each of them an access to their web site is provided.

Faculty of Science, Université de Paris-Sud, Orsay

These four laboratories are under joint administration with CNRS, and in one case, with Supélec.

• IEF, AXIS department (http://www.u-psud.fr/ief), Unité Mixte CNRS-UPS. In the AXIS Department researches are led in image processing, data fusion, architecture, networks, as well as on embedded micro-systems, sensors, and systems-on-chip.

• LIMSI (http://www.limsi.fr), Unité Propre CNRS with a convention with UPS. There are two departments, one working on fluid mechanics, energetics, acoustics and thermodynamics, the other one on human-machine communication: voice synthesis, spoken language and text processing, vision, virtual reality.

• LRI (http://www.lri.fr), Unité Mixte CNRS-UPS. The research themes addressed by LRI cover a wide spectrum of computer science ranging from fundamental to applied research: algorithms, complexity, quantum computing, graph theory, combinatorial optimisation, clusters and grids, software engineering, programming, human-computer interaction, databases, inference systems, data mining, machine learning, bioinformatics.

• L2S (http://www.lss.supelec.fr), Unité Mixte CNRS-Supélec-UPS. In this laboratory, research work is led on: signals and images considered as media for information transfer, with applications, for instance, in telecommunication; system modelling, analysis of properties of the models, and application for the control of physical systems ; modelling and simulation of the interactions between electromagnetic waves and materials, with applications, for instance, to the localisation of hidden objects.

More than 500 persons are working in these four laboratories. Among them, there are more than 200 Ph.D. students.

Ecole Supérieure d'Electricité

The Ecole Supérieure d'Electricité (SUPELEC) develops, in its own research departments, multidisciplinary decision methods (such as non-destructive evaluation, robust hybrid systems control…). The research activities focus on modelling, optimising and controlling concepts.

Software defined radio; multi-input/multi-output systems optimisation; uncertainties management and propagation within huge simulation codes; robust surrogate models; heterogeneous or hybrid systems models; data mining and information extraction from experimental data; optimum model choice for dedicated inference purposes; data and knowledge engineering; multi-technology micro-systems on-line identification; robust/predictive dynamical system control are some aspects of the conducted research. Applications such as bioinformatics, cognitive radio, and inverse problems have been significantly developed for several years.
SUPELEC, within the four departments that belong to Digiteo (Automatic Control, Computer Science, Signal Processing and Electronic Systems, Telecommunications) gathers 56 researchers and 54 PhD students and 4 post-doctoral colleagues.
http://www.supelec.fr/d2ri/Bienvenue.html

Neighbours

In addition to the above laboratories that are official partners of Digiteo, two neighbour laboratories must be mentioned, since some of their teams are thematically close to several partners, and involved in some projects of Digiteo. They are: LGEP, Unité Mixte CNRS-Supélec-UPS-UP6, for some works on sensors and activities in modelling and simulation, and LCFIO, Unité Mixte CNRS-Institut d'Optique-UPS, for some works on optical systems and sensors and their modelling.

CEA, Saclay centre

• CEA/LIST (http://www-list.cea.fr/) is primarily concerned with the development of technologies that combine software and hardware to form highly integrated complex systems: "software intensive systems". The research activities are structured into industry-driven objectives with a strong project-centred culture, around three major themes:

- Embedded systems: *Computing Systems* (multicore - multiprocessor architectures, reliability of architectures & networks, reconfigurable systems, execution platform, real-time & safety, OS & middleware); *Software Engineering* (methods and tools for software design & validation, static analysis of code, model-driven engineering, co-simulation); *Intelligent Vision Systems* (3D modelling & measure, augmented reality, tracking and detection).
- Interactive systems: *Knowledge Engineering* (multimedia indexing and retrieval, automatic ontology building, question/answer systems based on unstructured data, intelligent indexing of the semantic web); *Robotics* (adaptive, mesoscopic, poly-articulated structure mechanisms, perception, localisation and navigation of robots, remote control and supervision of robotic systems, new actuator technologies); *Virtual reality and sensorial interfaces* (real-time collision algorithms for large-scale models, modelling of deformable, tearable and divisible objects, dynamic simulation and distributed haptic control for collaborative work, dynamic and biomechanical models of virtual persons, gesture recognition, object recognition)
- Sensors and signal processing: *Instrumentation* (non-parametric estimation for the analysis of low data content model signals, spectrum analysis and stochastic algorithms, high-precision dynamic tomography of flow, micro-instrumentation of flow, modelling and simulation of systems, design of monocrystalline diamond-based sensors, bio-sensors); *Non-Destructive Testing* (design of simulation software, simulation of non-homogeneous anisotropic media for ultrasound testing, hybrid models (semi-analytical/digital), non-axisymmetrical NDT based on electromagnetic models, inversion in non-destructive testing, methods of reconstruction and multi-

element imagery, embedded simulation in sensors, shape-hugging flexible sensors, magnetic imager for non-destructive testing, monitoring of noised materials)

CEA LIST gathers 450 researchers, engineers and technicians, and 90 PhD and post doctoral fellows.

• CEA/DEN (http://www.cea.fr/energie)

The Nuclear Energy Division of the French Atomic Energy Commission (CEA/DEN) conducts research activities for the improvement of present nuclear industry, the validation of specific solutions for nuclear waste disposal and the development of nuclear energy for the future. Scientific activities related to Digiteo are conducted in the field of modelling and simulation, especially for coupled multi physics or multi scale problems, including software supervision, uncertainties propagation and optimisation methodologies. Physical fields for application are structural mechanics (static and dynamic), reactor physics (neutronics, thermo hydraulics and thermo mechanics), radiation protection and criticality, gaseous flow transport in geological media, global simulation of complex systems and software engineering.

As such, CEA/DEN develops software platforms based on reusable software components, making them open, flexible and versatile.

Within DEN, about 50 persons (permanent staff, and temporary ones) are involved in activities related to Digiteo.

INRIA

• INRIA Saclay-Ile-de France-Sud (http://www.inria.fr/saclay/). The creation of this new INRIA site took place more or less at the same time as the creation of Digiteo-labs.

Its expertise domains are: distributed computing and grids and networks, software engineering, embedded real-time critical systems, tools for system design, cryptography and security, information systems extended via Internet and information retrieval. The targeted application domains are: bioengineering, health, telecommunication, transportation.

Currently, INRIA-Saclay gathers 300 scientific staff organised in 24 research teams, including 135 Ph.D. students and 45 temporary staff.

• INRIA Rocquencourt (http://www-c.inria.fr/Internet/scientific-research/research/view?set_language=en) 19 teams are involved in Digiteo. In INRIA, research teams are structured around five scientific themes, namely: biological systems, cognitive systems, communicating systems, numerical systems, and symbolic systems. As INRIA Rocquencourt is partner of another RTRA, namely the “Sciences Mathématiques de Paris” one, most teams of the numerical systems theme are in this other RTRA, but the teams on control of complex systems. All the teams of the symbolic systems theme are in Digiteo, as well as most teams of the cognitive systems and communicating systems themes.

These 19 teams gather 112 researchers, 73 administrative and technical staff and 151 Ph.D. students and temporary staff.

Ecole Polytechnique

• LIX (http://www.lix.polytechnique.fr/), CNRS-X. The activities of LIX are organized into three areas: algorithmic, networks, and formal methods. They include research on: bioinformatics, algorithmic and optimisation, combinatorial models, algebraic modelling, cryptology, mobile networks, stochastic modelling of distributed systems, proof processing systems. LIX has a hundred members of which half are PhD students and forty are permanent researchers, equally distributed between CNRS, INRIA, and the Ecole Polytechnique.

• CMAP (www.**cmap**.**polytechnique**.fr/), CNRS-X. They lead research on shape optimisation, topological optimisation, image compression and signal theory, and learning.

• LPICM (http://www.lpicm.polytechnique.fr/), CNRS-X. This laboratory works on modelling and simulation of complex physical systems, sensor clusters, photovoltaic sensors.
Other physics laboratories (LLR, LPHT) lead some works that are at the boundaries of Digiteo scientific perimeter.

Ecole Normale Supérieure de Cachan

• The Farman Institute (www.farman.ens-cachan.fr) at ENS Cachan brings together five research laboratories with the aim to foster multidisciplinary research linked to modelling, simulation, control and validation of complex systems in the field of information, energy and transport. The member laboratories are the CMLA laboratory (www.cmla.ens-cachan.fr) in the field of applied mathematics, the LMT laboratory (www.lmt.ens-cachan.fr) in the field of solid mechanics and structural mechanics, the LSV laboratory (www.lsv.ens-cachan.fr) in the field of computer-aided verification of critical systems, the LURPA (www.lurpa.ens-cachan.fr) laboratory in the field of product design and manufacturing systems and the SATIE laboratory (www.satie.ens-cachan.fr) in the field of electrical engineering. Currently, the Farman Institute gathers nearly 400 people including 140 researchers and 180 doctoral students and 30 post-doctoral students. Research activities concern in particular modelling and simulation of multi-scale phenomena (CMLA, LMT), acquisition, extraction and analysis of experimental data (LURPA, CMLA, LSV, LMT), modelling of materials and structures in their environment (LMT, CMLA, SATIE), modelling of software components and industrial integration (LSV, SATIE, LURPA) and intensive computing (CMLA, LMT).

Ecole Centrale de Paris

• MAS (http://www.mas.ecp.fr/). The research works led in this laboratory cover the modelling, the design and the analysis of complex systems by technologies coming from applied mathematics (numerical analysis, EDP) or from computer science (formal methods, algorithmic, visualisation).
There are three research axes:
- System modelling, optimisation and simulation
- Data processing, extraction and visualisation
- Engineering and architecture of information systems
There are 12 permanent staff and 33 Ph.D. students

Université de Versailles Saint-Quentin

• PRISM (http://www.prism.uvsq.fr/), CNRS-UVSQ.
The staff includes 46 permanent members, 4 administrative staff and 75 doctoral students.
The laboratory has a long record in design, modelling, performance evaluation, and optimisation of computing and networking systems. These themes are studied in the contexts of next generation of the Internet, mobile networks, all-optical networks, systems with very high storage capacity, high performance parallel computer architecture, distributed object oriented applications, databases, bio-information technology, security systems and cryptography.
• LISV (www.lisv.uvsq.fr/), UVSQ. This laboratory is specialised in robotics and addresses different aspects: mobility (humanoid robots); exploration (multi-legged robots, football cup); assistance and handicap.