



RESEARCH CENTER

FIELD

Activity Report 2012

Section Application Domains

Edition: 2013-04-24

1. BYMOORE Exploratory Action (section vide)	9
2. POPIX Exploratory Action	10

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

3. ABSTRACTION Project-Team	12
4. ALF Project-Team	14
5. AOSTE Project-Team	15
6. ARIC Team	17
7. ATEAMS Project-Team (section vide)	18
8. CAIRN Project-Team	19
9. CAMUS Team	20
10. CARMEL Project-Team	21
11. CARTE Project-Team	23
12. CASCADE Project-Team	26
13. CASSIS Project-Team	28
14. CELTIQUE Project-Team (section vide)	30
15. COMETE Project-Team	31
16. COMPSYS Project-Team	32
17. CONTRAINTES Project-Team	33
18. CONVECS Team	35
19. DART Project-Team	36
20. DEDUCTEAM Team	37
21. ESPRESSO Project-Team	38
22. FORMES Team	39
23. GALAAD Project-Team	40
24. GALLIUM Project-Team	41
25. GEOMETRICA Project-Team	43
26. GRACE Team	44
27. LFANT Project-Team	45
28. MARELLE Project-Team (section vide)	46
29. MEXICO Project-Team	47
30. MUTANT Project-Team	49
31. PAREO Project-Team	50
32. PARKAS Project-Team	51
33. PARSIFAL Project-Team	52
34. PI.R2 Project-Team	54
35. POLSYS Project-Team	55
36. POP ART Project-Team	56
37. PROSECCO Project-Team	57
38. S4 Project-Team	58
39. SECRET Project-Team	60
40. SECSI Project-Team	61

41. TASC Project-Team	62
42. TOCCATA Team	63
43. TRIO Project-Team	64
44. TYPICAL Project-Team (section vide)	65
45. VEGAS Project-Team	66
46. VERIDIS Project-Team	67
47. VERTECS Project-Team	68

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

48. ALEA Project-Team	69
49. APICS Project-Team	70
50. ASPI Project-Team	74
51. BACCHUS Team	75
52. BIPOP Project-Team	80
53. CAD Team	82
54. CAGIRE Team	83
55. CALVI Project-Team	85
56. CASTOR Team (section vide)	87
57. CLASSIC Project-Team	88
58. COFFEE Project-Team	89
59. COMMANDS Project-Team	92
60. CONCHA Project-Team	93
61. CORIDA Project-Team	95
62. CQFD Project-Team	97
63. DEFI Project-Team	98
64. DISCO Project-Team	101
65. DOLPHIN Project-Team	102
66. GAMMA3 Project-Team (section vide)	106
67. GECO Team	107
68. GEOSTAT Project-Team	111
69. I4S Team	112
70. IPSO Project-Team	115
71. MATHRISK Team	117
72. MAXPLUS Project-Team	118
73. MC2 Project-Team	122
74. MCTAO Team	124
75. MICMAC Project-Team	128
76. MISTIS Project-Team	131
77. MODAL Project-Team	132
78. NACHOS Project-Team	133
79. NANO-D Team	136
80. NECS Project-Team	140

81. NON-A Project-Team	142
82. OPALE Project-Team	144
83. POEMS Project-Team	146
84. REALOPT Project-Team	148
85. REGULARITY Project-Team	151
86. SCIPORT Team	156
87. SELECT Project-Team	159
88. SEQUEL Project-Team	161
89. SIERRA Project-Team	164
90. SIMPAF Project-Team	165
91. TAO Project-Team	167
92. TOSCA Project-Team	168

COMPUTATIONAL SCIENCES FOR BIOLOGY, MEDICINE AND THE ENVIRONMENT

93. ABS Project-Team (section vide)	170
94. AMIB Project-Team (section vide)	171
95. ASCLEPIOS Project-Team (section vide)	172
96. ATHENA Project-Team	173
97. BAMBOO Project-Team	174
98. BANG Project-Team	175
99. BEAGLE Team	176
100. BIGS Project-Team	177
101. BIOCORE Project-Team	179
102. BONSAI Project-Team	181
103. CARMEN Team	182
104. CLIME Project-Team	183
105. CORTEX Project-Team	185
106. DEMAR Project-Team	186
107. DRACULA Project-Team	187
108. DYLISS Team	193
109. FLUMINANCE Project-Team	194
110. GALEN Team	195
111. GENSCALE Team	196
112. IBIS Project-Team (section vide)	197
113. MACS Project-Team	198
114. MAGIQUE-3D Project-Team	199
115. MAGNOME Project-Team	200
116. MASAIE Project-Team	202
117. MNEMOSYNE Team	203
118. MODEMIC Project-Team	204
119. MOISE Project-Team	205
120. MORPHEME Team (section vide)	208

121. NEUROMATHCOMP Project-Team (section vide)	209
122. NUMED Project-Team	210
123. PARIETAL Project-Team	215
124. POMDAPI Project-Team (section vide)	219
125. REO Project-Team	220
126. SAGE Project-Team	222
127. SERPICO Team	223
128. SHACRA Project-Team	225
129. SISYPHE Project-Team (section vide)	228
130. STEEP Exploratory Action	229
131. VIRTUAL PLANTS Project-Team (section vide)	231
132. VISAGES Project-Team	232

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

133. ACES Project-Team (section vide)	234
134. ADAM Project-Team	235
135. ALGORILLE Project-Team	237
136. ARLES Project-Team	238
137. ASAP Project-Team	239
138. ASCOLA Project-Team	240
139. ATLANMOD Team	242
140. AVALON Team	245
141. CEPAGE Project-Team	247
142. CIDRE Project-Team	252
143. DANTE Team	253
144. DIONYSOS Project-Team (section vide)	254
145. DISTRIBCOM Project-Team	255
146. FOCUS Project-Team	257
147. FUN Team (section vide)	258
148. GANG Project-Team	259
149. GRAND-LARGE Project-Team (section vide)	260
150. HIEPACS Project-Team	261
151. HIPERCOM Project-Team	265
152. INDES Project-Team	268
153. KERDATA Project-Team	269
154. LOGNET Team	271
155. MADYNES Project-Team	272
156. MAESTRO Project-Team	273
157. MASCOTTE Project-Team	274
158. MESCAL Project-Team	275
159. MOAIS Project-Team	276
160. MYRIADS Project-Team	280

161. OASIS Project-Team	281
162. PHOENIX Project-Team	283
163. PLANETE Project-Team	285
164. RAP Project-Team (section vide)	292
165. REGAL Project-Team	293
166. RMOD Project-Team (section vide)	294
167. ROMA Team	295
168. RUNTIME Project-Team	296
169. SARDES Project-Team (section vide)	297
170. SCORE Team	298
171. SOCRATE Team	299
172. TREC Project-Team	302
173. TRISKELL Project-Team	303
174. URBANET Team	304

PERCEPTION, COGNITION, INTERACTION

175. ALICE Project-Team (section vide)	306
176. ALPAGE Project-Team	307
177. AVIZ Project-Team	310
178. AXIS Project-Team	311
179. AYIN Team	314
180. COPRIN Project-Team	315
181. DAHU Project-Team	316
182. DREAM Project-Team	317
183. E-MOTION Project-Team	319
184. EXMO Project-Team	320
185. FLOWERS Project-Team	321
186. GRAPHIK Project-Team	322
187. IMAGINE Team	323
188. IMARA Project-Team	324
189. IMEDIA2 Team	326
190. IN-SITU Project-Team	327
191. LAGADIC Project-Team	328
192. LEAR Project-Team	329
193. MAGRIT Project-Team	330
194. MAIA Project-Team	331
195. MANAO Team (section vide)	333
196. MAVERICK Team	334
197. METISS Project-Team	336
198. MIMETIC Team	340
199. MINT Project-Team	343
200. MORPHEO Team	344

201. MOSTRARE Project-Team	345
202. OAK Team	346
203. ORPAILLEUR Project-Team	347
204. PAROLE Project-Team	349
205. PERCEPTION Team	350
206. POTIOC Team	352
207. PRIMA Project-Team	353
208. REVES Project-Team (section vide)	357
209. SEMAGRAMME Team	358
210. SIROCCO Project-Team	359
211. SMIS Project-Team	362
212. STARS Team	363
213. TEXMEX Project-Team	365
214. VR4I Team	367
215. WAM Project-Team	368
216. WILLOW Project-Team	370
217. WIMMICS Team	371
218. ZENITH Project-Team	373

BYMOORE Exploratory Action (section vide)

POPIX Exploratory Action

4. Application Domains

4.1. Pharmacometrics

Participants: Marc Lavielle, Kevin Bleakley, Maud Delattre, Cyprien Mbogning, Célia Barthelemy, Hector Mesa, Elodie Maillot, Elena Carvajal, Laura Brocco.

POPIX is directly implicated in the domain of pharmacology. Historically, Marc Lavielle was the driving force behind the pharmacological modeling software MONOLIX, now an industry standard. Lixoft, an Inria start-up, now develops and supports MONOLIX and the commercial side of things. POPIX collaborates closely with Lixoft to transfer research results into software improvements and the development of new user tools in MONOLIX.

POPIX is also majorly implicated in the 5-year DDMoRe (Drug and Disease Model Resources) European project financed by the IMI (Innovative Medicines Initiative), a public-private partnership. In particular, POPIX has the task of developing new tools and methods for this project regrouping researchers in pharmacometrics, biostatistics and biology from both the public and private sectors. Specific tools and methods being developed by POPIX include:

- a clinical trial simulator (see the Software section for more details)
- protocol optimization tools
- diagnostic tools
- model selection tools
- estimation techniques for complex models (eg, stochastic differential equations, partial differential equations)

4.2. Pharmacogenetics

Participants: Marc Lavielle, Kevin Bleakley.

Medicine, even when prescribed following dosage rules, is an important cause of illness and death. In essence, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Pharmacogenetics, a subdomain of pharmacology, is the study of the the relationship between genetic variability and the therapeutic outcome. The future goal is "personal medicine" whereby the drug and dose are chosen with respect to the individual's genetic make-up.

Currently, in the population approach followed by POPIX, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much more harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient. The subsequent model-selection problem is thus very complicated, and requires powerful classification and statistical learning tools.

4.3. Oncology

Participants: Marc Lavielle, Célia Barthelemy.

Despite great advances in the treatment and diagnosis of cancer, many steps remain to further improve prognoses and quality of life of cancer patients. Numerical models can be used to help adapt treatment protocol to the characteristics of each patient, ie, improve treatment efficacy by:

- choosing the best treatment
- choosing the best dose
- choosing the best drug-delivery protocol
- optimizing the above parameters to minimize toxicity

POPIX is part of the Inria pluridisciplinary project MONICA (MODèles Numériques et Imagerie pour le Cancer), including the NUMED, MC2 and ASCLEPIOS Inria teams, that aims to optimize the parameters listed above using numerical modeling.

Collaborations with NUMED and MC2 are already underway, in particular with the aim of extending the statistical methods developed by POPIX to partial differential equations based models. NUMED works on models of tumor growth and has already implemented an extension of MONOLIX to KPP-type reaction-diffusion models.

4.4. Change-point detection in signals

Participants: Marc Lavielle, Kevin Bleakley.

Change-point detection is historically a signal processing problem whereby we search for points at which a 1-dimensional noisy signal has an abrupt change in some way, eg. change in mean or variance. It turns out that similar methods can be developed for finding the genomic locations at which the DNA copy number changes in a cancer-stricken (or other) patient. Normally, we have two copies of DNA along the whole genome, so specific changes (gains or losses) in copy number can be associated with the specific cancer, hopefully leading to treatment possibilities. Kevin Bleakley collaborates with researchers at the Curie Institute in change-point detection in one or many DNA copy number profiles. Related to this, Marc Lavielle and Kevin Bleakley have developed methods to find change-points in histogram probability density functions using data sampled from the (unknown) density.

ABSTRACTION Project-Team

4. Application Domains

4.1. Certification of Safety Critical Software

Absence of runtime error, Abstract interpretation, Certified compilation, Static analysis, Translation validation, Verifier.

Safety critical software may incur great damage in case of failure, such as human casualties or huge financial losses. These include many kinds of embedded software, such as fly-by-wire programs in aircrafts and other avionic applications, control systems for nuclear power plants, or navigation systems of satellite launchers. For instance, the failure of the first launch of Ariane 5 (flight Ariane 501) was due to overflows in arithmetic computations. This failure caused the loss of several satellites, worth up to \$ 500 millions.

This development of safe and secure critical software requires formal methods so as to ensure that they do not go wrong, and will behave as specified. In particular, testing, bug finding methods, checking of models but not programs do not provide any guarantee that no failure will occur, even of a given type such as runtime errors; therefore, their scope is limited for certification purposes. For instance, testing can usually not be performed for *all* possible inputs due to feasibility and cost reasons, so that it does not prove anything about a large number of possible executions.

By contrast, program analysis methods such as abstract-interpretation-based static analysis are not subject to unsoundness, since they can *formally prove* the absence of bugs directly on the program, not on a model that might be erroneous. Yet, these techniques are generally incomplete since the absence of runtime errors is undecidable. Therefore, in practice, they are prone to false alarms (*i.e.*, they may fail to prove the absence of runtime errors for a program which is safe). The objective of certification is to ultimately eliminate all false alarms.

It should be noted that, due to the size of the critical codes (typically from 100 to 1000 kLOCs), only scalable methods can succeed (in particular, software model checking techniques are subject to state explosion issues). As a consequence, this domain requires efficient static analyses, where costly abstractions should be used only parsimoniously.

Furthermore, many families of critical software have similar features, such as the reliance on floating-point intensive computations for the implementation of control laws, including linear and non-linear control with feedback, interpolations, and other DSP algorithms. Since we stated that a proof of absence of runtime errors is required, very precise analyses are required, which should be able to yield no false alarm on wide families of critical applications. To achieve that goal, significant advantages can be found in the design of domain specific analyzers, such as **ASTRÉE** [31], [46], which has been initially designed specifically for synchronous embedded software.

Last, some specific critical software qualification procedures may require additional properties being proved. As an example, the DO-178 regulations (which apply to avionics software) require a tight, documented, and certified relation to be established between each development stage. In particular, compilation of high level programs into executable binaries should also be certified correct.

The ABSTRACTION project-team has been working on both proof of absence of runtime errors and certified compilation over the decade, using abstract interpretation techniques. Successful results have been achieved on industrial applications using the **ASTRÉE** analyzer. Following this success, **ASTRÉE** has been licensed to **AbsInt Angewandte Informatik GmbH** to be industrialized, and the ABSTRACTION project-team has strong plans to continue research on this topic.

4.2. Abstraction of Biological Cell Signaling Networks

Biology, Health, Static analysis.

Protein-protein interactions consist in complexations and post translational modifications such as phosphorylation. These interactions enable biological organisms to receive, propagate, and integrate signals that are expressed as proteins concentrations in order to make decisions (on the choice between cell division and cell death for instance). Models of such interaction networks suffer from a combinatorial blow up in the number of species (number of non-isomorphic ways in which some proteins can be connected to each others). This large number of species makes the design and the analysis of these models a highly difficult task. Moreover the properties of interest are usually quantitative observations on stochastic or differential trajectories, which are difficult to compute or abstract.

Contextual graph-rewriting systems allow a concise description of these networks, which leads to a scalable method for modeling them. Then abstract interpretation allows the abstraction of these systems properties. First qualitative abstractions (such as over approximation of complexes that can be built) provide both debugging information in the design phases (of models) and static information that are necessary in order to make other computations (such as stochastic simulations) scale up. Then qualitative invariants also drive efficient quantitative abstractions (such as the reduction of ordinary differential semantics).

The work of the ABSTRACTION project-team on biological cell signaling networks ranges from qualitative abstractions to quantitative abstractions.

ALF Project-Team

4. Application Domains

4.1. Application Domains

Performance, processor architecture, compilers, telecommunications, multimedia, biology, health, engineering, environment, transportation

The ALF team is working on the fundamental technologies for computer science: processor architecture and performance-oriented compilation. The research results have impacts on any application domain that requires high performance executions (telecommunication, multimedia, biology, health, engineering, environment ...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

AOSTE Project-Team

4. Application Domains

4.1. Multicore System-on-Chip design

Synchronous formalisms and GALS or multiclock extensions are natural model representations of hardware circuits at various abstraction levels. They may compete with HDLs (Hardware Description Languages) at RTL and even TLM levels. The main originality of languages built upon these models is to be based on formal *synthesis* semantics, rather than mere simulation forms.

The flexibility in formal Models of Computation and Communication allows specification of modular Latency-Insensitive Designs, where the interconnect structure is built up and optimized around existing IP components, respecting some mandatory computation and communication latencies prescribed by the system architect. This allows a real platform view development, with component reuse and timing-closure analysis. The design and optimization of interconnect fabric around IP blocks transform at modeling level an (untimed) asynchronous versions into a (scheduled) multiclock timed one.

Also, Network on Chip design may call for computable switching patterns, just like computable scheduling patterns were used in (predictable) Latency-Insensitive Design. Here again formal models, such as Cyclo-static dataflow graphs and extended Kahn networks with explicit routing schemes, are modeling elements of choice for a real synthesis/optimization approach to the design of systems.

Multicore embedded architecture platform may be represented as Marte UML component diagrams. The semantics of concurrent applications may also be represented as Marte behavior diagrams embodying precise MoCCs. Optimized compilations/syntheses rely on specific algorithms, and are represented as model transformations and allocation (of application onto architecture).

Our current work aims thus primarily at providing Theoretical Computer Science foundations to this domain of multicore embedded SoCs, with possibly efficient application in modeling, analysis and compilation wherever possible due to some natural assumptions. We also deal with a comparative view of Esterel and SystemC TLM for more practical modeling, and the relation between the Spirit IP-Xact interface standard in SoC domain with its Marte counterpart.

4.2. Automotive and avionic embedded systems

Model-Driven Engineering is in general well accepted in the transportation domains, where design of digital software and electronic parts is usually tightly coupled with larger aspects of system design, where models from physics are being used already. The formalisms **AADL** (for avionics) and **AutoSar** [55] (for automotive) are providing support for this, unfortunately not always with a clean and formal semantics. Thus there is a strong need here for approaches that bring closer together formal methods and tools on the one hand, engineering best practices on the other hand.

From a structural point of view AUTOSAR succeeded in establishing a framework that provides significant confidence in the proper integration of software components from a variety of distinct suppliers. But beyond those structural (interface) aspects, dynamic and temporal views are becoming more of a concern, so that AUTOSAR has introduced the AUTOSAR Specification of Timing Extension. AUTOSAR (discrete) timing models consist of timing descriptions, expressed by events and event chains, and timing constraints that are imposed on these events and event chains.

An important issue in all such formalisms is to mix in a single design framework heterogeneous time models and tasks: based on different timebases, with different triggering policy (event-triggered and time-triggered), and periodic and/or aperiodic tasks, with distinct periodicity if ever. Adequate modeling is a prerequisite to the process of scheduling and allocating such tasks onto complex embedded architectural platforms (see AAA approach in foundation section 3.3). Only then can one devise powerful synthesis/analysis/verification techniques to guide designers towards optimized solutions.

Traceability is also an important concern, to close the gap between early requirements and constraints modelling on the one hand, verification and correct implementation of these constraints at the different levels of the development on the other hand.

ARIC Team

4. Application Domains

4.1. Hardware Arithmetic

The application domains of hardware arithmetic operators are digital signal processing, image processing, embedded applications, reconfigurable computing, and cryptography.

4.2. Floating-point and Validated Numerics

Our expertise on validated numerics is useful to analyse and improve, and guarantee the quality of numerical results in a wide range of applications, from scientific simulation to global optimization or control theory. Much of our work, in particular the development of correctly rounded elementary functions, is critical to the reproducibility of floating-point computations.

4.3. Cryptography, Cryptology, Communication Theory

Lattice reduction algorithms have direct applications in public-key cryptography. A new and promising field of applications is communications theory.

ATEAMS Project-Team (section vide)

CAIRN Project-Team

4. Application Domains

4.1. Panorama

keywords: telecommunications, wireless communications, wireless sensor networks, content-based image retrieval, video coding, intelligent transportation systems, automotive, security

Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

The high complexity of the **Next-Generation (4G) Wireless Communication Systems** leads to the design of real-time high-performance specific architectures. The study of these techniques is one of the main field of applications for our research, based on our experience on WCDMA for 3G implementation.

In **Wireless Sensor Networks (WSN)**, where each wireless node has to operate without battery replacement for a long time, energy consumption is the most important constraint. In this domain, we mainly study energy-efficient architectures and wireless cooperative techniques for WSN.

Intelligent Transportation Systems (ITS), and especially Automotive Systems, more and more apply technology advances. While wireless transmissions allow a car to communicate with another or even with road infrastructure, **automotive industry** can also propose driver assistance and more secure vehicles thanks to improvements in computation accuracy for embedded systems.

Other important fields will also be considered: hardware cryptographic and security modules, specialized hardware systems for the filtering of the network traffic at high-speed, high-speed true-random number generation for security, content-based image retrieval and video processing.

4.2. 4G Wireless Communication Systems

With the advent of the next generation (4G) broadband wireless communications, the combination of MIMO (Multiple-Input Multiple-Output) wireless technology with Multi-Carrier CDMA (MC-CDMA) has been recognized as one of the most promising techniques to support high data rate and high performance. Moreover, future mobile devices will have to propose interoperability between wireless communication standards (4G, WiMax ...) and then implement MIMO pre-coding, already used by WiMax standard. Finally, in order to maximize mobile devices lifetime and guarantee quality of services to consumers, 4G systems will certainly use cooperative MIMO schemes or MIMO relays. Our research activity focuses on MIMO pre-coding and MIMO cooperative communications with the aim of algorithmic optimization and implementation prototyping.

4.3. Wireless Sensor Networks

Sensor networks are a very dynamic domain of research due, on the one hand, to the opportunity to develop innovative applications that are linked to a specific environment, and on the other hand to the challenge of designing totally autonomous communicating objects. Cross-layer optimizations lead to energy-efficient architectures and cooperative techniques dedicated to sensor networks applications. In particular, cooperative MIMO techniques are used to decrease the energy consumption of the communications.

4.4. Multimedia processing

In multimedia applications, audio and video processing is the major challenge embedded systems have to face. It is computationally intensive with power requirements to meet. Video or image processing at pixel level, like image filtering, edge detection and pixel correlation or at block-level such as transforms, quantization, entropy coding and motion estimation have to be accelerated. We investigate the potential of reconfigurable architectures for the design of efficient and flexible accelerators in the context of multimedia applications.

CAMUS Team

4. Application Domains

4.1. Application Domains

Performance being our main objective, our developments' target applications are characterized by intensive computation phases. Such applications are numerous in the domains of scientific computations, optimization, data mining and multimedia.

Applications involving intensive computations are necessarily high energy consumers. However this consumption can be significantly reduced thanks to optimization and parallelization. Although this issue is not our prior objective, we can expect some positive effects for the following reasons:

- Program parallelization tries to distribute the workload equally among the cores. Thus an equivalent performance, or even a better performance, to a sequential higher frequency execution on one single core, can be obtained.
- Memory and memory accesses are high energy consumers. Lowering the memory consumption, lowering the number of memory accesses and maximizing the number of accesses in the low levels of the memory hierarchy (registers, cache memories) have a positive consequence on execution speed, but also on energy consumption.

CAMEL Project-Team

4. Application Domains

4.1. Cryptology

The first application domain for our research is cryptology. This includes cryptography (constructive side) and cryptanalysis (breaking systems). For the cryptanalysis part, although it has practical implications, we do not expect any transfer in the classical sense of the term: it is more directed to governmental agencies and the end-users who build their trust, based on the cryptanalysis effort.

4.1.1. Cryptography

Our cryptographic contributions are related to multiple facets of the large realm of curve-based cryptology. While it is quite clear that enough algorithms exist in order to provide cryptographers with elliptic curves having a suitably hard discrete logarithm (as found in cryptographic standards for instance), one must bear in mind that refinements of the requirements and extensions to curves of higher genus raise several interesting problems. Our work contributes to expanding the cryptographer's capabilities in these areas.

In the context of genus-2 curves, our work aims at two goals. First, improvements on the group law on selected curves yield better speed for the associated cryptosystems. The cryptographic primitives, and then the whole suite of cryptographic protocols built upon such curves would be accelerated. The second goal is the expansion of the set of curves that can be built given a set of desired properties. Using point counting algorithms for arbitrary curves, a curve offering a 128-bit security level, together with nice properties for fast arithmetic, has been computed by CAMEL [3]. Another natural target for construction of curves for cryptography is also the suitability of curves for pairings. We expect to be able to compute such curves.

Implementations of curve-based cryptography, both in hardware and software, are a necessary step on the way to assessing cryptographic speed. We plan to provide such implementations. In particular, on the hardware side, one of our goals is the design of a complete cryptographic coprocessor, including all the primitives for curve-based and pairing-based cryptography, providing optimized and configurable efficiency vs area trade-off.

4.1.2. Cryptanalysis

Our research on cryptanalysis is important for the cryptographic industry: by detecting weak instances, and setting new records we contribute to the definition of recommended families of systems together with their key sizes. The user's confidence in a cryptographic primitive is also related to how well the underlying problem is studied by researchers.

In particular, our involvement in computations with "NFS-like" algorithms encompasses of course the task of assessing the computational limits for integer factorization and discrete-logarithm computations. The impact of the former is quite clear as it concerns the RSA algorithm; record-sized computations attract broad interest and determine updates on key-length recommendations. The latter are particularly important for pairing-based cryptography, since, in this context, one naturally encounters discrete-logarithm problems in extension fields of large degree.

4.2. Standardization

4.2.1. Floating-point arithmetic

The IEEE 754 standard for floating-point arithmetic was revised in 2008. The main new features are some new formats for decimal computations, and the recommendation of correctly rounded transcendental functions. The new decimal formats should not have an impact on our work, since we either use integer-only arithmetic, or arbitrary-precision binary floating-point arithmetic through the GNU MPFR library.

A new standard (P1788) is currently under construction for interval arithmetic. We are not officially involved in this standard, but we follow the discussions, to check in particular that the proposed standard will also cover arbitrary precision (interval) arithmetic.

4.3. Computer algebra systems

Some of our software libraries are being used by computer algebra systems. Most of those libraries are free software, with a license that allows proprietary systems to link them. This gives us a maximal visibility, with a large number of users.

4.3.1. Magma

Magma is a very large computational algebra package. It provides a mathematically rigorous environment for computing with algebraic, number-theoretic, combinatoric, and geometric objects. It is developed in Sydney, by the team around John Cannon. It is non-commercial (in the sense that its goal is not to make profit), but is not freely distributed and is not open-source.

Several members of the team have visited Sydney — a few years ago — to contribute to the development of Magma, by implementing their algorithms or helping in integrating their software. Our link to Magma exists also via the libraries it uses: it currently links GNU MPFR and GNU MPC for its floating-point calculations, and links GMP-ECM as part of its factorization suite.

4.3.2. Pari-GP

Pari/GP is a computational number theory system that is composed of a C library and an interpreter on top of it. It is developed in Bordeaux, where Karim Belabas from the LFANT project-team is the main maintainer. Its license is GPL. Although we do not directly contribute to this package, we have good contact with the developers.

4.3.3. Sage

Sage is a fairly large scale and open-source computer algebra system written in Python. Sage aggregates a large amount of existing free software, aiming at the goal of selecting the fastest free software package for each given task. The motto of Sage is that instead of “reinventing the wheel” all the time, Sage is “building the car”. To date, Sage links GNU MPFR, GMP-ECM, and GNU MPC as standard packages.

CARTE Project-Team

4. Application Domains

4.1. Computer Virology

4.1.1. *The theoretical track.*

It is rightful to wonder why there is only a few fundamental studies on computer viruses while it is one of the important flaws in software engineering. The lack of theoretical studies explains maybe the weakness in the anticipation of computer diseases and the difficulty to improve defenses. For these reasons, we do think that it is worth exploring fundamental aspects, and in particular self-reproducing behaviors.

4.1.2. *The virus detection track.*

The crucial question is how to detect viruses or self-replicating malwares. Cohen demonstrated that this question is undecidable. The anti-virus heuristics are based on two methods. The first one consists in searching for virus signatures. A signature is a regular expression, which identifies a family of viruses. There are obvious defects. For example, an unknown virus will not be detected, like ones related to a 0-day exploit. We strongly suggest to have a look at the independent audit [44] in order to understand the limits of this method. The second one consists in analysing the behavior of a program by monitoring it. Following [46], this kind of methods is not yet really implemented. Moreover, the large number of false-positive implies this is barely usable. To end this short survey, intrusion detection encompasses virus detection. However, unlike computer virology, which has a solid scientific foundation as we have seen, the IDS notion of “malwares” with respect to some security policy is not well defined. The interested reader may consult [65].

4.1.3. *The virus protection track.*

The aim is to define security policies in order to prevent malware propagation. For this, we need (i) to define what is a computer in different programming languages and setting, (ii) to take into consideration resources like time and space. We think that formal methods like rewriting, type theory, logic, or formal languages, should help to define the notion of a *formal immune system*, which defines a certified protection.

4.1.4. *The experimentation track.*

This study on computer virology leads us to propose and construct a “high security lab” in which experiments can be done in respect with the French law. This project of “high security lab” is one of the main project of the CPER 2007-2013.

4.2. Computations and Dynamical Systems

4.2.1. *Continuous computation theories*

Understanding computation theories for continuous systems leads to studying hardness of verification and control of these systems. This has been used to discuss problems in fields as diverse as verification (see e.g. [30]), control theory (see e.g. [37]), neural networks (see e.g. [66]), and so on.

We are interested in the formal decidability of properties of dynamical systems, such as reachability [56], the Skolem-Pisot problem [33], the computability of the ω -limit set [55]. Those problems are analogous to verification of safety properties.

Contrary to computability theory, complexity theory over continuous spaces is underdeveloped and not well understood. A central issue is the choice of the representation of objects by discrete data and its effects on the induced complexity notions. As for computability, it is well known that a representation is gauged by the topology it induces. However more structure is needed to capture the complexity notions: topologically equivalent representations may induce different classes of polynomial-time computable objects. For example, developing a sound complexity theory over continuous structures would enable us to make abstract computability results more applicable by analysing the corresponding complexity issues. We think that the preliminary step towards such a theory is the development of higher-order complexity, which we are currently carrying out.

In contrast with the discrete setting, it is of utmost importance to compare the various models of computation over the reals, as well as their associated complexity theories. In particular, we focus on the General Purpose Analog Computer of Claude Shannon [68], on recursive analysis [74], on the algebraic approach [63] and on computability in a probabilistic context [59].

A crucial point for future investigations is to fill the gap between continuous and discrete computational models. This is one deep motivation of our work on computation theories for continuous systems.

4.2.2. Analysis and verification of adversary systems

The other research direction on dynamical systems we are interested in is the study of properties of adversary systems or programs, i.e. of systems whose behavior is unknown or indistinct, or which do not have classical expected properties. We would like to offer proof and verification tools, to guarantee the correctness of such systems.

On one hand, we are interested in continuous and hybrid systems. In a mathematical sense, a hybrid system can be seen as a dynamical system, whose transition function does not satisfy the classical regularity hypotheses, like continuity, or continuity of its derivative. The properties to be verified are often expressed as reachability properties. For example, a safety property is often equivalent to (non-)reachability of a subset of unsure states from an initial configuration, or to stability (with its numerous variants like asymptotic stability, local stability, mortality, etc ...). Thus we will essentially focus on verification of these properties in various classes of dynamical systems.

We are also interested in rewriting techniques, as a mean to describe dynamical systems, in particular in the adversary context. As they were initially developed in the context of automated deduction, rewriting proof techniques, although now numerous, are not yet adapted to the complex framework of modelling and programming. An important stake in the domain is to enrich them to provide realistic validation tools, both in providing finer rewriting formalisms and their associated proof techniques, and in developing new validation concepts in the adversary case, i.e. when usual properties of the systems like, for example, termination are not verified.

For several years, we have been developing specific procedures for property proofs of rewriting, for the sake of programming, in particular with an inductive technique, already applied with success to termination under strategies [47], [48], [49], to weak termination [50], sufficient completeness [52] and probabilistic termination [51].

The last three results are in the context of adversary computations, since they allow to prove that a program can give the expected results, even when it diverges i.e., even when it has not the usual termination property. A common mechanism has been extracted from the above works, and provides a generic inductive proof framework for properties of reduction relations, which can be parametrized by the property to be proved [53], [54]. Provided that program code can be translated into rule-based specifications, this approach can be applied to correctness proof of software in a larger context.

A crucial element of safety and security of software systems is the problem of resources. We are working in the field of Implicit Computational Complexity. Interpretation based methods like Quasi-interpretations (QI) or sup-interpretations, are the approach we have been developing these last five years, see [60], [61], [62].

Implicit complexity is an approach to the analysis of the resources that are used by a program. Its tools come essentially from proof theory. The aim is to compile a program while certifying its complexity.

CASCADE Project-Team

4. Application Domains

4.1. Hash Functions

Since the previous section just ended on this topic, we start with it for the major problems to address within the next years. A NIST competition on hash functions has been launched late 2007 and finished a few months ago. In the first step, cryptographers had to build and analyze their own candidate; in a second step, cryptanalysts were solicited, in order to analyze and break all the proposals. The conclusion has been announced with the winner Keccak, on October 2nd, 2012.

The symmetric people of the Cascade team have worked these years on the development of a new hash function called SIMD that has been selected for the second round of the NIST SHA-3 competition. SIMD hash function is quite similar to members of the MD/SHA family. It is based on a familiar Merkle-Damgard design, where the compression function is built from a Feistel-like cipher in Davies-Meyer mode. However there are some innovations in this design: the internal state is twice as big as the output size, we use a strong message expansion, and we use a modified feed-forward in the compression function. The main design criteria was to follow the MD/SHA designs principle which are quite well understood, and to add some elements to avoid all known attacks. SIMD is particularly efficient on platforms with vector instructions (SIMD) which are available on many processors. Such instructions have been proposed since 1997 and are now widely deployed. Moreover, it is also possible to use two cores on multicore processors to boost the performances with a factor 1.8 by splitting the message expansion function and the hashing process.

More recently, we essentially worked on the other candidates, with some analyses and attacks. Even if the winner has been selected, there is still a lot of work to do on hash functions, as there is on block-ciphers, even if AES was selected a long time ago.

4.2. Anonymity and Privacy

A relatively new goal of growing importance of cryptography is *privacy*. In a digital world where data is ubiquitous, users are more and more concerned about confidentiality of their personal data. Cryptography makes it possible to benefit from the advantages of digital technology while at the same time providing means for privacy protection. An example is anonymous authentication: A user can convincingly prove that she has certain rights without however revealing her identity. Privacy and anonymity remains thus one of the main challenges for the next years.

4.3. Copyright Protection

Similarly to the privacy concern, the digital world makes easy the large-scale diffusion of information. But in some cases, this can be used in violation of some copyrights. Cryptography should help at solving this problem, which is actually two-fold: one can either mark the original document in order to be able to follow the distribution (and possibly trace the traitor who illegally made it public) or one can publish information in an encrypted way, so that authorized people only can access it.

4.4. Lattice-Based Cryptography

In 1996, Ajtai [66] showed that lattices, which up to that point had only been used as tools in cryptanalysis, can actually be used to *construct* cryptographic primitives. He proposed a cryptographic primitive whose security is based on the worst-case hardness of lattice problems: if one succeeds in breaking the primitive, even with some small probability, then one can also solve any instance of a certain lattice problem. This powerful property makes lattice-based cryptographic constructions very attractive. In contrast, virtually all other cryptographic constructions are based on some average-case assumption. Furthermore, there are currently

very few alternatives to traditional number-theoretic based cryptography such as RSA. Such alternatives will be needed in case an efficient algorithm for factoring integers is ever found, a possibility some leading number theorists consider as quite likely. In fact, efficient quantum algorithms for factoring integers and computing discrete logarithms already exist [87]. Although large-scale quantum computers are not expected to exist for at least a decade, this fact should already be regarded as a warning. In contrast, there are currently no known quantum algorithms for lattice problems. Finally, the computations involved in lattice-based cryptography are typically very fast and often require only modular additions, making them attractive for many applications.

For all these reasons, lattice-based cryptography has become a hot topic, especially in the last few years, and our group is playing an important part in this effort.

4.5. Cryptanalysis

As already explained, even with the *provable security* concept, cryptanalysis is still an important area, and attacks can be done at several levels. Algebraic tools (against integer factoring, discrete logarithm, polynomial multivariate systems, lattice reduction, etc) have thus to be studied and improved in order to further evaluation of the actual security level of cryptographic schemes.

At the hardware level, side-channel information has to be identified (time, power, radiation, noise, heat, etc) in order to securely protect embedded systems. But such information may also be used in a positive way....

4.6. Access Control

Access control policies describe the rights that different users have on the resources available and are used to protect systems against attacks by unauthorised users. Many authorisation and access control models have been proposed so far; we focus mainly on category-based access control. Amongst the most important issues that have to be addressed in this area are the development of techniques and tools for the analysis and verification of access control policies, and the design of transparent and efficient mechanisms to enforce access control policies.

CASSIS Project-Team

4. Application Domains

4.1. Verification of Security Protocols

Security protocols such as SET, TLS and Kerberos, are designed for establishing the confidence of electronic transactions. They rely on cryptographic primitives, the purpose of which is to ensure integrity of data, authentication or anonymity of participants, confidentiality of transactions, etc.

Experience has shown that the design of those protocols is often erroneous, even when assuming that cryptographic primitives are perfect, i.e., that an encoded message cannot be decrypted without the appropriate key. An intruder can intercept, analyze and modify the exchanged messages with very few computations and therefore, for example, generate important economic damage.

Analyzing cryptographic protocols is complex because the set of configurations to consider is very large, and can even be *infinite*: one has to consider any number of sessions, any size of messages, sessions interleaving, some algebraic properties of encryption or data structures.

Our objective is to automatize as much as possible the analysis of protocols starting from their specification. This consists in designing a tool that is easy to use, enables the specification of a large number of protocols thanks to a standard high-level language, and can either look for flaws in a given protocol or check whether it satisfies a given property. Such a tool is essential for verifying existing protocols, but also for helping in designing new ones. For our tool to be easy to use, it has to provide a graphical interface allowing a user to easily perform push-button verification.

Our tools for verifying security protocols are available as components of the AVISPA platform. As an extension of the AVISPA specification language, we are working on a new environment called *CASRUL* for handling more general protocols like e-business protocols for example.

4.2. Automated Boundary Testing from Formal Specifications

We have introduced a new approach for test generation from set-oriented formal specifications: the BZ-TT method. This method is based on Constraint Logic Programming (CLP) techniques. The goal is to test every operation of the system at every boundary state using all input boundary values of that operation. It has been validated in several industry case studies for smart card OS and application validation (GSM 11-11 standard [71] and Java Card Virtual Machine Transaction mechanism [73]), information system and for embedded software [80].

This test generation method can be summed up as follows: from the formal model, the system computes boundary values to create boundary states; test cases are generated by traversal of the state space with a preamble part (sequences of operations from the initial state to a boundary state), a body part (critical invocations), an identification part (observation and oracle state computation) and a post-amble part (return path to initial or boundary state). Then, an executable test script file is generated using a test pattern and a table of correspondence between abstract operations (from the model) and concrete ones. This approach differs in several main points from previous works (e.g. [78]): first, using boundary goals as test objectives avoids the complete construction of the reachability graph; second, this process is fully automated and the test engineer could just drive it at the boundary value computation level or for the path computation.

The BZ-TT method is fully supported by the BZ-Testing-Tools tool-set. This environment is a set of tools dedicated to animation and test cases generation from B, Z or State-Chart formal specifications. It is based on the CLPS constraint solver, able to simulate the execution of the specification. By execution, we mean that the solver computes a so-called constrained state by applying the pre- and post-condition of operations. A constrained state is a constraint store where state variables and also input and output variables support constraints.

One orientation of the current work is to go beyond the finiteness assumption limitations by using symbolic constraint propagation during the test generation process. Second orientation is to extend the result to object oriented specifications as UML/OCL. Third orientation is to extend the coverage of method for security aspect.

4.3. Program Debugging and Verification

Catching bugs in programs is difficult and time-consuming. The effort of debugging and proving correct even small units of code can surpass the effort of programming. Bugs inserted while “programming in the small” can have dramatic consequences for the consistency of a whole software system as shown, e.g., by viruses which can spread by exploiting buffer overflows, a bug which typically arises while coding a small portion of code. To detect this kind of errors, many verification techniques have been put forward such as static analysis and software model checking.

Recently, in the program verification community, there seems to be a growing demand for more declarative approaches in order to make the results of the analysis readily available to the end user. To meet this requirement, a growing number of program verification tools integrate some form of theorem proving.

The goals of our research are twofold. First, we perform theoretical investigations of various combinations of propositional and first-order satisfiability checking in order to automate the theorem proving activity required to solve a large class of program analysis problems which can be encoded as first-order formulae. Second, we experimentally investigate how our techniques behave on real problems in order to make program analysis more precise and scalable. Building tools capable of providing a good balance between precision and scalability is one of the crucial challenges to transfer theorem proving technology to the industrial domains.

4.4. Verification of Web Services

Driven by rapidly changing requirements and business needs, IT systems and applications are undergoing a paradigm shift: components are replaced by services, distributed over the network, and composed and reconfigured dynamically in a demand-driven way into service-oriented architectures. Exposing services in future network infrastructures means a wide range of trust and security issues need to be addressed. Solving them is extremely hard since making the service components trustworthy is not sufficient: composing services leads to new subtle and dangerous vulnerabilities due to interference between component services and policies, the shared communication layer, and application functionality. Thus, one needs validation of both the service components and their composition into secure service architectures. In this context, there is an obvious need of applying formal methods. Our project aims at applying our proof and constraint solving techniques to reason on web services. More precisely, we focus on the composition problem in the presence of security policies.

4.5. Model-Checking of Collaborative Systems

Collaborative systems constitute a class of distributed systems where real human interactions are predominant. In these systems, users at geographically distributed sites interact by simultaneously manipulating shared objects like, text documents, XML trees, filesystems, etc. To improve data availability, the shared objects are replicated so that the users update their local replicas and exchange their updates between them. One of the main challenges here is how to ensure the data consistency when the updates are executed in arbitrary orders at different replicas. Operational Transformation (OT) is an optimistic technique which has been proposed to overcome the consistency problem. This technique consists of an application-dependent protocol to enforce the out-of-order execution of updates even though these updates do not naturally commute. The data consistency relies crucially on the correctness of OT protocols whose proof is extremely hard. Indeed, possibly infinitely many cases should be tested. Our research work aims at applying symbolic model-checking techniques to automatically verify OT protocols. Most importantly, we are interested in finding under which conditions the model-checking problem can be reduced to a finite-state model.

CELTIQUE Project-Team (section vide)

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Miguel Andrés, Nicolás Bordenabe, Konstantinos Chatzikokolakis, Jérémy Dubreil, Catuscia Palamidessi.

The aim of our research is the specification and verification of protocols used in mobile distributed systems, in particular security protocols. We are especially interested in protocols for *information hiding*.

Information hiding is a generic term which we use here to refer to the problem of preventing the disclosure of information which is supposed to be secret or confidential. The most prominent research areas which are concerned with this problem are those of *secure information flow* and of *privacy*.

Secure information flow refers to the problem of avoiding the so-called *propagation* of secret data due to their processing. It was initially considered as related to software, and the research focussed on type systems and other kind of static analysis to prevent dangerous operations, Nowadays the setting is more general, and a large part of the research effort is directed towards the investigation of probabilistic scenarios and treaths.

Privacy denotes the issue of preventing certain information to become publicly known. It may refer to the protection of *private data* (credit card number, personal info etc.), of the agent's identity (*anonymity*), of the link between information and user (*unlinkability*), of its activities (*unobservability*), and of its *mobility* (*untraceability*).

The common denominator of this class of problems is that an adversary can try to infer the private information (*secrets*) from the information that he can access (*observables*). The solution is then to obfuscate the link between secrets and observables as much as possible, and often the use randomization, i.e. the introduction of *noise*, can help to achieve this purpose. The system can then be seen as a *noisy channel*, in the information-theoretic sense, between the secrets and the observables.

We intend to explore the rich set of concepts and techniques in the fields of information theory and hypothesis testing to establish the foundations of quantitative information flow and of privacy, and to develop heuristics and methods to improve mechanisms for the protection of secret information. Our approach will be based on the specification of protocols in the probabilistic asynchronous π -calculus, and the application of model-checking to compute the matrices associated to the corresponding channels.

COMPSYS Project-Team

4. Application Domains

4.1. Compilers for Embedded Computing Systems

The previous sections described our main activities in terms of research directions, but also places Compsys within the embedded computing systems domain, especially in Europe. We will therefore not come back here to the importance, for industry, of compilation and embedded computing systems design.

In terms of application domain, the embedded computing systems we consider are mostly used for multimedia: phones, TV sets, game platforms, etc. But, more than the final applications developed as programs, our main application is the computer itself: how the system is organized (architecture) and designed, how it is programmed (software), how programs are mapped to it (compilation and high-level synthesis).

The industry that can be impacted by our research is thus all the companies that develop embedded systems and processors, and those (the same plus other) than need software tools to map applications to these platforms, i.e., that need to use or even develop programming languages, program optimization techniques, compilers, operating systems. Compsys do not focus on all these critical parts, but our activities are connected to them.

CONTRAINTES Project-Team

4. Application Domains

4.1. Combinatorial optimization

The number and economic impact of combinatorial optimization problems found in the industrial world are constantly increasing. They cover:

- resource allocation;
- placement, bin packing;
- scheduling;
- planning;
- transport;
- etc.

The last fifty years have brought many improvements in Operations Research resolution techniques. In this context, Constraint Programming can be seen as providing, on the one hand, constraint propagation algorithms that can be applied to various numerical or symbolic constraints, and on the other hand, declarative languages to model real-life problems and express complex resolution strategies. The latter point is crucial for designing new algorithms that cannot be defined without a sufficiently high-level language to express them. It allowed for better results than traditional methods, for instance in scheduling, and is promised to an even better future when thinking about the cooperation of global resolution, local consistency techniques and search methods.

The European FP6 Strep project **Net-WMS** that we have coordinated, has shown the benefit of combining discrete geometry constraints with rules to express physical, common sense and packing business constraints to solve packing problems in the context of warehouse management systems for the automotive industry. In this context, we have developed a rule-based modeling language, called **Rules2CP**, to express requirements in a declarative and flexible manner, and compile them to efficient constraint programs using reified constraints and a global constraint dedicated to geometrical placement problems in high dimension.

4.2. Computational Systems Biology

In partnership with biologists, we develop and experiment our modeling methods in five main leading applications:

- **Cancer chronotherapy optimization.** This research initiated in 2004 in partnership with Jean Clairambault, EPI BANG, and Francis Lévi INSERM, Hopital Paul Brousse, Villejuif, aims at understanding fundamental mechanisms involved in cancer and chronotherapies through mathematical modeling. Following the EU STREP project (2006-2009) on “temporal genomics for patient tailored chronotherapeutics”, coordinated by Francis Lévi, and in the framework of the Era-Net SysBio **C5Sys** project (2010-2013) coordinated by Francis Lévi and David Rand, University of Warwick, UK, we develop coupled models of the cell cycle, the circadian clock, the DNA repair system, irinotecan metabolism and drug injection optimization, focussing on the interactions between the cell cycle and the circadian clock in mammalian cells.
- **Mammalian cell cycle regulation.** This theme that is closely related to the previous one has lead to a formal collaboration in the framework of the ANR Syscomm project **CALAMAR**, started in 2009 on the “Compositional modeling and Analysis of LArge Molecular Regulatory networks”. In partnership with Claudine Chaouiya, TAGC INSERM, Marseille, and Laurence Calzone, Institut Curie, Paris, this project aims at applying our computational techniques – both qualitative and quantitative – to the analysis of the large scale RB/E2F network, in order to elucidate various features of the human cell proliferation, especially in the case of healthy and bladder-tumor cells of different aggressiveness.

- **G-protein coupled receptor signal transduction.** This research initiated in 2004 in partnership with Eric Reiter, INRA Tours, and Frédérique Clément, EPI SISYPHE, aimed at understanding the structure and the dynamics of the follicle stimulating hormone (FSH) and angiotensine signal transduction in mammalian cells. It was first conducted in the INRA AgroBi project **INSIGHT** (2006-2009) and in the AE **REGATE**.

The article [4] concludes our fruitful collaboration over this period of eight years, with a tightly coupled formal and experimental study of GPCR signaling, of particular importance in medicine since these receptors are the most common drug target.

- **Real-time control of gene expression in yeast.** This research lead in the team by Grégory Batt investigates the possibilities to control gene expression in living cells. In collaboration with Pascal Hersen and Samuel Bottani, biophysicists at the Matière and Systèmes Complexes lab, CNRS/Paris Diderot University, we develop a microfluidic platform and control software for the real-time control of gene expression in yeast. In a larger initiative, we consider a similar problem but in mammalian cells, where the stochasticity of gene expression makes the control problem particularly challenging. The Iceberg Investissement d'Avenir project, coordinated by Grégory Batt, involves the MSC, BM2A, LIFL and PPS labs, and the Jacques Monod Institut. Similarly, the Contraintes research group is also involved in the Inria/INSERM large-scale initiative action **COLAGE** coordinated by Huges Berry, EPI COMBINING, with François Taddei, Ariel Lindner, INSERM Paris Necker, Hidde de Jong, Delphine Ropers, EPI IBIS, Jean-Luc Gouzé, and Madalena Chaves, EPI COMORE. In this project, we investigate the possibilities to control and reprogram growth and aging in bacteria *E. coli* using synthetic biology approaches.
- **Artificial tissue homeostasis in mammalian cells.** Artificial tissue design is a particularly challenging problem in synthetic biology since the system behavior results from the interplay between intra- and intercellular dynamics. In the framework of the **Syne2arti** ANR project, coordinated by Grégory Batt, and involving Dirk Draso, EPI BANG, Oded Maler, CNRS Verimag, and Ron Weiss, MIT, USA, we design and genetically-engineer mammalian cells to obtain a tissue having a desired cell density. The long-term correct functioning of the system relies several key aspects, including individual cell decisions, collective, spatial aspects, and cell-to-cell variability.
- **TGF β signaling and initiation of translation in sea urchin.** In the framework of the **BioTempo** ANR project, we recently started to apply the different algorithms available in the **BIOCHAM** platform to the modeling of the TGF β signaling network in collaboration with the SeRAIC lab (Rennes, France) and of the sea urchin's initiation of translation with Laboratoire Mer et Santé (Roscoff, France). In the first case, the main challenge is to compare and understand crosstalks between the SMAD-dependent fast pathway and the MAPK-dependent slower pathway that is often related to cancer. In the second case there is a whole issue of parametrization even for small models since the data is quite sparse. The different parameter learning features of BIOCHAM, notably based on temporal logics, are therefore put to good use.

CONVECS Team

4. Application Domains

4.1. Application Domains

The theoretical framework we use (automata, process algebras, bisimulations, temporal logics, etc.) and the software tools we develop are general enough to fit the needs of many application domains. They are applicable to virtually any system or protocol that consists of distributed agents communicating by asynchronous messages. The list of recent case studies performed with the CADP toolbox (see in particular § 6.5) illustrates the diversity of applications:

- *Bioinformatics*: genetic regulatory networks, nutritional stress response, metabolic pathways,
- *Consumer electronics*: home networking, video on-demand,
- *Databases*: transaction protocols, distributed knowledge bases, stock management,
- *Distributed systems*: virtual shared memory, distributed file systems, election algorithms, dynamic reconfiguration algorithms, fault tolerance algorithms, cloud computing,
- *Embedded systems*: smart-card applications, air traffic control, avionic systems,
- *Hardware architectures*: asynchronous circuits, multiprocessor architectures, systems on chip, networks on chip, bus arbitration protocols, cache coherency protocols, hardware/software codesign,
- *Human-machine interaction*: graphical interfaces, biomedical data visualization, plasticity,
- *Security protocols*: authentication, electronic transactions, cryptographic key distribution,
- *Telecommunications*: high-speed networks, network management, mobile telephony, feature interaction detection.

DART Project-Team

4. Application Domains

4.1. Gaspard2 for avionic hybrid test platform design

The emergence and the maturity of FPGA circuits for distributed and reconfigurable architectures offer the opportunity to explore real time problems in the field of avionic systems. FPGA becomes de facto a major processing element as same as general CPUs. As of now, the FPGA is widely used in the field of I/O component in order to connect the real equipment with the CPU host. Among the main features mapped into the FPGA in the original architecture, we quote the fast serial link and RAM IPs (Intellectual property) which are needed to ensure communication between CPU and FPGA. Additionally, the Base Time IP is needed for the global system synchronization. This minimal configuration based on FPGA can be duplicated several times and connected together to build bigger test system or a complete simulator. Eurocopter expectation for the above-described architecture is to prototype some models which can be eligible and relocated in the FPGA. The objective is to increase the performances of these models and to reduce the communication latencies by the means of embedding the different parts in the same chip. To do so, we studied in this first year a real avionic test loop in order to extract the complex models that will be implemented in the FPGA. Different hardware model configurations have been explored to reach an optimal well-balanced global system using the ML403 Virtex-4 Xilinx board. Different tradeoffs in terms of performance and resource occupation in the FPGA are obtained. Later, these results will be used for dynamically adapt the system functioning according to the available resources and performance requirements.

As a second part, we used the MARTE profile to represent an hybrid system (CPU/FPGA). In the MARTE specification, an application is a set of tasks connected through ports. Tasks are considered as mathematical functions reading data from their input ports and writing data on their output ports. This specification has been used to model the avionic test loop. In addition, MARTE allows describing the hardware architecture in a structural way. Typical components such as HwProcessor, HwFPGA and HwRAM can be specified with their non-functional properties. We used this subset of MARTE in order to represent an hybrid multiprocessor architecture. The main component of this architecture is composed of the Xeon-X3370 processor (multicore CPU) and the Virtex-4 Xilinx FPGA. Furthermore, MARTE provides the Allocate concept as well as the concept specially crafted for repetitive structures Distribute. This latter concept gives a way to express regular distribution of tasks onto a set of processors or FPGA resources. The mapping step relies on two types of distribution (timeScheduling and spatialDistribution) depending on the target hardware platform (CPU/FPGA). The different models of our avionic test loop can be mapped onto the host multicore processor, the embedded processor (Microblaze) or the hardware resources in the FPGA.

4.2. Electromagnetic modeling

We collaborate with the L2EP specialized in electromagnetic modeling, on algorithms definition and the parallelization of their computations, especially on GPUs.

For the first point, we have designed a parallel version of the Finite Integration Technique (F.I.T). This is used to simulate electromagnetic phenomena. This technique is efficient if the mesh is generated by a regular hexahedron. Moreover the matrix system, obtained from a regular mesh can be exploited to use the parallel direct solver. In fact, in reordering the unknowns by the nested dissection method, it is possible to construct directly the lower triangular matrix with many processors without assembling the matrix system. During this year, we have used our parallel direct solver as a preconditionner for a sparse linear system coming from a FEM problem with a good efficiency[25].

For the second point, we have include our Gaspard2 generated code in Code_CARMEL, a software for electromagnetic fields simulations. This GPGPU code is now robust enough to run most of the testbenches implemented inside this framework[23].

DEDUCTEAM Team

3. Application Domains

3.1. Safety of Aerospace systems

In parallel with this effort in logic and in the development of proof checkers and automated theorem proving systems, we always have been interested in using such tools. One of our favorite application domain is the safety of aerospace systems. Together with César Muñoz' team in Nasa-Langley, we have proved the correctness of several geometric algorithms used in air traffic control.

This has led us sometimes to develop such algorithms ourselves, and sometimes to develop tools for automating these proofs.

3.2. B-set theory

The B method allows the user to develop software correct by construction, going from abstract models to implementations via refinement. During the development process, proof obligations are generated. The formalism underlying B is based on predicate logic and B-set theory. Atelier B that supports the B method provides interactive and automatic provers. To increase automation the user may add proof rules which, if not correct, may corrupt the process. Siemens has developed a tool chain to verify such added proof rules. In particular we have to verify that any proof rule derives from B logic. This step has to be as automatic as possible. Furthermore confidence in these verification proofs is required. A first attempt using the first order prover Zenon allowed the verification of a large number of proof rules [10]. To go further we have experimented techniques such as super deduction and deduction modulo. B-set theory is an interesting benchmark for the tools developed by Deducteam since this theory contains numerous operators and predicates defined by equations or rewrite rules.

ESPRESSO Project-Team

4. Application Domains

4.1. Embedded systems

The application domains covered by the Polychrony toolbox are engineering areas where a system design-flow requires high-level model transformations and verifications to be applied during the development-cycle. The project-team has focused on developing such integrated design methods in the context of avionics applications, through the European IST projects Sacres, Syrf, Safeair, Speeds, and through the national ANR projects Topcased, OpenEmbeDD, Spacify. In this context, Polychrony is seen as a platform on which the architecture of an embedded system can be specified from the earliest design stages until the late deployment stages through a number of formally verifiable design refinements.

Along the way, the project adopted the policy proposed with project Topcased and continued with OpenEmbeDD to make its developments available to a large community in open-source. The Polychrony environment is now integrated in the OPEES/Polarsys platform and distributed under EPL and GPL v2.0 license for the benefits of a growing community of users and contributors, among which the most active are Virginia Tech's Fermat laboratory and Inria's project-teams Aoste, Dart.

FORMES Team

4. Application Domains

4.1. Simulation

Simulation is relevant to most areas where complex embedded systems are used, not only to the semiconductor industry for System-on-Chip modeling, but also to any application where a complex hardware platform must be assembled to run the application software. It has applications for example in industry automation, digital TV, telecommunications and transportation.

4.2. Certified Compilation for Embedded systems

Many frameworks have been designed in order to make the design and the development of embedded systems more rigorous and secure on the basis of some formal model. All these frameworks implicitly assume the *reliability of the translation* to executable code, in order to guarantee the verified properties in the design level are preserved in the implementation. In other words, they rely on a claim saying that the compilers from high level model description to the implementation perfectly will not introduce undesired behaviors or errors in silence. The only safe way to satisfy such a claim is to certify correctness of the compilers, that is, to prove that the code they produce has exactly the semantics of the source code or model.

4.3. Distributed Systems

Many embedded systems run in a distributed environment. Distributed systems raise extremely challenging issues, both for the design and the implementation, because decisions can be made only from a local knowledge, which is imperfect due to communication time and unreliability of transmissions.

4.4. Security

The convergence between embedded technologies and the Internet offers many opportunities to malicious people for breaking the privacy of consumers or of organisations. Using cryptography is not enough for ensuring the protection of data, because of possible flaws in protocols and interfaces, providing opportunities for many well-known attacks. This area is therefore an important target of formal methods.

GALAAD Project-Team

4. Application Domains

4.1. Shape modeling

Geometric modeling is increasingly familiar for us (synthesized images, structures, vision by computer, Internet, ...). Nowadays, many manufactured objects are entirely designed and built by means of geometric software which describe with accuracy the shape of these objects. The involved mathematical models used to represent these shapes have often an algebraic nature. Their treatment can be very complicated, for example requiring the computations of intersections or isosurfaces (CSG, digital simulations, ...), the detection of singularities, the analysis of the topology, etc. Optimising these shapes with respect to some physical constraints is another example where the choice of the models and the design process are important to lead to interesting problems in algebraic geometric modeling and computing. We propose the developments of methods for shape modeling that take into account the algebraic specificities of these problems. We tackle questions whose answer strongly depends on the context of the application being considered, in direct relationship with the industrial contacts that we are developing in Computer Aided Geometric Design.

4.2. Shape processing

Many problems encountered in the application of computer sciences start from measurement data, from which one wants to recover a curve, a surface, or more generally a shape. This is typically the case in image processing, computer vision or signal processing. This also appears in computer biology where the geometry of distances plays a significant role, for example, in the reconstruction from NMR (Nuclear Magnetic Resonance) experiments, or the analysis of realizable or accessible configurations. In another domain, scanners which tend to be more and more easily used yield large set of data points from which one has to recover compact geometric model. We are working in collaboration with groups in agronomy on the problems of reconstruction of branching models (which represent trees or plants). We are investigating the application of algebraic techniques to these reconstruction problems. Geometry is also highly involved in the numerical simulation of physical problems such as heat conduction, ship hull design, blades and turbines analysis, mechanical stress analysis. We apply our algebraic-geometric techniques in the isogeometric approach which uses the same (bspline) formalism to represent both the geometry and the solutions of partial differential equations on this geometry.

GALLIUM Project-Team

4. Application Domains

4.1. High-assurance software

A large part of our work on programming languages and tools focuses on improving the reliability of software. Functional programming and static type-checking contribute significantly to this goal.

Because of its proximity with mathematical specifications, pure functional programming is well suited to program proof. Moreover, functional programming languages such as Caml are eminently suitable to develop the code generators and verification tools that participate in the construction and qualification of high-assurance software. Examples include Esterel Technologies's KCG 6 code generator, the Astrée static analyzer, the Caduceus/Jessie program prover, and the Frama-C platform. Our own work on compiler verification combines these two aspects of functional programming: writing a compiler in a pure functional language and mechanically proving its correctness.

Static typing detects programming errors early, prevents a number of common sources of program crashes (null references, out-of bound array accesses, etc), and helps tremendously to enforce the integrity of data structures. Judicious uses of generalized abstract data types (GADTs), phantom types, type abstraction and other encapsulation mechanisms also allow static type checking to enforce program invariants.

4.2. Software security

Static typing is also highly effective at preventing a number of common security attacks, such as buffer overflows, stack smashing, and executing network data as if it were code. Applications developed in a language such as Caml are therefore inherently more secure than those developed in unsafe languages such as C.

The methods used in designing type systems and establishing their soundness can also deliver static analyses that automatically verify some security policies. Two examples from our past work include Java bytecode verification [48] and enforcement of data confidentiality through type-based inference of information flows and noninterference properties [51].

4.3. Processing of complex structured data

Like most functional languages, Caml is very well suited to expressing processing and transformations of complex, structured data. It provides concise, high-level declarations for data structures; a very expressive pattern-matching mechanism to destructure data; and compile-time exhaustiveness tests. Languages such as CDuce and OCamlDuce extend these benefits to the handling of semi-structured XML data [44]. Therefore, Caml is an excellent match for applications involving significant amounts of symbolic processing: compilers, program analyzers and theorem provers, but also (and less obviously) distributed collaborative applications, advanced Web applications, financial modeling tools, etc.

4.4. Rapid development

Static typing is often criticized as being verbose (due to the additional type declarations required) and inflexible (due to, for instance, class hierarchies that must be fixed in advance). Its combination with type inference, as in the Caml language, substantially diminishes the importance of these problems: type inference allows programs to be initially written with few or no type declarations; moreover, the OCaml approach to object-oriented programming completely separates the class inheritance hierarchy from the type compatibility relation. Therefore, the Caml language is highly suitable for fast prototyping and the gradual evolution of software prototypes into final applications, as advocated by the popular "extreme programming" methodology.

4.5. Teaching programming

Our work on the Caml language has an impact on the teaching of programming. Caml Light is one of the programming languages selected by the French Ministry of Education for teaching Computer Science in *classes préparatoires scientifiques*. OCaml is also widely used for teaching advanced programming in engineering schools, colleges and universities in France, USA, and Japan.

GEOMETRICA Project-Team

4. Application Domains

4.1. Geometric Modeling and Shape Reconstruction

Modeling 3D shapes is required for all visualization applications where interactivity is a key feature since the observer can change the viewpoint and get an immediate feedback. This interactivity enhances the descriptive power of the medium significantly. For example, visualization of complex molecules helps drug designers to understand their structure. Multimedia applications also involve interactive visualization and include e-commerce (companies can present their products realistically), 3D games, animation and special effects in motion pictures. The uses of geometric modeling also cover the spectrum of engineering, computer-aided design and manufacture applications (CAD/CAM). More and more stages of the industrial development and production pipeline are now performed by simulation, due to the increased performance of numerical simulation packages. Geometric modeling therefore plays an increasingly important role in this area. Another emerging application of geometric modeling with high impact is medical visualization and simulation.

In a broad sense, shape reconstruction consists of creating digital models of real objects from points. Example application areas where such a process is involved are Computer Aided Geometric Design (making a car model from a clay mockup), medical imaging (reconstructing an organ from medical data), geology (modeling underground strata from seismic data), or cultural heritage projects (making models of ancient and or fragile models or places). The availability of accurate and fast scanning devices has also made the reproduction of real objects more effective such that additional fields of applications are coming into reach. The members of GEOMETRICA have a long experience in shape reconstruction and contributed several original methods based upon the Delaunay and Voronoi diagrams.

4.2. Scientific Computing

Meshes are the basic tools for scientific computing using finite element methods. Unstructured meshes are used to discretize domains bounded by complex shapes while allowing local refinements. GEOMETRICA contributes to mesh generation of 2D and 3D possibly curved domains. Most of our methods are based upon Delaunay triangulations, Voronoi diagrams and their variants. Anisotropic meshes are also investigated. We investigate in parallel both greedy and variational mesh generation techniques. The greedy algorithms consist of inserting vertices in an initial coarse mesh using the Delaunay refinement paradigm, while the variational algorithms consists of minimizing an energy related to the shape and size of the elements. Our goal is to show the complementarity of these two paradigms. Quadrangle surface meshes are also of interest for reverse engineering and geometry processing applications. Our goal is to control the final edge alignment, the mesh sizing and the regularity of the quadrangle tiling.

GRACE Team

3. Application Domains

3.1. Cryptology

We want to establish the security of practical proposals relying on computational problems, be they standardized (like RSA or Elliptic Curve Cryptography), or more exotic (like Hyperelliptic Curve Cryptography). We do not work with abstract cryptographic primitives. On the design side, building efficient near-optimal codes impacts directly on the security of basic operations in symmetric primitives. We also investigate other applications, such as secret sharing schemes, universal hash functions, and message authentication, revisiting them in the context of Algebraic Geometry codes.

3.2. Codes in Computer Science

We do not want to do basic forward error correction, dealing with bit error rates and signal-to-noise ratios. Rather, we aim to deal with higher models of communication and computation, including peer-to-peer systems and distributed storage. We also consider adversarial noise, or distributed computations with byzantine faults. List decoding deals precisely with these kinds of “difficult”, non-random errors. In a related spirit, one can deal with “computationally bounded channels”, where the errors are generated by an adversarial machine or algorithm that is computationally bounded.

LFANT Project-Team

4. Application Domains

4.1. Number theory

Being able to compute quickly and reliably algebraic invariants is an invaluable aid to mathematicians: It fosters new conjectures, and often shoots down the too optimistic ones. Moreover, a large body of theoretical results in algebraic number theory has an asymptotic nature and only applies for large enough inputs; mechanised computations (preferably producing independently verifiable certificates) are often necessary to finish proofs.

For instance, many Diophantine problems reduce to a set of Thue equations of the form $P(x, y) = a$ for an irreducible, homogeneous $P \in \mathbb{Z}[x, y]$, $a \in \mathbb{Z}$, in unknown integers x, y . In principle, there is an algorithm to solve the latter, provided the class group and units of a rupture field of P are known. Since there is no other way to prove that the full set of solutions is obtained, these algebraic invariants must be computed and certified, preferably without using the GRH.

Deeper invariants such as the Euclidean spectrum are related to more theoretical concerns, e.g., determining new examples of principal, but not norm-Euclidean number fields, but could also yield practical new algorithms: Even if a number field has class number larger than 1 (in particular, it is not norm-Euclidean), knowing the upper part of the spectrum should give a *partial* gcd algorithm, succeeding for almost all pairs of elements of \mathcal{O}_K . As a matter of fact, every number field which is not a complex multiplication field and whose unit group has rank strictly greater than 1 is almost norm-Euclidean [37], [38].

Algorithms developed by the team are implemented in the free PARI/GP system for number theory maintained by K. Belabas, which is a reference and the tool of choice for the worldwide number theory community.

4.2. Cryptology

Public key cryptology has become a major application domain for algorithmic number theory. This is already true for the ubiquitous RSA system, but even more so for cryptosystems relying on the discrete logarithm problem in algebraic curves over finite fields [7]. For the same level of security, the latter require smaller key lengths than RSA, which results in a gain of bandwidth and (depending on the precise application) processing time. Especially in environments that are constrained with respect to space and computing power such as smart cards and embedded devices, algebraic curve cryptography has become the technology of choice. Most of the research topics of the LFANT team concern directly problems relevant for curve-based cryptology: The difficulty of the discrete logarithm problem in algebraic curves determines the security of the corresponding cryptosystems. Complex multiplication, point counting and isogenies provide, on one hand, the tools needed to create secure instances of curves. On the other hand, isogenies have been found to have direct cryptographic applications to hash functions [39] and encryption [43]. Pairings in algebraic curves have proved to be a rich source for novel cryptographic primitives. Class groups of number fields also enter the game as candidates for algebraic groups in which cryptosystems can be implemented. However, breaking these systems by computing discrete logarithms has proved to be easier than in algebraic curves; we intend to pursue this cryptanalytic strand of research.

Apart from solving specific problems related to cryptology, number theoretic expertise is vital to provide cryptologic advice to industrial partners in joint projects. It is to be expected that continuing pervasiveness and ubiquity of very low power computing devices will render the need for algebraic curve cryptography more pressing in coming years.

MARELLE Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Panorama

telecommunications, multimedia, transportation systems, web services

MExiCo's research is motivated by problems on system management in several domains:

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize *adaptators* for every component in order to steer its interaction with the surrounding processes.
- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

This list is likely to grow over the next years as we continue our research.

4.2. Autonomous Telecommunications Systems: In-Band Supervision

Participants: Stefan Haar, Serge Haddad.

In the context of traditional hard-wired communication networks, supervision structures for managing faults, configuration, provisioning etc could be developed with a fixed infrastructure, and perform the communication between sensors, supervisors, policy enforcement points etc over a separate network using separate hardware. This rigid, **out-of-band** technology does not survive passing to today's and tomorrow's services and networks. In fact, the dynamic mobility of services combined across sites and domains cannot be captured unless the network used for supervision evolves in the same way and simultaneously, which rules out static solutions; but providing out-of-band infrastructure that grows with the networks to be supervised would be prohibitively expensive, if at all technically feasible. *Heterogeneity* is the other feature of modern networks that forces a change, since different domains are not likely to agree on a pervasive third-party supervision. Rather, the providers will keep control over the internal state and evolution of their domain, and accept only exchange through standardized outward interfaces.

Supervision has thus to be re-invented on an *in-band, autonomous* base: monitoring probes deployed on the web, dysfunctions on one peer node diagnosed by another peer in a network with changing configuration, enhanced supervisor and actor capacities of services, etc. *MExiCo* will work on improving the interoperability of service components through continued application of e.g. distributed techniques for control and diagnosis.

4.3. Web Services

Participants: Stefan Haar, Serge Haddad.

Specific applications targeted by *MExiCo* include the problem of adaptation in Service-Oriented Computing (SOC). The challenge is here twofold, stemming both from the distributed nature of services (scattered over the entire web) and their heterogeneous origins.

4.3.1. Context

Web services have become the most frequently used model of design and programming based on components for business applications. Web service languages like BPEL have useful constructors that manage for instance exceptions, (timed guarded) waiting of messages, parallel execution of processes, distant service invocations, etc. Interoperability of components is based on interaction protocols associated with them and often published on public or private registers. In the framework of Web services, these protocols are called abstract processes by contrast with business processes (i.e. services). Composition of components must be analyzed for several reasons and at least to avoid deadlocks during execution. This has led to numerous works that focus on compositional verification, substitution of a component by another one, synthesis of adaptators, etc., and triggered a push towards a unifying theoretical framework (see e.g. [115], [119])

4.3.2. Problems

Interoperability requires that when a user or a program wants to interact with the component, the knowledge of the interaction protocol is enough. Our previous works have shown that the interaction protocols can be inherently ambiguous: no client can conduct a correct interaction with the component in every scenario. This problem is even more complex when the protocol can evolve during execution due to adaptation requirements. The composition of components also raises interesting problems. When composing optimal components (w.r.t. the number of states for instance) the global component can be non optimal. So one aims at reducing a posteriori or better on the fly the global component. At last, the dynamical insertion of a component in a business process requires to check whether this insertion is behaviorally consistent [121], [109]

We do not intend to check global properties based on a modular verification technique. Rather, given an interaction protocol per component and a global property to ensure, we want to synthesize an adaptator per component such that this property is fulfilled or to detect that there cannot exist such adaptators [106]. In another research direction, one can introduce the concept of utility of a service and then optimize a system i.e. keeping the same utility value while reducing the resources (states, transitions, clocks, etc.).

MUTANT Project-Team

4. Application Domains

4.1. Application Domains

- **Authoring and Performing Interactive Music.** The combination of both realtime machine listening systems and reactive programming paradigms has enabled the *authoring* of interactive music systems as well as their realtime performance within a coherent synchronous framework called *Antescofo*. The module, developed since 2008 by the team members, has gained increasing attention within the user community worldwide with more than 30 prestigious public performances yearly. The outcomes of the proposed research will enhance the interactive and reactive aspects of this emerging paradigm as well as creating novel authoring tool for such purposes. The outcome of the **ANR Project INEDIT** (with LABRI and GRAME and coordinated by team leader), will further extend the use-cases of *Antescofo* for interactive multimedia pieces with more complex temporal structures and computational paradigms.
- **Music Post-Production.** Outcomes of our recognition and alignment paradigms can improve and ease existing workflows employed by audio engineers for mixing and editing using commercial Digital Audio Workstations (DAW) in post-production. We have recently initiated collaborations with audio engineers at Ircam and Paris Superior Music Conservatory (CNSMDP) to define the framework [8] and we will continue to develop and integrate our tools into their daily workflow.
- **Realtime Music Information Retrieval** We will apply our information geometric approach to well-known and complex MIR problems. A glance of such problems is presented in [5]. Such applications can be used as front-end of many high-level MIR applications such as audio summarisation, audio finger printing, and automatic annotation tools. Besides such low-level enhancements, our information geometric approach can address the well-known (and still to be solved) problem of audio queries over a database.
- **Automatic Accompaniment/Creative Tools for Entertainment Industry** Technologies developed by MUTANT can find their way with general public (besides professional musicians) and within the entertainment industry. Recent trends in music industry show signs of tendencies towards more intelligent and interactive interfaces for music applications. Among them is reactive and adaptive automatic accompaniment and performance assessment as commercialized by companies such as *MakeMusic* and *Tonara*. Technologies developed around *Antescofo* can enhance interaction between user and the computer for such large public applications. We hope to pursue this by licensing our technologies to third-party companies.

PAREO Project-Team

4. Application Domains

4.1. Application Domains

Beside the theoretical transfer that can be performed via the cooperations or the scientific publications, an important part of the research done in the *Pareo* group team is published within software. *Tom* is our flagship implementation. It is available via the Inria Gforge (<http://gforge.inria.fr>) and is one of the most visited and downloaded projects. The integration of high-level constructs in a widely used programming language such as Java may have an impact in the following areas:

- Teaching: when (for good or bad reasons) functional programming is not taught nor used, *Tom* is an interesting alternative to exemplify the notions of abstract data type and pattern-matching in a Java object oriented course.
- Software quality: it is now well established that functional languages such as Caml are very successful to produce high-assurance software as well as tools used for software certification. In the same vein, *Tom* is very well suited to develop, in Java, tools such as provers, model checkers, or static analyzers.
- Symbolic transformation: the use of formal anchors makes possible the transformation of low-level data structures such as C structures or arrays, using a high-level formalism, namely pattern matching, including associative matching. *Tom* is therefore a natural choice each time a symbolic transformation has to be implemented in C or Java for instance. *Tom* has been successfully used to implement the Rodin simplifier, for the B formal method.
- Prototyping: by providing abstract data types, private types, pattern matching, rules and strategies, *Tom* allows the development of quite complex prototypes in a short time. When using Java as the host-language, the full runtime library can be used. Combined with the constructs provided by *Tom*, such as strategies, this procures a tremendous advantage.

One of the most successful transfer is certainly the use of *Tom* made by Business Objects/SAP. Indeed, after benchmarking several other rule based languages, they decided to choose *Tom* to implement a part of their software. *Tom* is used in Paris, Toulouse and Vancouver. The standard representation provided by *Tom* is used as an exchange format by the teams of these sites.

PARKAS Project-Team

4. Application Domains

4.1. Application Domains

The project addresses the design, semantics and implementation of programming languages together with compilation techniques to develop provably safe and efficient computing systems. Traditional applications can be found in safety critical embedded systems with hard real-time constraints such as avionics (e.g., fly-by-wire command), railways (e.g., on board control, engine control), nuclear plants (e.g., emergency control of the plant). While embedded applications have been centralized, they are now massively parallel and physically distributed (e.g., sensor networks, train tracking, distributed simulation of factories) and they integrate computationally intensive algorithms (e.g., video processing) with a mix of hard and soft real-time constraints. Finally, systems are heterogeneous with discrete devices communicating with physical ones (e.g., interface between analog and digital circuits). Programming and simulating a whole system from a unique source code, with static guarantees on the reproducibility of simulations together with a compiler to generate target embedded code is a scientific and industrial challenge of great importance.

PARSIFAL Project-Team

4. Application Domains

4.1. Automated Theorem Proving

Automated theorem proving has traditionally focused on classical first-order logic, but non-classical logics are increasingly becoming important in the specification and analysis of software. Most type systems are based on (possibly second-order) propositional intuitionistic logic, for example, while resource-sensitive and concurrent systems are most naturally expressed in linear logic.

The members of the Parsifal team have a strong expertise in the design and implementation of performant automated reasoning systems for such non-classical logics. In particular, the Linprover suite of provers [35] continue to be the fastest automated theorem provers for propositional and first-order linear logic.

Any non-trivial specification, of course, will involve theorems that are simply too complicated to prove automatically. It is therefore important to design semi-automated systems that allow the user to give high level guidance, while at the same time not having to write every detail of the formal proofs. High level proof languages in fact serve a dual function – they are more readily comprehended by human readers, and they tend to be more robust with respect to maintenance and continued evolution of the systems. Members of the Parsifal team, in association with other Inria teams and Microsoft Research, have been building a heterogeneous semi-automatic proof system for verifying distributed algorithms [36].

On a more foundational level, the team has been developing many new insights into the structure of proofs and the proof search spaces. Two directions, in particular, present tantalizing possibilities:

- The concept of *multi-focusing* [37] can be used to expose concurrency in computational behavior, which can in turn be exploited to prune areas of the proof search space that explore irrelevant interleavings of concurrent actions.
- The use of *bounded search*, where the bounds can be shown to be complete by meta-theoretic analysis, can be used to circumvent much of the non-determinism inherent in resource-sensitive logics such as linear logic. The lack of proofs of a certain bound can then be used to justify the presence or absence of properties of the encoded computations.

Much of the theoretical work on automated reasoning has been motivated by examples and implementations, and the Parsifal team intends to continue to devote significant effort in these directions.

4.2. Mechanized Metatheory

There has been increasing interest in the use of formal methods to provide proofs of properties of programs and programming languages. Tony Hoare’s Grand Challenge titled “Verified Software: Theories, Tools, Experiments” has as a goal the construction of “verifying compilers” for a world where programs would only be produced with machine-verified guarantees of adherence to specified behavior. Guarantees could be given in a number of ways: proof certificates being one possibility.

The POPLMark challenge [33] envisions “a world in which mechanically verified software is commonplace: a world in which theorem proving technology is used routinely by both software developers and programming language researchers alike.” The proposers of this challenge go on to say that a “crucial step towards achieving these goals is mechanized reasoning about language metatheory.”

The Parsifal team has developed several tools and techniques for reasoning about the meta-theory of programming languages. One of the most important requirements for programming languages is the ability to reason about data structures with binding constructs up to α -equivalence. The use of higher-order syntax and nominal techniques for such data structures was pioneered by Miller, Nadathur and Tiu. The Abella system (see Section 3.2) implements a refinement of a number of these ideas and has been used to give full solutions to sections of the POPLMark challenge in addition to fully formal proofs of a number of other theorems in the meta-theory of the λ -calculus.

Now that the Abella system has been in circulation among colleagues during the past couple of years, there are many aspects of the methodology that now need to be addressed. During the summer of 2011, the team employed three interns Carnegie Mellon University and McGill University to work on different aspects of Abella. Particular focus was given to better ways to manipulate specification-logic contexts in the reasoning-logic and with finding ways to have Abella output a proper proof object (different from the scripts that are used to find a proof).

Our colleague Alwen Tiu from the Australian National University has also been building on our Bedwyr model checking tool so that we can build on top of it his SPEC system for doing model checking of spi-calculus expressions. We have adopted his enhancements to Bedwyr and are developing further improvements within the context of the BATT project (see Section 5.2).

4.3. Proof Certificates

Members of the Parsifal team have shown how to specify a large variety of proof systems—including natural deduction, the sequent calculus, and various tableau and free deduction systems—uniformly using either focused linear logic [52], [50] or focused intuitionistic logic [43] as the meta-language. In the presence of induction and co-induction, arbitrary finite computations can be embedded into single synthetic steps [34]. Additional work [8] shows that this same framework can also capture resolution refutations as well as Pratt primality certificates.

An important application then of this work in designing synthetic inference systems based on classical and intuitionistic logic is that of designing a *broad spectrum proof certificate*. The definition of proof certificates can be remarkably flexible within the simple setting of focused proofs.

The most important implications of such a certificate format would be that most of the worlds theorem provers should be able to print out their proofs and communicate them to other provers: these other provers could then check such certificates by expanding the synthetic connectives they contain down into a small and fixed set of “micro” inference rules.

4.4. Automated reasoning and SMT solving

Automated reasoning uses a broad range of techniques whose soundness and completeness relate to the existence of proofs. The research programme of the ANR PSI project at Parsifal is to build a finer-grained connection by specifying automated reasoning techniques as the step-by-step construction of proofs, as we know it from proof theory and logic programming. The goal is to do this in a unifying framework, namely proof-search in a polarized and focussed logic. One of the advantages of this is to combine those techniques more easily. Another one is to envisage extending those techniques.

For instance, SAT-modulo-Theory problems require the combination of logical reasoning with domain-specific decision procedures. So in the PSI project we study how to incorporate the call to decision procedures in proof-theoretical framework like the focussed sequent calculus, and the proof-search mechanisms that are related to it.

In the same spirit we also study how to handle existential variables and equality, for which specific automated reasoning techniques have been designed (superposition / paramodulation calculi).

PL.R2 Project-Team

4. Application Domains

4.1. The impact of Coq

Coq is one of the 8 most used proof assistants in the world. In Europe, its main challengers are Isabelle (developed in Munich, Germany), HOL (developed in Cambridge, UK) and Mizar (developed in Białystok, Poland).

Coq is used in various research contexts and in a few industrial contexts. It is used in the context of formal mathematics at the University of Nijmegen (constructive algebra and analysis), Inria Sophia-Antipolis (number theory and algebra), Inria-MSR joint lab (group theory), the University of Nice (algebra). It is used in France in the context of computer science at Inria-Rocquencourt (certified compilation), Inria-Saclay (certification of imperative programs), LORIA, Strasbourg (certification of geometry algorithms). Outside France, it is used in the context of computer science e.g. at U. Penn, Harvard (programming languages, semantics), Yale, Ottawa and Berkeley Universities (building of a certified platform for proof-carrying code), University of Princeton (certified compilation), AIST at Tokyo (certification of cryptographic protocols), Microsoft Research Cambridge (proof of imperative programs), ... In the industry, it is used by Gemalto and Trusted Logic (JavaCard formal model and commercial applets certification).

All in all, it is difficult to evaluate how much Coq is used. Two indicators are the readership of the textbook on Coq by Yves Bertot and Pierre Castéran [35] and the number of subscribers to the Coq-club mailing list. More than 1200 copies of the book have been sold. There has been a second printing, and a Chinese translation of the book has been published. There are around 600 subscribers to the mailing list. Coq is taught or used for teaching in many universities: Paris, Bordeaux, Lyon, Nice, Strasbourg, CNAM, Nottingham, Ottawa, U. Penn, Harvard, MIT, Princeton, Yale, Berkeley, Warsaw, Krakow, Rosario in Argentina, ...

Users of the assistant are also disseminating the use of the tool: A collaborative effort led by B. Pierce's team at U. Penn gave rise to a set of courses named Software Foundations (<http://www.seas.upenn.edu/~cis500/current/sf/index.html>) [66] on basic logic and computer science in Coq, that is used by many universities and individuals throughout the world. A. Chlipala wrote an advanced textbook on "Certified Programming with Dependent Types" in Coq, freely available on the web and soon to be published by MIT Press [37].

POLSYS Project-Team

4. Application Domains

4.1. Cryptology

We propose to develop a systematic use of *structured systems* in Algebraic Cryptanalysis. We want to improve the efficiency and to predict the theoretical complexity of such attacks. We plan to demonstrate the power of algebraic techniques in new areas of cryptography such as Algebraic Number Theory (typically, in curve based cryptography).

4.2. Engineering sciences

Solving polynomial systems over the reals arise as a critical issue in wide range of problems coming from engineering sciences (biology, physics, control theory, etc.). We will focus on developing general enough software that may impact on these domains with a particular focus on control theory.

POP ART Project-Team

4. Application Domains

4.1. Industrial Applications

Our applications are in the embedded system area, typically: transportation, energy production, robotics, telecommunications, systems on chip (SoC). In some areas, safety is critical, and motivates the investment in formal methods and techniques for design. But even in less critical contexts, like telecommunications and multimedia, these techniques can be beneficial in improving the efficiency and the quality of designs, as well as the cost of the programming and the validation processes.

Industrial acceptance of formal techniques, as well as their deployment, goes necessarily through their usability by specialists of the application domain, rather than of the formal techniques themselves. Hence our orientation towards the proposal of domain-specific (but generic) realistic models, validated through experience (*e.g.*, control tasks systems), based on formal techniques with a high degree of automation (*e.g.*, synchronous models), and tailored for concrete functionalities (*e.g.*, code generation).

4.2. Industrial Design Tools

The commercially available design tools (such as UML with real-time extensions, MATLAB/ SIMULINK/ dSPACE¹²) and execution platforms (OS such as VxWORKS, QNX, real-time versions of LINUX ...) starts now to provide besides their core functionalities design or verification methods. Some of them, founded on models of reactive systems, come close to tools with a formal basis, such as for example STATEMATE by iLOGIX.

Regarding the synchronous approach, commercial tools are available: SCADE¹³ (based on LUSTRE), CONTROLBUILD and RT-BUILDER (based on SIGNAL) from GEENSY¹⁴ (part of DASSAULTSYSTEMES), specialized environments like CELLCONTROL for industrial automatism (by the INRIA spin-off ATHYS– now part of DASSAULTSYSTEMES). One can observe that behind the variety of actors, there is a real consistency of the synchronous technology, which makes sure that the results of our work related to the synchronous approach are not restricted to some language due to compatibility issues.

4.3. Current Industrial Cooperations

Regarding applications and case studies with industrial end-users of our techniques, we cooperate with STMicroelectronics on dynamic data-flow models of computation for streaming applications, dedicated to high definition video applications for their new STHORM manycore chip.

¹²<http://www.dspaceinc.com>

¹³<http://www.esterel-technologies.com>

¹⁴<http://www.geensoft.com>

PROSECCO Project-Team

4. Application Domains

4.1. Cryptographic protocol implementations

Cryptographic protocols such as TLS, SSH, IPSec, and Kerberos are the trusted base on which the security of modern distributed systems is built. Our work enables the analysis and verification of such protocols, both in their design and implementation. Hence, for example, we build and verify models and reference implementations for well-known protocols such as TLS, as well as analyze their popular implementations such as OpenSSL.

4.2. Hardware-based security APIs

Cryptographic devices such as Hardware Security Modules (HSMs) and smartcards are used to protect long-term secrets in tamper-proof hardware, so that even attackers who gain physical access to the device cannot obtain its secrets. These devices are used in a variety of scenarios ranging from bank servers to transportation cards (e.g. Navigo). Our work investigates the security of commercial cryptographic hardware and evaluates the APIs they seek to implement.

4.3. Web application security

Web applications use a variety of cryptographic techniques to securely store and exchange sensitive data for their users. For example, a website may authenticate and authorize users using a single sign-on protocol such as OAuth, a cloud storage service may encrypt user files on the server-side using XML encryption, and a password manager may encrypt passwords in the browser using a JavaScript cryptographic library. We build verification tools that can analyze such usages in commercial web applications and evaluate their security against sophisticated web-based attacks.

S4 Project-Team

4. Application Domains

4.1. Modular design of embedded systems with interface theories

In 2006, with the opportunity of the SPEEDS European project on embedded system design, we decided to open a new research track on contract-/interface-based design. Our objective was to provide theory, methods and tools to support the design of embedded software in transport system industries. According to our understanding of industrial needs, gained during the SPEEDS European project, the following requirements apply to the notions of *contract* and *interface* and have been used to guide our research on this topic:

- Complex embedded and reactive systems are generally developed under a multi-layered OEM-supplier chain. Hence, a contract-based methodology should offer provision for formalizing the technical part of contractual relations. This should be achieved by formalizing, for a considered subsystem: 1/ its context of use (*assumptions*), and 2/ what is expected from the subsystem (*guarantees*). Assumptions and guarantees can be specified separately, or in a single automata-theoretic structure called interface.
- When developed under a contract-/interface-based methodology, subsystems or components should be designable in isolation, by including the needed information regarding possible future contexts of use. Subsystems or components should be substitutable to their specifications, meaning that their integration should raise no problem.
- Large systems are concurrently developed for their different *aspects* or *viewpoints* by different teams using different frameworks and tools. Examples of such aspects include the functional, reliability, timing, memory and power aspects. Each of these aspects requires specific frameworks and tools for their analysis and design. Yet, they are not totally independent but rather interact. The issue of dealing with multiple aspects or multiple viewpoints is thus essential. This implies that several contracts or interfaces are associated with a same system, sub-system, or component, namely at least one per viewpoint. These contracts/interfaces are to be interpreted in a conjunctive way and modular reasoning methods have to be developed to support large sets of contracts.
- The need for supporting conjunctive contracts/interfaces also follows from the current practice in which early requirement capture results in many elementary requirements. These requirements typically consist of English text, semi-formal languages whose sentences are translatable into predefined behavioral patterns, or even graphical scenario languages.
- It is highly desirable that designing by contracts and interfaces has the mildest possible impact on the design process, a key proprietary asset to all major companies.

4.2. Opacity, Supervision, and Petri Nets

Our activities on components emerged from a larger basis of competences developed in the past of S4 on supervisory control and Petri net synthesis. Components and their interfaces are intimately tied to supervisory control, and Petri net synthesis is a possible approach to controller synthesis. In the last four years, we have carried on work on both themes, but refocussed our research on fresh topics. A major contribution has been to study supervisory control for secrecy objectives, with promising results. Another contribution has been to study supervisory control for finite abstractions of services. The fusion of both topics, that would increase the interest of the results for Web applications, is not yet done. A different topic that we continued to investigate is the synthesis of distributed controllers based on the synthesis of distributed Petri nets. Our progress on this difficult topic is limited, but we feel we should pursue the effort.

Opacity is an abstract property that includes non-interference and that can cover confidentiality, authenticity and many other specific security concepts. Our project-team has inaugurated research on supervisory control of discrete event systems for opacity, which became soon a theme of cooperation with project-team Vertecs and subsequently attracted concurrent researches at Wayne State U., Kyoto Inst. of Tech., and U. Illinois. We have some advance over these concurrent teams.

The rest of our work on supervision focusses on minimizing communication between decentralized controllers, on asynchronous and distributed control, and on the enforcement of modal specifications. Decreasing communication between decentralized controllers was studied at Michigan U. but we could further show that minimizing communication reduces to a classical optimization problem. As regards asynchronously communicating control, the only current attempts we are aware of are those of project-teams S4 and Vertecs. As regards supervisory control w.r.t. modal specifications, the closest work is Lohmann and Wolf's synthesis of communication partners for Web services.

The approach which we propose towards distributed control relies upon the synthesis of distributed Petri nets. We have been leaders for fifteen years on the synthesis of P/T nets, on a par with the Petrify team focussed on Elementary (or safe) net synthesis. The algorithms which we have defined have been reused or adapted by many other researchers in Europe, in the US, and in China, to respond to three types of problems: controller synthesis, process mining, and concise representation of services. We are currently writing a book covering all aspects of the theory and applications of Petri net synthesis. We also pursue research on structure theory of Petri nets, in cooperation with U. Oldenburg, with focus set recently on non-interference.

4.3. Hybrid Systems Modelers

This is an opportunistic objective, not part of the plans stated when the team was formed. It results from a series of events: in 2008, Benoît Caillaud was part of the *Synchronics* large scale initiative (see section 7.1.1), dedicated to “embedded systems programming in 2020”. Hybrid Systems Modelers were part of the research program. Such tools are nowadays absolutely central in the development of Cyber Physical Systems (CPS), which are physical systems in closed loop with embedded control. Hybrid Systems Modelers support the modeling of physical systems (with Ordinary Differential Equations, ODE, and Differential Algebraic Equations, DAE): Matlab-Simulink and Modelica are the main players. Our vision was that these tools should deserve similar effort in theory as synchronous languages did for the programming of embedded systems. About one year after *Synchronics* started (focusing mostly on other topics), the PhD thesis of Simon Bliudze came to our knowledge. This thesis contained a long chapter on the use of *non-standard analysis* as a semantic framework for hybrid systems. The exposure relied on a recent presentation of non-standard analysis, not axiomatic in style, due to the mathematician Lindström. That attracted the attention of Albert Benveniste, so he joined the group of *Synchronics* working on hybrid systems. This was the beginning of a deeply novel and exciting research track.

The computer science community has devoted significant efforts to the analysis and verification of hybrid automata. The framework of hybrid automata is, however, much less flexible than what actual Hybrid Systems Modelers offer. The only ongoing effort towards modeling has been developed by Edward Lee and his team as part of the Ptolemy II project. This has led to the proposal of *super-dense time semantics*, in which cascades of successive instants can occur in zero time by using $R_+ \times N$ as a time index. It turns out that the set $T = \{n\partial \mid n \in N^*\}$, where ∂ is an *infinitesimal* and N^* is the set of *non-standard integers* is such that $1/T$ is dense in R_+ , making it “continuous”, and $2/$ every $t \in T$ has a predecessor in T and a successor in T , making it “discrete” (le beurre et l'argent du beurre, as we say in french). Although non-effective from the operational point of view, the *non-standard semantics* of hybrid systems provides a framework that is very familiar to the computer scientist (who is afraid of continuous time) and at the same time efficient as a symbolic abstraction. This makes it an excellent candidate for the development of compilation schemes.

SECRET Project-Team

4. Application Domains

4.1. Application domains

Our main application domains are:

- cryptology,
- error-correcting codes, especially codes for quantum communications and fault-tolerant quantum computing,
- reverse-engineering of communication systems.

We also investigate some cross-disciplinary domains, which require a scientific competence coming from other areas, mainly social aspects of cryptology, cryptology for large databases.

SECSI Project-Team

4. Application Domains

4.1. Application Domains

Here are a few examples of applications of research done in SECSI:

- Security of electronic voting schemes: the case of the Helios protocol, used in particular at University of Louvain-la-Neuve (2010) and at the International Association for Cryptographic Research (IACR).
- Security of the protocols involved in the TPM (Trusted Platform Module) chip, a chip present in most PC laptops today, and which is meant to act as a trusted base.
- Security of the European electronic passport—and the discovery of an attack on the French implementation of it.
- The Tookan tool allows one to assess the security of security tokens. These tokens are meant as safes holding secret keys, which should never be permitted to get out unencrypted. Several vulnerabilities discovered. Several interesting customers in banking (Barclays), in aeronautics (Boeing), notably.
- Intrusion detection with the Orchids tool: several interested partners, among which EADS Cassidian, Thales, Galois Inc. (USA), the French Direction Générale de l'Armement (DGA).

TASC Project-Team

4. Application Domains

4.1. Introduction

Constraint programming deals with the resolution of decision problems by means of rational, logical and computational techniques. Above all, constraint programming is founded on a clear distinction between, on the one hand the description of the constraints intervening in a problem, and on the other hand the techniques used for the resolution. The ability of constraint programming to handle in a flexible way heterogeneous constraints has raised the commercial interest for this paradigm in the early eighties. Among his fields of predilection, one finds traditional applications such as computer aided decision-making, scheduling, planning, placement, logistics or finance, as well as applications such as electronic circuits design (simulation, checking and test), DNA sequencing and phylogeny in biology, configuration of manufacturing products or web sites, formal verification of code.

4.2. Panorama

In 2012 the **TASC** team was involved in the following application domains:

- *Planning and replanning* in Data Centres (**SelfXL** project).
- *Packing complex shapes* in the context of a warehouse (NetWMS2 project).
- Building decision support system for *city development planning with evaluation of energy impacts* (**SUSTAINS** project).
- *Optimizing electricity production* in the context of the **Gaspard Monge call program for Optimisation and Operation Research**. We extract global constraints from daily energy production temporal series issued from all productions plants of **EDF** over a period of several years.

TOCCATA Team

4. Application Domains

4.1. Application Domains

Keywords: embedded software, smartcards, avionics, telecommunication, transportation systems

The application domains we target involve safety-critical software, that is where a high level guarantee of soundness of functional execution of the software is wanted. The domains of application include

- Transportation: aeronautics, railroad, space flight, automotive
- Communications: mobile phones, smart phones, Web applications
- Financial applications, banking
- Medicine: diagnostic devices, computer-assisted surgery
- Databases with confidentiality requirements (e.g. health records, electronic voting)

Currently our industrial collaborations mainly belong the first of these domains: transportation. These include, in the context of the ANR U3CAT project (Airbus France, Toulouse; Dassault Aviation, Saint-Cloud; Sagem Défense et Sécurité):

- proof of C programs via *Frama-C/Jessie/Why* ;
- proof of floating-point programs ;
- use of the *Alt-Ergo* prover via CAVEAT tool (CEA) or *Frama-C/WP*.

In the context of the FUI project Hi-Lite, the Adacore (Paris) uses *Why3* and *Alt-Ergo* as back-end to GnatProve, an environment for verification of Ada programs. This is applied in the domain of aerospace (Thales).

In the context of a new ANR project BWare, we investigate the use of *Why3* and *Alt-Ergo* as an alternative back-end for checking proof obligation generated by *Atelier B*, whose main applications are railroad-related software (http://www.methode-b.com/documentation_b/ClearSy-Industrial_Use_of_B.pdf, collaboration with Mitsubishi Electric R&D Centre Europe, Rennes; ClearSy, Aix-en-Provence)

Apart from the domain of transportation, the Cubicle model checker modulo theories based on the *Alt-Ergo* SMT prover (collaboration with Intel Strategic Cad Labs, Hillsboro, OR, USA) can be applied to verification of concurrent programs and protocols (<http://cubicle.lri.fr/>).

TRIO Project-Team

4. Application Domains

4.1. Domain

Three main application domains can be underlined.

- In-vehicle embedded systems. A lot of work developed in TRIO is oriented towards transportation systems (cars, autonomous vehicles, etc.). They mainly cover two points. The first one is the specification of what must be modeled in such a system and how to reach a good accuracy of a model; this leads to investigate topics like Architecture Description Languages and automatic generation of models. The second point concerns the verification of dependability properties and temporal properties required by these applications and, consequently, the development of new fault tolerant on-line mechanisms to include in an application or the automatic generation of a standard middleware.
- Compilation, memory management and low-power issues for real time embedded systems. It becomes mandatory to design embedded systems that respect performances and reliability constraints while minimizing the energy consumption. Hence, TRIO is involved, on the one hand, in the definition of ad-hoc memory management at compilation time and on the other hand, in joint study of memory management strategies and tasks scheduling for real time critical systems.
- Code analyses and software visualization for embedded systems. Despite important advances, it is still impossible to develop and optimize automatically all the programs with all their variety, especially when deployment constraints are considered. Software design and implementation thus remain highly ad-hoc, poorly automated activities, with a human being in the loop. TRIO is thus involved in the design of better tools for software engineering focusing on helping the human developer understand and develop the system, thanks to powerful automated program analyses and advanced visualizations techniques.

TYPICAL Project-Team (section vide)

VEGAS Project-Team

3. Application Domains

3.1. Computer graphics

We are interested in the application of our work to virtual prototyping, which refers to the many steps required for the creation of a realistic virtual representation from a CAD/CAM model.

When designing an automobile, detailed physical mockups of the interior are built to study the design and evaluate human factors and ergonomic issues. These hand-made prototypes are costly, time consuming, and difficult to modify. To shorten the design cycle and improve interactivity and reliability, realistic rendering and immersive virtual reality provide an effective alternative. A virtual prototype can replace a physical mockup for the analysis of such design aspects as visibility of instruments and mirrors, reachability and accessibility, and aesthetics and appeal.

Virtual prototyping encompasses most of our work on effective geometric computing. In particular, our work on 3D visibility should have fruitful applications in this domain. As already explained, meshing objects of the scene along the main discontinuities of the visibility function can have a dramatic impact on the realism of the simulations.

3.2. Solid modeling

Solid modeling, i.e., the computer representation and manipulation of 3D shapes, has historically developed somewhat in parallel to computational geometry. Both communities are concerned with geometric algorithms and deal with many of the same issues. But while the computational geometry community has been mathematically inclined and essentially concerned with linear objects, solid modeling has traditionally had closer ties to industry and has been more concerned with curved surfaces.

Clearly, there is considerable potential for interaction between the two fields. Standing somewhere in the middle, our project has a lot to offer. Among the geometric questions related to solid modeling that are of interest to us, let us mention: the description of geometric shapes, the representation of solids, the conversion between different representations, data structures for graphical rendering of models and robustness of geometric computations.

3.3. Fast prototyping

We work in collaboration with **CIRTES** on rapid prototyping. **CIRTES**, a company based in Saint-Dié-des-Vosges, has designed a technique called Stratoconception[®] where a prototype of a 3D computer model is constructed by first decomposing the model into layers and then manufacturing separately each layer, typically out of wood of standard thickness (e.g. 1 cm), with a three-axis CNC (Computer Numerical Controls) milling machine. The layers are then assembled together to form the object. The Stratoconception[®] technique is cheap and allows fast prototyping of large models.

When the model is complex, for example an art sculpture, some parts of the models may be inaccessible to the milling machine. These inaccessible regions are sanded out by hand in a post-processing phase. This phase is very consuming in time and resources. We work on minimizing the amount of work to be done in this last phase by improving the algorithmic techniques for decomposing the model into layers, that is, finding a direction of slicing and a position of the first layer.

VERIDIS Project-Team

4. Application Domains

4.1. Application Domains

Our work focuses on distributed algorithms and protocols. These are or will be found at all levels of computing infrastructure, from many-core processors and systems-on-chip to wide-area networks. We are particularly interested in novel paradigms, for example ad-hoc networks that underly mobile and low-power computing or overlay networks and peer-to-peer networking that provide services for telecommunication or cloud computing services. Distributed protocols underly computing infrastructure that must be highly available and mostly invisible to the end user, therefore correctness is important. One should note that standard problems of distributed computing such as consensus, group membership or leader election have to be reformulated for the dynamic context of these modern systems. We are not ourselves experts in the design of distributed algorithms, but work together with domain experts on the modeling and verification of these protocols. These cooperations help us focus on concrete algorithms and ensure that our work is relevant to the distributed algorithm community.

Formal verification techniques that we study can contribute to certify the correctness of systems. In particular, they help assert under which assumptions an algorithm or system functions as required. For example, the highest levels of the Common Criteria for Information Technology Security Evaluation require code analysis, based on mathematically precise foundations. While initially the requirements of certified development have mostly been restricted to safety-critical systems, they are becoming more and more common due to the cost associated with malfunctioning system components and software. We are in particular working on modeling and verifying medical devices that require closed-loop models of both the system and its environment.

VERTECS Project-Team

4. Application Domains

4.1. Overview

The methods and tools developed by the VERTECS project-team for test generation and control synthesis of reactive systems are intended to be as generic as possible. This allows us to apply them in many application domains where the presence of software is predominant and its correctness is essential. In particular, we apply our research in the context of telecommunication systems, for embedded systems, for smart-cards application, and control-command systems.

4.2. Telecommunication systems

Our research on test generation was initially proposed for conformance testing of telecommunication protocols. In this domain, testing is a normalized process [21], and formal specification languages are widely used (SDL in particular). Our test generation techniques have already proved useful in this context, going up to industrial transfer. New standardized component-based design methodologies such as UML and OMG's MDE increase the need for formal techniques in order to ensure the compositionality of components, by verification and testing. Our techniques, by their genericity and adaptativity, have also proved useful at different levels of these methodologies, from component testing to system testing. The telecommunication industry now also tries to provide more and more services to the users. These services must be validated.

4.3. Software embedded systems

In the context of transport, software embedded systems are increasingly predominant. This is particularly important in automotive systems, where software replaces electronics for power train, chassis (e.g. engine control, steering, brakes) and cabin (e.g. wiper, windows, air conditioning) or new services to passengers are increasing (e.g. telematics, entertainment). Car manufacturers have to integrate software components provided by many different suppliers, according to specifications. One of the problems is that testing is done late in the life cycle, when the complete system is available. Faced with these problems, but also with the complexity of systems, compositionality of components, distribution, etc, car manufacturers now try to promote standardized interfaces and component-based design methodologies. They also develop virtual platforms which allow for testing components before the system is complete. It is clear that software quality and trust are one of the problems that have to be tackled in this context. This is why we believe that our techniques (testing and control) can be useful.

4.4. Control-command systems

The main application domain for our techniques is control-command systems. In general, such systems control costly machines (see, e.g., robotic systems, flexible manufacturing systems), that are connected to an environment (e.g., a human operator). Such systems are often critical systems and errors occurring during their execution may have dramatic economical or human consequences. In this field, the controller synthesis methodology (CSM) is useful to ensure by construction the interaction between 1) the different components, and 2) the environment and the system itself. For the first point, the CSM is often used as a safe scheduler, whereas for the second one, the supervisor can be interpreted as a safe discrete tele-operation system. Also in the context of the Vaccim ANR project, we investigate the testing, monitoring and verification of control-command systems.

ALEA Project-Team

4. Application Domains

4.1. Application Domains

This short section is only concerned with the list of concrete application domains developed by our team project on Bayesian inference and unsupervised learning, nonlinear filtering and rare event analysis. Most of these application areas result from fruitful collaborations with other national institutes or industries.

Three application domains are directly related to evolutionary computing, particle filtering and Bayesian inference. They are currently investigated by our team project:

1. **Multi-object tracking.** Multi-object tracking deals with the task of estimating the states of a set of moving objects from a set of measurements obtained sequentially. These measurements may either arise from one of the targets or from clutter and the measurement-to-target association is generally unknown. This problem can then be recast as a dynamic clustering one where the clusters are the clutter and the different targets. The targets actually move in time, some targets may appear/disappear over time and the number of targets is generally unknown and time-varying. The ALEA team has been involved in the ANR project Propagation (2010-2012) with DCNS, Thalés and Exavision.
2. **Finance.** The Team ALEA is interested in the design and analysis of new advanced particle methods for option pricing, partial observation problems, and sensitivity measures computation. An **international workshop** has been jointly organized by ALEA and **CMAP (Polytechnique)** on this topic in October 2012 (organizers: E. Gobet, P. Del Moral, P. Hu).
3. **Epidemiology.** Our team is interested in the development and analysis of particle mean field models for the calibration and uncertainty propagations in complex kinetic population models. The ALEA team is involved in an interdisciplinary exploratory research project with Laboratory Ecologie & Evolution, and co-organized an **international workshop** on this topic in 2011.

APICS Project-Team

4. Application Domains

4.1. Introduction

These domains are related to the problems described in sections 3.2.1 and 3.2.2 . They are handled using the techniques described in section 3.3 .

4.2. Inverse problems for elliptic PDE

Participants: Laurent Baratchart, Juliette Leblond, Ana-Maria Nicu, Dmitry Ponomarev.

This work is done in collaboration with Maureen Clerc and Théo Papadopoulo from the Athena project-team. Solving overdetermined Cauchy problems for the Laplace equation on a spherical layer (in 3-D) in order to extrapolate incomplete data (see section 3.2.1) is a necessary ingredient of the team’s approach to inverse source problems, in particular for applications to EEG since the latter involves propagating the initial conditions through several layers of different conductivities, from the boundary down to the center of the domain where the singularities (*i.e.* the sources) lie. Actually, once propagated to the innermost sphere, it turns out that that traces of the boundary data on 2-D cross sections (disks) coincide with analytic functions in the slicing plane, that has branched singularities inside the disk [4]. These singularities are related to the actual location of the sources (namely, they reach in turn a maximum in modulus when the plane contains one of the sources). Hence, we are back to the 2-D framework of section 3.3.3 where approximately recovering these singularities can be performed using best rational approximation.

Numerical experiments gave very good results on simulated data and we are now proceeding with real experimental magneto-encephalographic data, see also sections 5.6 and 6.1 . The PhD thesis of A.-M. Nicu [13] was concerned with these applications, see [16], in collaboration with the Athena team at Inria Sophia Antipolis, and neuroscience teams in partner-hospitals (hosp. Timone, Marseille).

Similar inverse potential problems appear naturally in magnetic reconstruction. A particular application, which is the object of a joint NSF project with Vanderbilt University and MIT, is to geophysics. There, the remanent magnetization of a rock is to be analysed to draw information on magnetic reversals and to reconstruct the rock history. Recently developed scanning magnetic microscopes measure the magnetic field down to very small scales in a “thin plate” geological sample at the Laboratory of planetary sciences at MIT, and the magnetization has to be recovered from the field measured on a plane located at small distance above the slab.

Mathematically speaking, EEG and magnetization inverse problems both amount to recover the (3-D valued) quantity m (primary current density in case of the brain or magnetization in case of a thin slab of rock) from measurements of the vector potential:

$$\int_{\Omega} \frac{\operatorname{div} m(x') dx'}{|x-x'|},$$

outside the volume Ω of the object, from Maxwell’s equations.

The team is also getting engaged in problems with variable conductivity governed by a 2-D conjugate-Beltrami equation, see [5], [58], [35]. The application we have in mind is to plasma confinement for thermonuclear fusion in a Tokamak, more precisely with the extrapolation of magnetic data on the boundary of the chamber from the outer boundary of the plasma, which is a level curve for the poloidal flux solving the original div-grad equation. Solving this inverse problem of Bernoulli type is of importance to determine the appropriate boundary conditions to be applied to the chamber in order to shape the plasma [54]. These issues are the topics of the PhD theses of S. Chaabi and D. Ponomarev [27], and of a joint collaboration with the Laboratoire J.-A. Dieudonné at the Univ. of Nice-SA (and the Inria team Castor), and the CMI-LATP at the Univ. of Aix-Marseille I (see section 6.2).

4.3. Identification and design of microwave devices

Participants: Laurent Baratchart, Sylvain Chevillard, Martine Olivi, Fabien Seyfert.

This work is done in collaboration with Stéphane Bila (XLim, Limoges) and Jean-Paul Marmorat (Centre de mathématiques appliquées (CMA), École des Mines de Paris).

One of the best training grounds for the research of the team in function theory is the identification and design of physical systems for which the linearity assumption works well in the considered range of frequency, and whose specifications are made in the frequency domain. This is the case of electromagnetic resonant systems which are of common use in telecommunications.

In space telecommunications (satellite transmissions), constraints specific to on-board technology lead to the use of filters with resonant cavities in the microwave range. These filters serve multiplexing purposes (before or after amplification), and consist of a sequence of cylindrical hollow bodies, magnetically coupled by irises (orthogonal double slits). The electromagnetic wave that traverses the cavities satisfies the Maxwell equations, forcing the tangent electrical field along the body of the cavity to be zero. A deeper study (of the Helmholtz equation) states that essentially only a discrete set of wave vectors is selected. In the considered range of frequency, the electrical field in each cavity can be seen as being decomposed along two orthogonal modes, perpendicular to the axis of the cavity (other modes are far off in the frequency domain, and their influence can be neglected).



Figure 1. Picture of a 6-cavities dual mode filter. Each cavity (except the last one) has 3 screws to couple the modes within the cavity, so that 16 quantities must be optimized. Quantities such as the diameter and length of the cavities, or the width of the 11 slits are fixed during the design phase.

Each cavity (see Figure 1) has three screws, horizontal, vertical and midway (horizontal and vertical are two arbitrary directions, the third direction makes an angle of 45 or 135 degrees, the easy case is when all cavities show the same orientation, and when the directions of the irises are the same, as well as the input and output slits). Since the screws are conductors, they act more or less as capacitors; besides, the electrical field on the surface has to be zero, which modifies the boundary conditions of one of the two modes (for the other mode, the electrical field is zero hence it is not influenced by the screw), the third screw acts as a coupling between the two modes. The effect of the iris is to the contrary of a screw: no condition is imposed where there is a hole, which results in a coupling between two horizontal (or two vertical) modes of adjacent cavities (in fact the iris

is the union of two rectangles, the important parameter being their width). The design of a filter consists in finding the size of each cavity, and the width of each iris. Subsequently, the filter can be constructed and tuned by adjusting the screws. Finally, the screws are glued. In what follows, we shall consider a typical example, a filter designed by the CNES in Toulouse, with four cavities near 11 Ghz.

Near the resonance frequency, a good approximation of the Maxwell equations is given by the solution of a second order differential equation. One obtains thus an electrical model for our filter as a sequence of electrically-coupled resonant circuits, and each circuit will be modelled by two resonators, one per mode, whose resonance frequency represents the frequency of a mode, and whose resistance represent the electric losses (current on the surface).

In this way, the filter can be seen as a quadripole, with two ports, when plugged on a resistor at one end and fed with some potential at the other end. We are then interested in the power which is transmitted and reflected. This leads to defining a scattering matrix S , that can be considered as the transfer function of a stable causal linear dynamical system, with two inputs and two outputs. Its diagonal terms $S_{1,1}$, $S_{2,2}$ correspond to reflections at each port, while $S_{1,2}$, $S_{2,1}$ correspond to transmission. These functions can be measured at certain frequencies (on the imaginary axis). The filter is rational of order 4 times the number of cavities (that is 16 in the example), and the key step consists in expressing the components of the equivalent electrical circuit as a function of the S_{ij} (since there are no formulas expressing the lengths of the screws in terms of parameters of this electrical model). This representation is also useful to analyze the numerical simulations of the Maxwell equations, and to check the design, particularly the absence of higher resonant modes.

In fact, resonance is not studied via the electrical model, but via a low-pass equivalent circuit obtained upon linearising near the central frequency, which is no longer conjugate symmetric (*i.e.* the underlying system may not have real coefficients) but whose degree is divided by 2 (8 in the example).

In short, the identification strategy is as follows:

- measuring the scattering matrix of the filter near the optimal frequency over twice the pass band (which is 80Mhz in the example).
- Solving bounded extremal problems for the transmission and the reflection (the modulus of the response being respectively close to 0 and 1 outside the interval measurement, cf. section 3.3.1). This provides us with a scattering matrix of order roughly 1/4 of the number of data points.
- Approximating this scattering matrix by a rational transfer-function of fixed degree (8 in this example) via the Endymion or RARL2 software (cf. section 3.3.2.2).
- A realization of the transfer function is thus obtained, and some additional symmetry constraints are imposed.
- Finally one builds a realization of the approximant and looks for a change of variables that eliminates non-physical couplings. This is obtained by using algebraic-solvers and continuation algorithms on the group of orthogonal complex matrices (symmetry forces this type of transformation).

The final approximation is of high quality. This can be interpreted as a validation of the linearity hypothesis for the system: the relative L^2 error is less than 10^{-3} . This is illustrated by a reflection diagram (Figure 2). Non-physical couplings are less than 10^{-2} .

The above considerations are valid for a large class of filters. These developments have also been used for the design of non-symmetric filters, useful for the synthesis of repeating devices.

The team also investigates problems relative to the design of optimal responses for microwave devices. The resolution of a quasi-convex Zolotarev problems was for example proposed, in order to derive guaranteed optimal multi-band filter's responses subject to modulus constraints [11]. This generalizes the classical single band design techniques based on Tchebychev polynomials and elliptic functions. These techniques rely on the fact that the modulus of the scattering parameters of a filters, say $|S_{1,2}|$, admits a simple expression in terms of the filtering function $D = |S_{1,1}|/|S_{1,2}|$ namely,

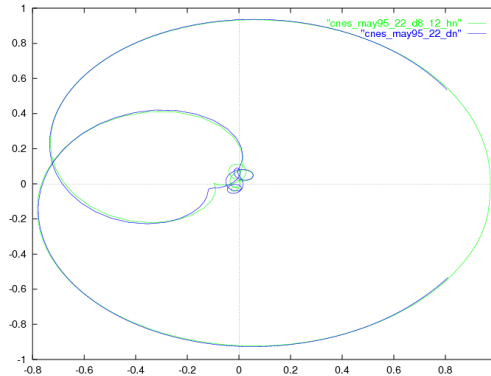


Figure 2. Nyquist Diagram. Rational approximation (degree 8) and data - S_{22} .

$$|S_{1,2}|^2 = \frac{1}{1 + D^2}.$$

The filtering function appears to be the ratio of two polynomials p_1/p_2 , the numerator of the reflection and transmission scattering factors, that can be chosen freely. The denominator q is obtained as the unique stable and unitary polynomial solving the classical Feldtkeller spectral equation:

$$qq^* = p_1p_1^* + p_2p_2^*.$$

The relative simplicity of the derivation of filter's responses under modulus constraints is due to this ability to "forget" about latter spectral equation, and express all design constraints on the filtering functions D . This no longer the case when considering the synthesis N -port devices for $N > 3$, like multiplexers, routers power dividers or when considering the synthesis of filters under matching conditions. The efficient derivation of multiplexers responses is one of the team's active recent research area, where technique based on constrained Nevanlinna-Pick interpolation problems are under study (see section 6.4.1).

ASPI Project-Team

4. Application Domains

4.1. Localisation, navigation and tracking

See 5.9 .

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [53], [47], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

- a prior dynamical model of typical evolutions of the mobile, such as inertial estimates and prior model for inertial errors,
- measurements provided by sensors,
- and possibly a digital map providing some useful feature (terrain altitude, power attenuation, etc.) at each possible position.

In some applications, another useful source of information is provided by

- a map of constrained admissible displacements, for instance in the form of an indoor building map,

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [52], [67], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

target tracking, integrated navigation, points and / or objects tracking in video sequences, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.

ASPI is contributing (or has contributed recently) to several applications of particle filtering in positioning, navigation and tracking, such as geolocalisation and tracking in a wireless network, terrain-aided navigation, and data fusion for indoor localisation.

4.2. Rare event simulation

See 3.3 , and 5.1 , 5.6 , 5.10 and 5.11 .

Another application domain of particle methods, or interacting Monte Carlo methods, that ASPI has decided to focus on is the estimation of the small probability of a rare but critical event, in complex dynamical systems. This is a crucial issue in industrial areas such as

nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly un-efficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [60], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss.

ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

BACCHUS Team

4. Application Domains

4.1. Introduction

We are working on problems that can be written in the following form

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}_e(\mathbf{U}) - \nabla \cdot \mathbf{F}_v(\mathbf{U}, \nabla \mathbf{U}) = 0 \quad (1)$$

in a domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, subjected to initial and boundary conditions. The variable \mathbf{U} is a vector in general, the flux \mathbf{F}_e is a tensor, as well as \mathbf{F}_v which also depends on the gradient of \mathbf{U} . The subsystem

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}_e(\mathbf{U}) = 0$$

is assumed to be hyperbolic, the subsystem

$$\frac{\partial \mathbf{U}}{\partial t} - \nabla \cdot \mathbf{F}_v(\mathbf{U}, \nabla \mathbf{U}) = 0$$

is assumed to be elliptic. Last, (1) is supposed to satisfy an entropy inequality. The coefficients or models that define the flux and the boundary conditions can be deterministic or random.

The systems (1) are discretised mesh made of conformal elements. The tessellation is denoted by \mathcal{T}_h . The simplices are denoted by K_j , $j = 1, n_e$, and $\cup_j K_j = \Omega_h$, an approximation of Ω . The mesh is assumed to be adapted to the boundary conditions. In our methods, we assume a globally continuous approximation of \mathbf{U} such that $\mathbf{U}|_{K_j}$ is either a polynomial of degree k or a more complex approximation such as a Nurbs. For now k is uniform over the mesh, and let us denote by V_h the vector space spanned by these functions, taking into account the boundary conditions.

The schemes we are working on have a variational formulation: find $\mathbf{U} \in V_h$ such that for any $\mathbf{V} \in V_h$,

$$a(\mathbf{U}, \mathbf{V}; \mathbf{U}) = 0.$$

The variational operator $a(\mathbf{U}, \mathbf{V}; \mathbf{W})$ is a sum of local operator that use only data within elements and boundary elements: it is very local. Boundary conditions can be implemented in a variational formulation of via penalisation, see figure 3. The third argument \mathbf{W} stands for the way are implemented the non oscillatory properties of the method.

This leads to highly non linear systems to solve, we use typically non linear Krylov space techniques. The cost is reduced thanks to a parallel implementation, the domain is partitionned via Scotch. Mesh balancing, after mesh refinement, is handled via PaMPA. These schemes are implemented in Realfluids and, partially, Aerosol. An example of such a simulation is given by Figure 2.

In case of non deterministic problems, we have a semi-intrusive strategy. The randomness is expressed via N scalar random parameters (that might be correlated), $X = (x_1, \dots, x_N)$ with probability measure $d\mu$ which support is in a subset of \mathbb{R}^N . The idea of non intrusive methods is to approximate $d\mu$ either by $d\mu \approx \sum_j \omega_j \delta_{X_j}$ for $\omega_j \geq 0$ that sum up to unity, for ‘‘well chosen’’ samples X_j or by $d\mu \approx \sum_l \mu(\Omega_l) 1_{X_j} dX$ where the sets Ω_j covers the support of $d\mu$ and are non overlapping.

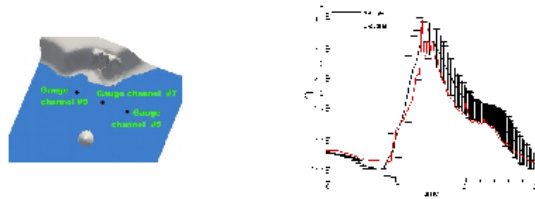


Figure 1. Topography for the Okushiri tsunami with random amplitude and phase. The mean and variance of the average at gauge 9.

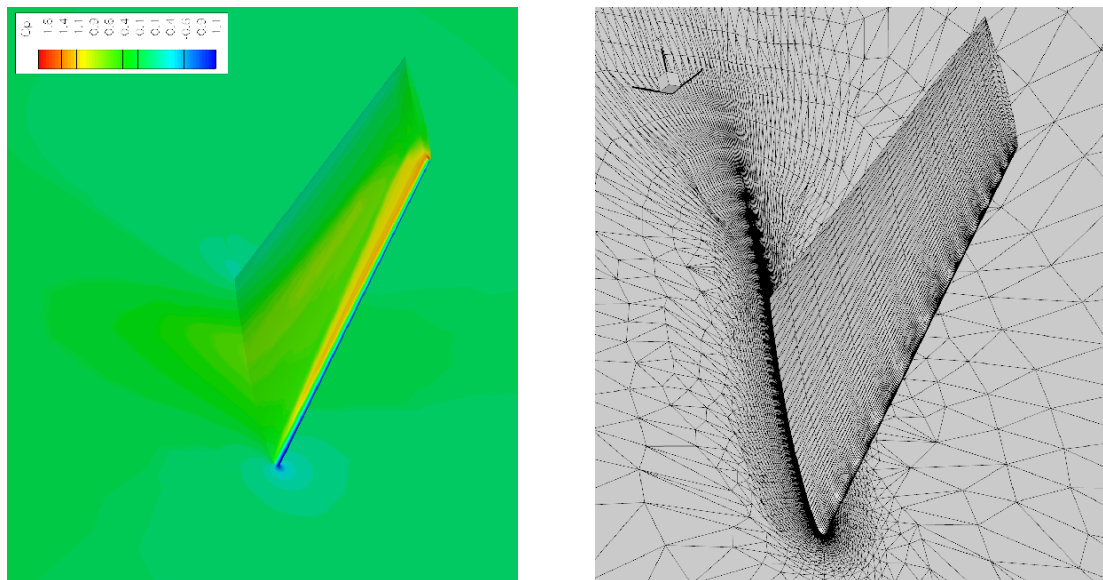


Figure 2. Flow over a turbulent M6 wing (pressure coefficient, mesh by Dassault Aviation).

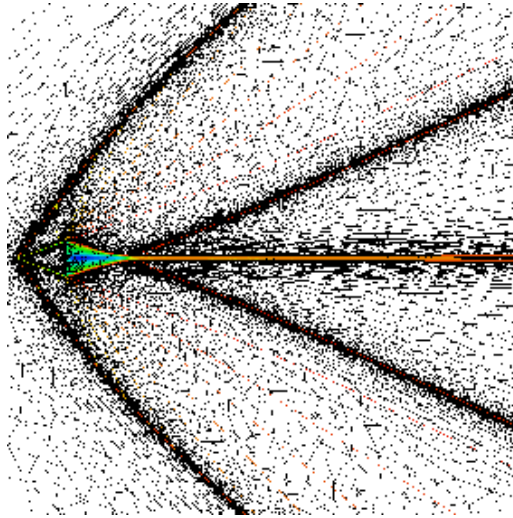


Figure 3. Adapted mesh for a viscous flow over a triangular wedge.

Starting from a discrete approximation of (1), we can implement randomness in the scheme. An example is given on figure 1 applied to the shallow water equations with dry shores, when the amplitude of the incoming tsunami wave is not known.

4.2. External and internal Aerodynamics

A classical application is the simulation of internal and external flows, with perfect or real gas equation of states, in complex geometries. This requires often the use of meshes having heterogeneous structures. We are working with unstructured meshes, either with simplicial elements or mixtures of hex, tets, pyramids and prisms. Mesh refinement can be enable in order to better resolve either the discontinuous flow structures or the capture of boundary layers.

4.3. Multiphase flows

Another domain of application is the simulation of multiphase flows. Here, the system (1) need to be supplemented by at least a PDE describing the phase volume changes, and the equation of states of the phases. The system is in most case writetn in non conservative form, so that additional difficulies need to be handled.

Multiphase fows occur in many applications: for petroleum industry, nuclear industry (accident management), engines, pipes, etc.

4.4. Inflight icing and ice shedding

Actual concerns about greenhouse gases lead to changes in the design of aircraft with an increase use of composite materials. This in turns offers new possibilities for design of ice protection systems, thus renewing interest in de-icing simulation tools. To save fuel burn, aircraft manufacturers are investigating ice protection systems such as electro-thermal or electro-mechanical de-icing systems to replace anti-icing systems. By reducing the adhesive shear strength between ice and surface, de-icing systems remove ice formed on the protected surfaces following a periodic cycle. This cycle is defined such that inter cycle ice shapes remain acceptable from a performance point of view. One of the drawbacks of de-icing device is the ice pieces shed into the flow. The knowledge of ice shedding trajectories could allow assessing the risk of impact/ingestion

on/in aircraft components located downstream. When the pieces leave the aircraft surface, they become projectiles that can hit and cause severe damage to aircraft surface or other components, such as aircraft horizontal and vertical tails, or aircraft engine. Aircraft certification authorities, such as FAA, have specific requirements for large ice fragment ingestion during engine certification. Control surfaces or wing flaps are also sensitive to ice shedding because they can be blocked by ice fragments. Aircraft manufacturers rely mainly on flight tests to evaluate the potential negative effects of ice shedding because of the lack of appropriate numerical tools. The random shape and size taken by ice shed particles together with their rotation as they move make it difficult for classical CFD tools to predict trajectories. The numerical simulation of a full unsteady viscous flow, with a set of moving bodies immersed within, shows several difficulties for grid based methods. Drawbacks income from the meshing procedure for complex geometries and the re-gridding procedure in tracing the body motion. A new approach that take into account the effect of ice accretion on flow field is used to solve the ice trajectory problem. The approach is based on mesh adaptation, penalization method and level sets.

4.5. Modeling study of Silicic acid flux to the ocean from tidal permeable sediments (in collaboration with EPOC)

Sandy sediments of tidal beaches are poor in reactive substances because they are regularly flushed by significant flow caused by tidal forcing. This transport process may significantly affect the flux of reactive solutes to the ocean. A two dimensional model coupling the Richards equation that describes the flow in permeable sediments and the conservation equation of the silicic acid was developed to simulate the evolution of the silicic acid concentration into a variably saturated porous media submitted to tidal forcing.

4.6. ORCs cycles

ORCs are Rankine cycles that use properly chosen low-boiling molecularly heavy organic compounds to drive the turbine in place of steam. This makes them suitable for the exploitation of low grade heat sources like biomass combustion, geothermal reservoirs and heat recovery from industrial processes. ORC turbines mainly use a single (less frequently, two) stage to expand the fluid. Up till present, no experimental data are available for flows of heavy fluids in the dense gas region. Experiments are difficult because of high temperature and pressure conditions, and fluid decomposition or inflammability in presence of air. This has motivated the use of numerical simulation as a preferential tool for dense gas flow analysis, but only a limited number of papers have been devoted to the computation of dense gas flows. With no experimental validation yet available for any of these configurations, care must be taken in the analysis of the computed flow fields because of their sensitivity to the thermodynamic model and to the numerical ingredients of the discretization scheme. Since no comparison with experimental data is possible, particular attention is devoted to code validation and model assessment. We created the platform ORComp, for computing some global performance metrics, and we applied some UQ and numerical methods for taking into account the sun variability in the design of ORCs cycles.

4.7. Cavitation

Cavitation consists in a local pressure drop below the vapor pressure at the liquid temperature, thus creating a phase change and vapor bubbles formation. Their collapse in high-pressure region can dramatically lead to failure, erosion and other undesirable effects. For this reason, there is a strong effort devoted to develop predictive numerical tools for cavitating flows in industrial applications. Unfortunately, an accurate description of interactions between the vapour and liquid phases requires accurate physical models and a way to take into account the dynamics of the interface. Moreover, multiscale effects, turbulence and thermodynamics should be also considered. Cavitation models are typically dependent on two types of parameters: first, on some physical parameters, such as for example the number of bubbles, that is not usually well measured; secondly, on some empiric parameters, useful for fitting and calibration procedures with respect to the experimental data. Therefore, model parameters represent an important source of uncertainty. Moreover, it is not an easy task to

well define boundary and initial conditions, because of difficulties encountered in order to control accurately experiments in cavitating flows. As a result, conditions imposed for the setting of a numerical simulation, are affected by a dramatic randomness. We performed a systematic study for considering the probabilistic properties of the input parameters permitting to capture non-linearities in uncertainty propagation. Moreover, the DEM method has been modified to take into account cavitation phenomena with real-gas effects.

4.8. Atmospheric entries of spacecraft

Simulation of atmospheric entries of spacecraft is a challenging problem involving many complex physical phenomena, including rarefied gas effects, aerothermochemistry, radiation, and the response of thermal protection materials to extreme conditions. The post-flight analysis of a space mission requires accurate determination of the freestream conditions for the trajectory, that is, temperature and pressure conditions and the Mach number in front of the shock. The latter can be rebuilt from the pressure and heat flux measured on the spacecraft by means of a Flush Air Data System (FADS). This instrumentation comprises a set of sensors flush mounted in the thermal protection system to measure the static pressure (pressure taps) and heat flux (calorimeters). In this context, Computational Fluid Dynamics (CFD) supplied with UQ tools permits to take into account chemical effects and to include both measurement errors and epistemic uncertainties on the chemical model parameters in the bulk and at the wall (surface catalysis). Rebuilding the freestream conditions from the FADS data therefore amounts to solving a stochastic inverse problem. In this context, we proposed a new methodology for solving the inverse problem based on a Bayesian setting, that is, probability densities of possible values of freestream conditions are rebuilt from stagnation-point pressure and heat flux measurements. A Bayesian setting offers a rigorous foundation for inferring input parameters from noisy data and uncertain forward models, a natural mechanism for incorporating prior information, and a quantitative assessment of uncertainty on the inferred results.

4.9. Shallow Water

Prediction of shallow water equations in realistic application depends on the level of complexity used for the physical modelling (such as for example, for the friction coefficient) and on a set of empirical coefficients that are usually chosen in order to fit the experimental data. Then, input environmental conditions, topography and modelling involve a certain degree of uncertainty. The capability to take into account these uncertainties in the numerical simulation is of great importance in order to correctly predict extreme flood events. Stochastic modeling of long-wave propagation demands a robust shallow-water model in order to characterize the physical processes. We coupled a residual distribution scheme for shallow water equations with some stochastic methods in order to take into account uncertainties in the numerical simulation. Preliminary results showed that influence of uncertainties is stronger after the phase interaction indicating the need for a stochastic simulation in order to have a correct prediction of the numerical solution.

BIPOP Project-Team

4. Application Domains

4.1. Introduction

Many systems (either actual or abstract) can be represented by (1). Some typical examples are:

- Mechanical systems with unilateral constraints and dry friction (the biped robot is a typical example, hair and fiber dynamics is another example).
- Electrical circuits with ideal diodes and/or transistors Mos.
- Optimal control with constraints on the state, closed loop of a system controlled by an MPC algorithm², discontinuous feedback controllers like sliding-mode control, etc.

This class of models is not too large (to allow thorough studies), yet rich enough to include many applications. This goes in contrast to a study of general hybrid systems. Note for example that (1) is a “continuous” hybrid system, in that the continuous variables x and u prevail in the evolution (there is no discrete control to commute from a mode to the other: only the input u can be used). The main tools for the analysis and simulation of such dynamical systems come from Convex Analysis, Non-smooth Analysis, Complementarity Theory (we make a strong use of complementarity problems solvers for numerical simulation), Variational Inequalities. Let us cite some specific applications.

4.2. Computational neuroscience

Modeling in neuroscience makes extensive use of nonlinear dynamical systems with a huge number of interconnected elements. Our current theoretical understanding of the properties of neural systems is mainly based on numerical simulations, from single cell models to neural networks. To handle correctly the discontinuous nature of integrate-and-fire networks, specific numerical schemes have to be developed. Our current works focus on event-driven, time-stepping and voltage-stepping strategies, to simulate accurately and efficiently neuronal networks. Our activity also includes a mathematical analysis of the dynamical properties of neural systems. One of our aims is to understand neural computation and to develop it as a new type of information science.

4.3. Electronic circuits

Whether they are integrated on a single substrate or as a set of components on a board, electronic circuits are very often a complex assembly of many basic components with non linear characteristics. The IC technologies now allow the integration of hundreds of millions of transistors switching at GHz frequencies on a die of 1cm^2 . It is out of question to simulate a whole such IC with standard tools such as the SPICE simulator. We currently work on a dedicated plug-in able to simulate a whole circuit comprising various components, some modelled in a nonsmooth way.

4.4. Walking robots

As compared to rolling robots, the walking ones – for example hexapods – possess definite advantages whenever the ground is not plane or free: clearing obstacles is easier, holding on the ground is lighter, adaptivity is improved. However, if the working environment of the system is adapted to man, the biped technology must be preferred, to preserve good displacement abilities without modifying the environment. This explains the interest displayed by the international community in robotics toward humanoid systems, whose aim is to back man in some of his activities, professional or others. For example, a certain form of help at home to disabled persons could be done by biped robots, as they are able to move without any special adaptation of the environment.

²model predictive control

4.5. Optimization

Optimization exists in virtually all economic sectors. Simulation tools can be used to optimize the system they simulate. Another domain is parameter *identification* (Idopt or Estime teams), where the deviation between measurements and theoretical predictions must be minimized. Accordingly, giving an exhaustive list of applications is impossible. Some domains where Inria has been implied in the past, possibly through the former Promath and Numopt teams are: production management, geophysics, finance, molecular modeling, robotics, networks, astrophysics, crystallography, ...Our current applicative activity includes: the management of electrical production (deterministic or stochastic), the design and operation of telecommunication networks.

4.6. Computer graphics Animation

A new application in Bipop is the simulation of complex scenes involving many interacting objects. Whereas the problem of collision detection has become a mature field those recent years, simulating the collision response (in particular frictional contacts) in a realistic, robust and efficient way, still remains an important challenge. Another related issue we began to study is the simulation of heterogeneous objects such as granular or fibrous materials, which requires the design of new high-scales models for dynamics and contacts; indeed, for such large systems, simulating each interacting particle/fiber individually would be too much time-consuming for typical graphics applications. We also pursue some study on the design of high-order models for slender structures such as rods, plates or shells. Finally, our current activity includes the static inversion of mechanical objects, which is of great importance in the field of artistic design, for the making of movies and video games for example. Such problems typically involve geometric fitting and parameters identification issues, both resolved with the help of constrained optimization.

CAD Team

4. Application Domains

4.1. Introduction

Our scientific results in geometry have been tested in Aircraft industry, Ceramic industry, N-C simulation, and Computer Graphics as well.

4.2. Geometry

The cooperation with EADS, based on our new B-Spline surface formulation, is very promising, both for complex shape modelling and numerical simulations. Tolerance problems are currently studied in the sino-french Tsinghua PLM Center (supported by Dassault System).

Furthermore, this project also allowed to promoting several Associated Professors and Post Doctors in Chinese and French Research Institutes.

4.3. Computer Graphics

In Computer Graphics, our work in Computer Animation (fluid dynamics) has been tested in the Mr. Zhiyi Zhang's company.

CAGIRE Team

4. Application Domains

4.1. Effusion cooling of aeronautical combustion chambers walls

The industrial applications of our project is the cooling of the walls of the combustion chambers encountered in the helicopter engines, and more precisely, we wish to contribute to the improvement of effusion cooling.

Effusion cooling is nowadays very widespread, especially in the aeronautical context. It consists in piercing holes on the wall of the combustion chamber. These holes induce cold jets that enter inside the combustion chamber. The goal of this jet is to form a film of air that will cool the walls of the chamber, see Figure 2 .

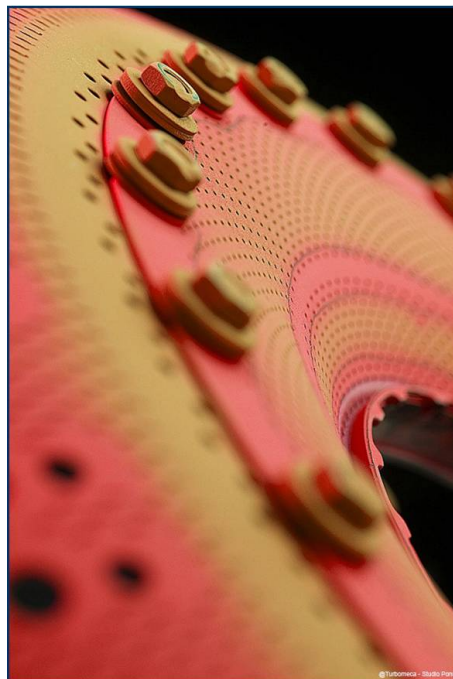


Figure 2. Effusion cooling of aeronautical combustion chambers: close view of a typical perforated chamber wall

Effusion cooling in a combustion chamber takes at the wall where thousands of small holes allow cool air to enter inside the combustion chamber. This induces jets in crossflow in charge of cooling the walls, whatever the heat and the acoustic waves present inside the chamber. Nevertheless, this technique is not straightforward to put in practice: the size, design and position of the holes can have an important effect on the cooling efficiency. For a safe and efficient functioning of the combustion chamber, it is required that the cooling jets and the combustion effects be as much independent as possible. For example, this means that

- The jets of cool air should not mix too much with the internal flow. Otherwise it will decrease the efficiency of the combustion.
- The jets should be as much stable as possible when submitted to waves emitted in the combustion chamber, e.g. acoustic waves induced by combustion instabilities. Otherwise the jets may not cool enough the walls of the combustion chamber which can then undergoes severe damages.

The first point is what we aim at simulate in this project. As the model chosen is the fully compressible Navier Stokes system, there should not be any problem in the future for being able to simulate the effect of an acoustic forcing on the jet in crossflow.

Having a database of Direct Numerical Simulations is also fundamental for testing closure laws that are used in turbulence models encountered in RANS and LES models. With such models, it is possible for example to perform optimisation.

A last aspect, that will not be dealt with in this project, but that could be dealt with in the future, is the interaction between the flow and the wall. The aim is to understand the effect of coupling between the heat propagation in the wall and the flow near the wall. A careful study of this interaction can allow to determine the exchange coefficients, and so the efficiency of the cooling by the jet. Such determination may particularly useful to develop one or multidimensional models of wall-fluid interaction [19].

From the application point of view, compressibility effects must be taken into account since the Mach number of the flow can reach values equal to 0.3, hence/or acoustic waves may be present inside the combustion chamber. This can raise a problem, because upwind numerical schemes are known to be less accurate in the low Mach limit.

CALVI Project-Team

4. Application Domains

4.1. Thermonuclear fusion

Inertial fusion, magnetic fusion, ITER, particle accelerators, laser-matter interaction

Controlled fusion is one of the major prospects for a long term source of energy. Two main research directions are studied: magnetic fusion where the plasma is confined in tokamaks using a large external magnetic field and inertial fusion where the plasma is confined thanks to intense laser or particle beams. The simulation tools we develop can be applied for both approaches.

Controlled fusion is one of the major challenges of the 21st century that can answer the need for a long term source of energy that does not accumulate wastes and is safe. The nuclear fusion reaction is based on the fusion of atoms like Deuterium and Tritium. These can be obtained from the water of the oceans that is widely available and the reaction does not produce long-term radioactive wastes, unlike today's nuclear power plants which are based on nuclear fission.

Two major research approaches are followed towards the objective of fusion based nuclear plants: magnetic fusion and inertial fusion. In order to achieve a sustained fusion reaction, it is necessary to confine sufficiently the plasma for a long enough time. If the confinement density is higher, the confinement time can be shorter but the product needs to be greater than some threshold value.

The idea behind magnetic fusion is to use large toroidal devices called tokamaks in which the plasma can be confined thanks to large applied magnetic field. The international project ITER ¹ is based on this idea and aims to build a new tokamak which could demonstrate the feasibility of the concept.

The inertial fusion concept consists in using intense laser beams or particle beams to confine a small target containing the Deuterium and Tritium atoms. The Laser Mégajoule which is being built at CEA in Bordeaux will be used for experiments using this approach.

Nonlinear wave-wave interactions are primary mechanisms by which nonlinear fields evolve in time. Understanding the detailed interactions between nonlinear waves is an area of fundamental physics research in classical field theory, hydrodynamics and statistical physics. A large amplitude coherent wave will tend to couple to the natural modes of the medium it is in and transfer energy to the internal degrees of freedom of that system. This is particularly so in the case of high power lasers which are monochromatic, coherent sources of high intensity radiation. Just as in the other states of matter, a high laser beam in a plasma can give rise to stimulated Raman and Brillouin scattering (respectively SRS and SBS). These are three wave parametric instabilities where two small amplitude daughter waves grow exponentially at the expense of the pump wave, once phase matching conditions between the waves are satisfied and threshold power levels are exceeded. The illumination of the target must be uniform enough to allow symmetric implosion. In addition, parametric instabilities in the underdense coronal plasma must not reflect away or scatter a significant fraction of the incident light (via SRS or SBS), nor should they produce significant levels of hot electrons (via SRS), which can preheat the fuel and make its isentropic compression far less efficient. Understanding how these deleterious parametric processes function, what non uniformities and imperfections can degrade their strength, how they saturate and interdepend, all can benefit the design of new laser and target configuration which would minimize their undesirable features in inertial confinement fusion. Clearly, the physics of parametric instabilities must be well understood in order to rationally avoid their perils in the varied plasma and illumination conditions which will be employed in the National Ignition Facility or LMJ lasers. Despite the thirty-year history of the field, much remains to be investigated.

¹ <http://www.iter.org>

Our work in modelling and numerical simulation of plasmas and particle beams can be applied to problems like laser-matter interaction, the study of parametric instabilities (Raman, Brillouin), the fast ignitor concept in the laser fusion research as well as for the transport of particle beams in accelerators. Another application is devoted to the development of Vlasov gyrokinetic codes in the framework of the magnetic fusion programme in collaboration with the Department of Research on Controlled Fusion at CEA Cadarache. Finally, we work in collaboration with the American Heavy Ion Fusion Virtual National Laboratory, regrouping teams from laboratories in Berkeley, Livermore and Princeton on the development of simulation tools for the evolution of particle beams in accelerators.

4.2. Nanophysics

Kinetic models like the Vlasov equation can also be applied for the study of large nano-particles as approximate models when *ab initio* approaches are too costly.

In order to model and interpret experimental results obtained with large nano-particles, *ab initio* methods cannot be employed as they involve prohibitive computational times. A possible alternative resorts to the use of kinetic methods originally developed both in nuclear and plasma physics, for which the valence electrons are assimilated to an inhomogeneous electron plasma. The LPMIA (Nancy) possesses a long experience on the theoretical and computational methods currently used for the solution of kinetic equation of the Vlasov and Wigner type, particularly in the field of plasma physics.

Using a Vlasov Eulerian code, we have investigated in detail the microscopic electron dynamics in the relevant phase space. Thanks to a numerical scheme recently developed by Filbet et al. [63], the fermionic character of the electron distribution can be preserved at all times. This is a crucial feature that allowed us to obtain numerical results over long times, so that the electron thermalization in confined nano-structures could be studied.

The nano-particle was excited by imparting a small velocity shift to the electron distribution. In the small perturbation regime, we recover the results of linear theory, namely oscillations at the Mie frequency and Landau damping. For larger perturbations nonlinear effects were observed to modify the shape of the electron distribution.

For longer time, electron thermalization is observed: as the oscillations are damped, the center of mass energy is entirely converted into thermal energy (kinetic energy around the Fermi surface). Note that this thermalization process takes place even in the absence of electron-electron collisions, as only the electric mean-field is present.

CASTOR Team (section vide)

CLASSIC Project-Team

4. Application Domains

4.1. Forecasting of the electricity consumption

Our partner is EDF R&D. The goal is to aggregate in a sequential fashion the forecasts made by some (about 20) base experts in order to predict the electricity consumption at a global level (the one of all French customers) at a half-hourly step. We need to abide by some operational constraints: the predictions need to be made at noon for the next 24 hours (i.e., for the next 48 time rounds).

4.2. Forecasting of the air quality

Our partner is the Inria project-team CLIME (Paris-Rocquencourt). The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to output field prediction of the concentration of some pollutants (typically, the ozone) over Europe. The results were and will be transferred to the public operator INERIS, which uses and will use them in an operational way.

4.3. Forecasting of the production data of oil reservoirs

Our partner is IFP Energies nouvelles. The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to predict some behaviors (gas/oil ratio, cumulative oil extracted, water cut) of the exploitation of some oil wells.

4.4. Data mining, massive data sets

Our partner is the start-up Safety Line. The purpose of this application is to investigate statistical learning strategies for mining massive data sets originated from aircraft high-frequency recordings and improve security.

4.5. Computational linguistics

We propose and study new language models that bridge the gap between models oriented towards the statistical analysis of large corpora and grammars oriented towards the description of syntactic features as understood by academic experts. We have conceived a new kind of grammar, based on some cut and paste mechanism and some label aggregation principle, that can be fully learnt from a corpus. We are currently testing this model and studying its mathematical properties.

4.6. Statistical inference on biological data

The question is about understanding how interactions between neurons can be detected. A mathematical modeling is given by multivariate Hawkes processes. Lasso-type methods can then be used to estimate interaction functions in the nonparametric setting by using fast algorithms, providing inference of the unitary event activity of individual neurons.

COFFEE Project-Team

4. Application Domains

4.1. Analysis of PDEs

We are led to consider coupled non linear systems of PDEs. We are mainly interested here in systems exhibiting hyperbolic structures, possibly with (partial) diffusive corrections. It includes non linear kinetic equations (of Vlasov-Fokker-Planck or Vlasov-Boltzmann type, say). It is also possible that the coefficients depend on the unknown in a non-local way. Generally speaking, we are concerned in the project with flows where the multiphase character is crucial and makes part of the difficulty. Models involve PDEs systems of hybrid type, and we expect to be able to exhibit common structures between several models coming from different physical motivation.

4.2. Numerical analysis of FV methods

We wish to design and analyze new numerical schemes, mostly in the FV framework. For hyperbolic systems the theory is well-advanced, but there remain many challenging questions, of crucial relevance for the applications:

- While the general FV framework is clear for conservation laws, the design of numerical fluxes and the discussion of stability issues can be “application-dependent”. In particular, we wish to use the underlying microscopic description of particulate and mixtures flows to design dedicated kinetic schemes.
- Stability issues become more intricate when we try to increase the consistency accuracy and when we deal with complex meshes. For instance preserving positivity of certain fields could be absolutely crucial not only for physical reasons, but also to preserve the stability of the simulation of a coupled system. We will therefore continue our work on MUSCL-like methods working on vertex-based discretization, with limiters defined by using multislope analysis.
- Finally, source terms have to be considered appropriately, in particular in order to preserve equilibrium states and to capture correct asymptotic regimes. It requires to UpWind the numerical fluxes by taking into account the source terms. Among others, we are particularly interested in Asymptotic High Order (AHO) schemes.

Besides, the conception of new methods based on Finite Volume discretizations for diffusion has known very intense activities after pioneering works at the beginning of the '00s. The FV framework, dealing with very general meshes, is very appropriate to deal with complex flows in highly heterogeneous media. The difficulty consists in defining additional unknowns to evaluate diffusion fluxes on the interfaces of the control volumes: using as unique numerical unknown the cell average of the continuous unknown requires unrealistic conditions on the mesh geometry. We are highly involved in the developments of such methods, which are strongly motivated by industrial needs (and our research is enhanced by a 12 years-long experience in the industrial context). We plan to investigate in particular the VAG (Vertex Approximate Gradient) method, based on cell center discretizations involving additional unknowns stored at the vertices, and the DDFV (Discrete Duality Finite Volume) method, which uses a dual mesh.

4.3. Asymptotic analysis

Asymptotic analysis has a crucial role in our activity. We are interested in the derivation of hierarchies of models based on asymptotic arguments: it allows to design reduced models, that can be relevant under well-identified assumptions of the physical coefficients. In particular, the discussion of hydrodynamic regimes which start from kinetic equations and lead to macroscopic equations will be the object of a specific attention.

4.4. Particulate flows, Mixture flows

We will investigate fluid mechanics models referred to as “multi-fluids” flows where a disperse phase interacts with a dense phase. Such flows arise in numerous applications, like for pollutant transport and dispersion, the combustion of fuel particles in air, the modeling of fluidized beds, the dynamic of sprays and in particular biosprays with medical applications, engine fine particles emission... There are many possible modeling of such flows: microscopic models where the two phases occupy distinct domains and where the coupling arises through intricate interface conditions; macroscopic models which are of hydrodynamic (multiphase) type, involving non standard state laws, possibly with non conservative terms, and the so-called mesoscopic models. The latter are based on Eulerian–Lagrangian description where the disperse phase is described by a particle distribution function in phase space. Following this path we are led to a Vlasov-like equation coupled to a system describing the evolution of the dense phase that is either the Euler or the Navier-Stokes equations

4.5. Interface problems

We are faced difficulties related to the coupling of models, methods and codes at interfaces. We distinguish several questions, which are nevertheless intimately connected on the technical viewpoint.

- As far as we consider flows in porous media, interface conditions and domain decomposition methods have been thoroughly investigated. However, the question of their numerical treatment left many questions open, which naturally depend on the underlying discretization techniques. A typical question, which is strongly motivated by collaborations with our industrial partners, relies on the simulations of mass and heat exchanges between a porous medium and an adjacent free-flow region. It is also remarkable that, despite the existing literature on the subject, techniques have still a reduced spreading out of the academics: in many industrial simulations, two different (commercial) codes are used and the interface coupling is managed more or less “manually” at each time step! Therefore, the challenge consists in dealing with compositional non-isothermal two-phase systems, including vaporization effects, and, again, the application fields make the use of FV schemes in the porous media more appropriate.
- The situation is a bit different for hyperbolic problems because the design of the interface condition itself is less clear. We are concerned with the coupling of hyperbolic systems involving different propagation properties, and possibly different set of unknowns. It raises modeling issues in order to design the coupling condition, problems of mathematical analysis, as well as complicated questions in order to match numerical methods. The framework of kinetic schemes might be a possible way to define consistent numerical fluxes. The problem is again strongly motivated by industrial needs, with the additional constraint of not to modify too much existing computational tools. We are involved in a collaboration with physicists specifically dedicated to such problems of wave propagations through complex interfaces.

4.6. Biological degradation

biodegradation of monuments is due, in part, to the formation of biofilms, namely a colony of bacteria embedded within an extra-cellular matrix. Biofilms can also be used as a protection device against corrosion of well cement in CO₂ storage reservoir. More generally the formation of biofilms is a common feature of the behavior of bacteria and has potentially many applications in medical and industrial settings; for instance, the cyanobacteria are seriously considered in order to produce energy as bio-fuel and there are also researches to set up bio-devices to avoid human or plant diseases. We are particularly interested in mathematical models of such phenomena based on arguments coming from mixture theory (thus with a natural connection to the previous item); it leads to a complex multi-dimensional hydrodynamic-type system, with polyphasic features.

Besides, when considering proliferation of micro-algae in a large domain, it is relevant to distinguish two phases : a development one on the sea bed as a biofilm and a spreading one in water which can be described thanks to kinetic equations subject to coagulation-fragmentation dynamics. We wish to derive a complete system, describing the two phases, including the design of coupling interface conditions. This is definitely an

interesting and original modeling challenge. We also wish to identify scaling parameters which will allow to bring out hierarchies of reduced models. Of course, the program has to be completed with the conception of the corresponding numerical schemes, so that we will be able to validate, at least on qualitative grounds, our approach, which, in turn, will be decisive to strengthen the collaborations with biologists.

Another question, which is equally related technically to the other problems addressed in the project, is concerned with the analysis and simulation of equations of hyperbolic type in inhomogeneous media, like porous media or networks. This is a direction to improve the existing models in biology and it can give rise in analytical and numerical viewpoints to fruitful exchanges between the biological domain and the environmental one. We are particularly interested in PDEs describing chemotactic behaviors, namely the movement of cells in response to a chemical signal and have potential applications, for example to model the movement of fibroblasts on scaffold

COMMANDS Project-Team

4. Application Domains

4.1. Introduction

Commands is a team with a strong commitment in tackling real-life applications in addition to theoretical challenges. This shows in our long history of contracts with industrial partners. In the recent years, we have mainly contributed to the following fields of application.

4.2. Aerospace applications

In the framework of a long-term partnership with the Cnes, and more recently Astrium, we have studied trajectory optimization for space launcher problems. This kind of problems typically involves hard constraints (thermal flux, mechanical efforts) and inexact models (atmosphere, aerodynamic forces). The two main achievements were to study when singular arcs may occur, and to show the effectiveness of a HJB approach on a reduced model. Singular arcs are flight phases with a non-maximal thrust, induced by a tradeoff between speed and atmospheric drag; they cause difficulties of both theoretical and practical nature. The latter point is the first step in the process of applying global methods to this class of difficult problems.

4.3. Trading applications

In a partnership with Total, we have studied problems dealing with the trading of Liquefied Natural Gas. We have computed maximizing revenue policies, by combining the Stochastic Dual Dynamic Programming approach (SDDP) with a quantization method for the noise that enters in prices. We have also given partial results for the case of integer decision.

4.4. Energy applications

With Renault, we have studied problems of energy management for hybrid vehicles. Hybrid vehicles include an auxiliary thermal (gas) engine that is used as a range extender for the main electric propulsion. We are interested in determining the optimal policies for energy management, taking into account some stochastic uncertainties, as well as execution delay and decision lags.

CONCHA Project-Team

4. Application Domains

4.1. Aerodynamics

Aerodynamics provide a challenging field for numerical simulations in fluid dynamics with a wide range of applications. Robustness of the simulation software with respect to physical parameters as the Reynolds and Mach numbers is necessary condition. In general, realistic simulations need to be done in three dimensions, which makes the efficiency of the numerical approach and implementation a question of feasibility. Therefore, different efforts are made in this project in order to tackle these subjects.

4.2. Red blood cells

Hammou El-Otmany started his PhD thesis in October 2012 in our group, supervised by D. Capatina and D. Graebler. The thesis is financed by UPPA (50%) and CDAPP (50%) and concerns the numerical simulation of biological fluids flows. We will focus more particularly on the physical and numerical modeling of red blood cells.

Clinically, some pathologies such as drepanocytosis or sickle cell anemia are due to the abnormal form of red blood cells. In the microcirculation, where cells must deform to pass through narrow capillaries, the deformability of individual red blood cells is a major determinant of resistance to flow.

The goal is twofold. On the one hand, we want to propose a realistic modeling of red blood cells in artery flow, by taking into account the membrane's viscoelasticity and thus, its deformability. The latter is essentially linked to its structure (i.e. its cellular geometry, membrane properties and cytoplasmic viscosity); thus structural abnormalities, as found in some haematological disorders can be expected to affect blood flow in the microcirculation and/or red cell lifespan.

On the other hand, we want to develop an efficient and stable numerical method in order to treat the coupling between the different models involved: Navier-Stokes for the matrix (blood) and for the cytoplasm (interior of the cell) and a non-Newtonian fluid (for instance, Giesekus) for the membrane. We will use the NXFEM method to take into account the interfaces between fluids.

4.3. Heat transfer

Heat transfer problems involve the coupling of the flow field of the fluid with temperature inside the flow and possibly on the boundary of the flow domain. A typical example of a heat transfer problem is the cooling of a combustion engine, see the project Optimal described in Section 7.1 .

4.4. Turbulence

Turbulent flows are ubiquitous in industrial applications. Direct numerical simulation (DNS), which aims at complete resolution of the flow field up to the Kolmogorov scale, has historically been limited to very simple geometries. The increase of computational power and the development of specialized numerical methods open the door to a wider range of applications. However, for most applications of practical interest, the use of some kind of turbulence modeling is unavoidable in order to obtain the prediction of averaged values and commercial software is in general based on such approaches combined with wall laws. In many applications, such as the project Optimal, see Section 7.1 , the Reynolds number is at an intermediate level, which means that the turbulence is not fully developed, and the heuristics behind most turbulence models are questionable. Especially, in heat transfer problems, the usage of wall laws seems to considerably lower the accuracy of the predicted mean values. In order to improve the computation of such values, we are particularly interested in variational multiscale methods and its relations to stabilized finite element methods.

4.5. Flows in porous media

Flows in fractured porous media are very important in petroleum engineering. They represent a good framework for the application of the tools developed in the CONCHA library such as the NXFEM method, goal oriented adaptivity, multiscale coupling of different models and multilevel solvers.

CORIDA Project-Team

4. Application Domains

4.1. Panorama

As we already stressed in the previous sections the robust control of infinite dimensional systems is an emerging theory. Our aim is to develop tools applicable to a large class of problems which will be tested on models of increasing complexity. We describe below only the applications in which the members of our team have recently obtained important achievements.

4.2. Biology and Medicine

4.2.1. Medicine

We began this year to study a new class of applications of observability theory. The investigated issues concern inverse problems in Magnetic Resonance Imaging (MRI) of moving bodies with emphasis on cardiac MRI. The main difficulty we tackle is due to the fact that MRI is, comparatively to other cardiac imaging modalities, a slow acquisition technique, implying that the object to be imaged has to be still. This is not the case for the heart where physiological motions, such as heart beat or breathing, are of the same order of magnitude as the acquisition time of an MRI image. Therefore, the assumption of sample stability, commonly used in MRI acquisition, is not respected. The violation of this assumption generally results in flow or motion artifacts. Motion remains a limiting factor in many MRI applications, despite different approaches suggested to reduce or compensate for its effects Welch et al. [68]. Mathematically, the problem can be stated as follows: can we reconstruct a moving image by measuring at each time step a line of its Fourier transform? From a control theoretic point of view this means that we want to identify the state of a dynamical system by using an output which is a small part of its Fourier transform (this part may change during the measurement).

There are several strategies to overcome these difficulties but most of them are based on respiratory motion suppression with breath-hold. Usually MRI uses ECG information to acquire an image over multiple cardiac cycles by collecting segments of Fourier space data at the same delay in the cycle Lanzer et al. [60], assuming that cardiac position over several ECG cycles is reproducible. Unfortunately, in clinical situations many subjects are unable to hold their breath or maintain stable apnea. Therefore breath-holding acquisition techniques are limited in some clinical situations. Another approach, so called real-time, uses fast, but low resolution sequences to be faster than heart motion. But these sequences are limited in resolution and improper for diagnostic situations, which require small structure depiction as for coronary arteries.

4.3. Simulation of viscous fluid-structure interactions

Participants: Bruno Pinçon, Jean-François Scheid [correspondant].

A number of numerical codes for the simulation for fluids and fluid-structure problems has been developed by the team. These codes are mainly written in MATLAB Software with the use of C++ functions in order to improve the sparse array process of MATLAB. We have focused our attention on 3D simulations which require large CPU time resources as well as large memory storage. An efficient 3D Stokes sparse solver for MATLAB is now available. An important work has been performed for the study and the development of a class of preconditioners for iterative solver of 3D Stokes problem. Efficient preconditioner of block preconditioned conjugate gradient type (BPCG) is now implemented. The use of this preconditioner significantly reduces the CPU time for the solution of linear system coming from the Stokes equations. This work has been developed in collaboration with Marc Fuentes, research engineer at Inria Nancy Grand Est. M. Fuentes has also written a PYTHON version of the 3D Stokes solver. A 3D characteristics method for the nonlinear Navier-Stokes equations is now in progress

4.4. Biohydrodynamics MATLAB Toolbox (BHT)

Participants: Alexandre Munnier [correspondant], Bruno Pinçon.

Understanding the locomotion of aquatic animals fascinated the scientific community for a long time. This constant interest has grown from the observation that aquatic mammals and fishes evolved swimming capabilities superior to what has been achieved by naval technology. A better understanding of the biomechanics of swimming may allow one to improve the efficiency, manoeuvrability and stealth of underwater vehicles. During the last fifty years, several mathematical models have been developed. These models make possible the qualitative analysis of swimming propulsion as a continuation of the previously developed quantitative theories. Based on recent mathematical advances, Biohydrodynamics MATLAB Toolbox (BHT) is a collection of M-Files for design, simulation and analysis of articulated bodies' motions in fluid. More widely, BHT allows also to perform easily any kind of numeric experiments addressing the motion of solids in ideal fluids (simulations of so-called fluid-structure interaction systems).

This software is available at <http://bht.gforge.inria.fr/>.

CQFD Project-Team

4. Application Domains

4.1. Dependability and safety

Our abilities in probability and statistics apply naturally to industry in particular in studies of dependability and safety.

An illustrative example which gathers all the topics of team is a collaboration started in May 2010 with Thales Optronique on the subject of *optimization of the maintenance of a digital camera equipped with HUMS* (Health Unit Monitoring Systems). This subject is very interesting for us because it combines many aspects of our project. Classification tools will be used to select significant variables as the first step in the modeling of a digital camera. The model will then be analysed and estimated in order to optimize the maintenance.

A second example concerns the optimization of the maintenance date for an aluminum metallic structure subject to corrosion. It is a structure of strategic ballistic missile that is stored in a nuclear submarine missile launcher in peace-time and inspected with a given periodicity. The requirement for security on this structure is very strong. The mechanical stress exerted on the structure depends on its thickness. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the break.

A third example is the minimization of the acoustic signature of a submarine. The submarine has to chose its trajectory in order to minimize at each time step its observability by a surface ship following an unknown random trajectory.

However the spectrum of applications of the topics of the team is larger and may concern many other fields. Indeed non parametric and semi-parametric regression methods can be used in biometry, econometrics or engineering for instance. Gene selection from microarray data and text categorization are two typical application domains of dimension reduction among others. We had for instance the opportunity via the scientific program PRIMEQUAL to work on air quality data and to use dimension reduction techniques as principal component analysis (PCA) or positive matrix factorization (PMF) for pollution sources identification and quantization.

DEFI Project-Team

4. Application Domains

4.1. Radar and GPR applications

Conventional radar imaging techniques (ISAR, GPR, etc.) use backscattering data to image targets. The commonly used inversion algorithms are mainly based on the use of weak scattering approximations such as the Born or Kirchhoff approximation leading to very simple linear models, but at the expense of ignoring multiple scattering and polarization effects. The success of such an approach is evident in the wide use of synthetic aperture radar techniques.

However, the use of backscattering data makes 3-D imaging a very challenging problem (it is not even well understood theoretically) and as pointed out by Brett Borden in the context of airborne radar: “In recent years it has become quite apparent that the problems associated with radar target identification efforts will not vanish with the development of more sensitive radar receivers or increased signal-to-noise levels. In addition it has (slowly) been realized that greater amounts of data - or even additional “kinds” of radar data, such as added polarization or greatly extended bandwidth - will all suffer from the same basic limitations affiliated with incorrect model assumptions. Moreover, in the face of these problems it is important to ask how (and if) the complications associated with radar based automatic target recognition can be surmounted.” This comment also applies to the more complex GPR problem.

Our research themes will incorporate the development, analysis and testing of several novel methods, such as sampling methods, level set methods or topological gradient methods, for ground penetrating radar application (imaging of urban infrastructures, landmines detection, underground waste deposits monitoring, ...) using multistatic data.

4.2. Biomedical imaging

Among emerging medical imaging techniques we are particularly interested in those using low to moderate frequency regimes. These include Microwave Tomography, Electrical Impedance Tomography and also the closely related Optical Tomography technique. They all have the advantage of being potentially safe and relatively cheap modalities and can also be used in complementarity with well established techniques such as X-ray computed tomography or Magnetic Resonance Imaging.

With these modalities tissues are differentiated and, consequentially can be imaged, based on differences in dielectric properties (some recent studies have proved that dielectric properties of biological tissues can be a strong indicator of the tissues functional and pathological conditions, for instance, tissue blood content, ischemia, infarction, hypoxia, malignancies, edema and others). The main challenge for these functionalities is to build a 3-D imaging algorithm capable of treating multi-static measurements to provide real-time images with highest (reasonably) expected resolutions and in a sufficiently robust way.

Another important biomedical application is brain imaging. We are for instance interested in the use of EEG and MEG techniques as complementary tools to MRI. They are applied for instance to localize epileptic centers or active zones (functional imaging). Here the problem is different and consists into performing passive imaging: the epileptic centers act as electrical sources and imaging is performed from measurements of induced currents. Incorporating the structure of the skull is primordial in improving the resolution of the imaging procedure. Doing this in a reasonably quick manner is still an active research area, and the use of asymptotic models would offer a promising solution to fix this issue.

4.3. Non destructive testing and parameter identification

One challenging problem in this vast area is the identification and imaging of defaults in anisotropic media. For instance this problem is of great importance in aeronautic constructions due to the growing use of composite materials. It also arises in applications linked with the evaluation of wood quality, like locating knots in timber in order to optimize timber-cutting in sawmills, or evaluating wood integrity before cutting trees. The anisotropy of the propagative media renders the analysis of diffracted waves more complex since one cannot only relies on the use of backscattered waves. Another difficulty comes from the fact that the micro-structure of the media is generally not well known a priori.

Our concern will be focused on the determination of qualitative information on the size of defaults and their physical properties rather than a complete imaging which for anisotropic media is in general impossible. For instance, in the case of homogeneous background, one can link the size of the inclusion and the index of refraction to the first eigenvalue of so-called interior transmission problem. These eigenvalues can be determined from the measured data and a rough localization of the default. Our goal is to extend this kind of idea to the cases where both the propagative media and the inclusion are anisotropic. The generalization to the case of cracks or screens has also to be investigated.

In the context of nuclear waste management many studies are conducted on the possibility of storing waste in a deep geological clay layer. To assess the reliability of such a storage without leakage it is necessary to have a precise knowledge of the porous media parameters (porosity, tortuosity, permeability, etc.). The large range of space and time scales involved in this process requires a high degree of precision as well as tight bounds on the uncertainties. Many physical experiments are conducted *in situ* which are designed for providing data for parameters identification. For example, the determination of the damaged zone (caused by excavation) around the repository area is of paramount importance since microcracks yield drastic changes in the permeability. Level set methods are a tool of choice for characterizing this damaged zone.

4.4. Diffusion MRI

In biological tissues, water is abundant and magnetic resonance imaging (MRI) exploits the magnetic property of the nucleus of the water proton. The imaging contrast (the variations in the grayscale in an image) in standard MRI can be from either proton density, T1 (spin-lattice) relaxation, or T2 (spin-spin) relaxation and the contrast in the image gives some information on the physiological properties of the biological tissue at different physical locations of the sample. The resolution of MRI is on the order of millimeters: the grayscale value shown in the imaging pixel represents the volume-averaged value taken over all the physical locations contained that pixel.

In diffusion MRI, the image contrast comes from a measure of the average distance the water molecules have moved (diffused) during a certain amount of time. The Pulsed Gradient Spin Echo (PGSE) sequence is a commonly used sequence of applied magnetic fields to encode the diffusion of water protons. The term 'pulsed' means that the magnetic fields are short in duration, and the term gradient means that the magnetic fields vary linearly in space along a particular direction. First, the water protons in tissue are labelled with nuclear spin at a precession frequency that varies as a function of the physical positions of the water molecules via the application of a pulsed (short in duration, lasting on the order of ten milliseconds) magnetic field. Because the precessing frequencies of the water molecules vary, the signal, which measures the aggregate phase of the water molecules, will be reduced due to phase cancellations. Some time (usually tens of milliseconds) after the first pulsed magnetic field, another pulsed magnetic field is applied to reverse the spins of the water molecules. The time between the applications of two pulsed magnetic fields is called the 'diffusion time'. If the water molecules have not moved during the diffusion time, the phase dispersion will be reversed, hence the signal loss will also be reversed, the signal is called refocused. However, if the molecules have moved during the diffusion time, the refocusing will be incomplete and the signal detected by the MRI scanner is weaker than if the water molecules have not moved. This lack of complete refocusing is called the signal attenuation and is the basis of the image contrast in DMRI. The pixels showing more signal attenuation is associated with further water displacement during the diffusion time, which may be linked to physiological factors, such as higher cell membrane permeability, larger cell sizes, higher extra-cellular volume fraction.

We model the nuclear magnetization of water protons in a sample due to diffusion-encoding magnetic fields by a multiple compartment Bloch-Torrey partial differential equation, which is a diffusive-type time-dependent PDE. The DMRI signal is the integral of the solution of the Bloch-Torrey PDE. In a homogeneous medium, the intrinsic diffusion coefficient D will appear as the slope of the semi-log plot of the signal (in appropriate units). However, because during typical scanning times, $50 - 100ms$, water molecules have had time to travel a diffusion distance which is long compared to the average size of the cells, the slope of the semi-log plot of the signal is in fact a measure of an 'effective' diffusion coefficient. In DMRI applications, this measured quantity is called the 'apparent diffusion coefficient' (ADC) and provides the most commonly used form the image contrast for DMRI. This ADC is closely related to the effective diffusion coefficient obtainable from mathematical homogenization theory.

DISCO Project-Team

4. Application Domains

4.1. Control of engineering systems

The team considers control problems in the aeronautic area and studies delay effects in automatic visual tracking on mobile carriers.

4.2. Analysis and Control of life sciences systems

The team is also involved in life sciences applications. The two main lines are the analysis of bioreactors models and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML).

4.3. Energy Management

The team is interested in Energy management and considers optimization and control problems in energy networks.

DOLPHIN Project-Team

4. Application Domains

4.1. Academic generic problems

In this project, some well known optimization problems are re-visited in terms of multi-objective modelization and resolution:

- **Workshop optimization problems:**

Workshop optimization problems deal with optimizing the production. In this project, two specific problems are under study.

- **Flow-shop scheduling problem:** The flow-shop problem is one of the most well-known scheduling problems. However, most of the works in the literature use a mono-objective model. In general, the minimized objective is the total completion time (makespan). Many other criteria may be used to schedule tasks on different machines: maximum tardiness, total tardiness, mean job flowtime, number of delayed jobs, maximum job flowtime, etc. In the DOLPHIN project, a bi-criteria model, which consists in minimizing the makespan and the total tardiness, is studied. A tri-criteria flow-shop problem, minimizing in addition the maximum tardiness, is also studied. It allows us to develop and test multi-objective (and not only bi-objective) exact methods.
- **Cutting problems:** Cutting problems occur when pieces of wire, steel, wood, or paper have to be cut from larger pieces. The objective is to minimize the quantity of lost material. Most of these problems derive from the classical one-dimensional cutting-stock problem, which have been studied by many researchers. The problem studied by the DOLPHIN project is a two-dimensional bi-objective problem, where rotating a rectangular piece has an impact on the visual quality of the cutting pattern. First we have to study the structure of the cutting-stock problem when rotation is allowed, then we will develop a method dedicated to the bi-objective version of the problem.

- **Logistics and transportation problems:**

- **Packing problems:** In logistic and transportation fields, packing problems may be a major issue in the delivery process. They arise when one wants to minimize the size of a warehouse or a cargo, the number of boxes, or the number of vehicles used to deliver a batch of items. These problems have been the subjects of many papers, but only few of them study multi-objective cases, and to our knowledge, never from an exact point of view. Such a case occurs for example when some pairs of items cannot be packed in the same bin. The DOLPHIN project is currently studying the problem in its one-dimensional version. We plan to generalize our approach to two and three dimensional problems, and to more other conflict constraints, with the notion of *distance* between items.
- **Routing problems:** The vehicle routing problem (VRP) is a well-known problem and it has been studied since the end of the 50's. It has a lot of practical applications in many industrial areas (ex. transportation, logistics, etc). Existing studies of the VRP are almost all concerned with the minimization of the total distance only. The model studied in the DOLPHIN project introduces a second objective, whose purpose is to balance the length of the tours. This new criterion is expressed as the minimization of the difference between the length of the longest tour and the length of the shortest tour. As far as we know, this model is one of the pioneer work in the literature.

The second routing problem is a generalization of the covering tour problem (CTP). In the DOLPHIN project, this problem is solved as a bi-objective problem where a set of constraints are modeled as an objective. The two objectives are: i) minimization of the length of the tour; ii) minimization of the largest distance between a node to be covered and a visited node. As far as we know, this study is among the first works that tackle a classic mono-objective routing problem by relaxing constraints and building a more general MOP.

The third studied routing problem is the Ring Star Problem (RSP). This problem consists in locating a simple cycle through a subset of nodes of a graph while optimizing two kinds of costs. The first objective is to minimize a ring cost that is related to the length of the cycle. The second one is to minimize an assignment cost from non-visited nodes to visited ones. In spite of its natural bi-criteria formulation, this problem has always been studied in a single-objective form where either both objectives are combined or one objective is treated as a constraint.

Now our goal is to propose VRP libraries that are able to cope with a large number of constraints. The libraries are conceived in such a way that totally new sets of constraints can be added dynamically, similarly to what is done in constraint programming. This code will be used in future industrial collaborations, and already includes algorithms to use GIS.

For all studied problems, standard benchmarks have been extended to the multi-objective case. The benchmarks and the obtained results (optimal Pareto front, best known Pareto front) are available on the Web pages associated to the project and from the MCDM (International Society on Multiple Criteria Decision Making) Web site. This is an important issue to encourage comparison experiments in the research community.

4.2. Application to mobile telecommunication networks

With the extraordinary success of mobile telecommunication systems, service providers have been affording huge investments for network design and infrastructure. Mobile network design is of outmost importance, and is thus a major issue in mobile telecommunication systems. In fact, with the continuous and rapid growth of communication traffic, large scale planning becomes more and more difficult. Hence, automatic, interactive and self-adaptive optimization algorithms and tools would be very useful and helpful. Advances in this area will certainly lead to important improvements in terms of quality of service, network management and cost deployment.

In the past, the DOLPHIN team has initiated solid industrial collaborations within the domain of mobile networks. In fact, the problem of network design and frequency assignment was studied in collaboration with France Telecom. In particular, a new formulation/resolution of the problem as a multi-objective constrained combinatorial optimization problem was considered. In collaboration with Mobinets, the DOLPHIN team has also addressed the problem of access network design. The problem consists in minimizing the cost of the access network and maximizing its availability.

More recently, the DOLPHIN team has been interested in new optimization models and algorithms to address new difficult problems raised by new emerging technologies in wireless networks. In fact, wireless communications are evolving from inflexible and monolithic systems to a composite radio environment made of cognitive radio devices and networks of different technologies. Within this context, the challenge is to design new optimization techniques which are not only resource, power, scale, and applications aware, but which are self-adaptive and fully distributed in order to allow the dynamic optimization of radio-devices behaviors depending on the environment constraints e.g., spectrum availability, network traffic, user demand, etc. To achieve this goal, distributed and nature-inspired algorithms, such as ant-colony and bees, will be investigated in order to dynamically and distributively optimize predefined criterion such as throughput, fairness, quality of service to cite a few. It is expected that the techniques developed in this work will lead to the design of new models and algorithms for opportunistic/dynamic spectrum access and cross layer network optimization which are at the core of future generation wireless networks.

4.3. Application to Bioinformatics

Bioinformatic research is a great challenge for our society and numerous research entities of different specialities (biology, medical or information technology) are collaborating on specific themes.

4.3.1. Genomic and post-genomic studies

Previous studies of the DOLPHIN project mainly deal with genomic and postgenomic applications. These have been realized in collaboration with academic and industrial partners (IBL: Biology Institute of Lille; IPL: Pasteur Institute of Lille; IT-Omics firm).

First, genomic studies aim to analyze genetic factors which may explain multi-factorial diseases such as diabetes, obesity or cardiovascular diseases. The scientific goal was to formulate hypotheses describing associations that may have any influence on diseases under study.

Secondly, in the context of post-genomic, a very large amount of data are obtained thanks to advanced technologies and have to be analyzed. Hence, one of the goals of the project was to develop analysis methods in order to discover knowledge in data coming from biological experiments.

These problems can be modeled as classical datamining tasks (Association rules, feature selection). As the combinatoric of such problems is very high and the quality criteria not unique, we proposed to model these problems as multi-objective combinatorial optimization problems. Evolutionary approaches have been adopted in order to cope with large scale problems.

Nowadays the technology is still going fast and the number of data increases rapidly. Within the new collaboration, started in 2010, with Genes Diffusion, specializes in genetics and animal reproduction for bovine, swine, equine and rabbit species, we will study combinations of Single Nucleotide Polymorphisms (SNP) that can explain some phenotypic characteristics.

4.3.2. Docking and conformational sampling

In molecular modelling, conformational sampling and docking procedures provide help for understanding the interaction mechanisms between (macro)molecules involved in physiological processes. The processes to be simulated are of a combinatorial complexity (molecule size, number of degrees of freedom) which represents an important challenge for the currently available computing power. Such a challenge can be expressed by three major objectives: (1) the proposition of mathematical models of maximum simplicity that nevertheless provide a relevant description of molecular behavior, (2) the development of powerful distributed optimization algorithms (evolutionary algorithms, local search methods, hybrid algorithms) for sampling the molecular energy surface for stable, populated conformations, and (3) the deployment of those intrinsic distributed algorithms on computational Grids.

Within the framework of ANR DOCK and Decrypton projects, the focus was to propose multi-objective formulations of the conformational and docking problems. The goal was to take into account different criteria characteristics of the complex docking process. Furthermore, in order to deal with the multimodal nature of the problems it is important to define new hybrid mechanisms allowing us to provide algorithms with both diversification and intensification properties. Finally, to deal with the exponential combinatorics of these problems when large proteins are concerned parallel and grid computing is highly required. Using grid computing is not straightforward, so a "gridification" process is necessary. Such process allows us to adapt the proposed algorithms to the characteristics of the grid. The gridification process must be exploited by the user in a transparent way. Therefore, coupling ParadisEO-PEO with a generic grid middleware such as Globus is important to provide robust and efficient algorithms to be exploited transparently.

New contacts with the Servier company show that these questions are really challenging ones for the design of new drug molecules.

4.3.3. Optimization for health care

The new collaboration (PhD thesis started in october 2010) with Alicante company, major actor in the hospital decision making, will deal with knowledge extraction by optimization methods for improving the process of inclusion in clinical trials. Indeed, conducting a clinical trial, allowing for example to measure the effectiveness of a treatment, involves selecting a set of patients likely to participate to this test. Currently existing selection processes are far from optimal, and many potential patients are not considered. The objective of this collaboration consists in helping the practitioner to quickly determine if a patient is interesting for a clinical trial or not. Exploring different data sources (from a hospital information system, patient data...), a set of decision rules have to be generated. For this, approaches from combinatorial optimization will be implemented, requiring extensive work to model the problem, to define criteria optimization and to design specific optimization methods.

GAMMA3 Project-Team (section vide)

GECO Team

4. Application Domains

4.1. Quantum control

The issue of designing efficient transfers between different atomic or molecular levels is crucial in atomic and molecular physics, in particular because of its importance in those fields such as photochemistry (control by laser pulses of chemical reactions), nuclear magnetic resonance (NMR, control by a magnetic field of spin dynamics) and, on a more distant time horizon, the strategic domain of quantum computing. This last application explicitly relies on the design of quantum gates, each of them being, in essence, an open loop control law devoted to a prescribed simultaneous control action. NMR is one of the most promising techniques for the implementation of a quantum computer.

Physically, the control action is realized by exciting the quantum system by means of one or several external fields, being them magnetic or electric fields. The resulting control problem has attracted increasing attention, especially among quantum physicists and chemists (see, for instance, [88], [93]). The rapid evolution of the domain is driven by a multitude of experiments getting more and more precise and complex (see the recent review [48]). Control strategies have been proposed and implemented, both on numerical simulations and on physical systems, but there is still a large gap to fill before getting a complete picture of the control properties of quantum systems. Control techniques should necessarily be innovative, in order to take into account the physical peculiarities of the model and the specific experimental constraints.

The area where the picture got clearer is given by finite dimensional linear closed models.

- **Finite dimensional** refers to the dimension of the space of wave functions, and, accordingly, to the finite number of energy levels.
- **Linear** means that the evolution of the system for a fixed (constant in time) value of the control is determined by a linear vector field.
- **Closed** refers to the fact that the systems are assumed to be totally disconnected from the environment, resulting in the conservation of the norm of the wave function.

The resulting model is well suited for describing spin systems and also arises naturally when infinite dimensional quantum systems of the type discussed below are replaced by their finite dimensional Galerkin approximations. Without seeking exhaustiveness, let us mention some of the issues that have been tackled for finite dimensional linear closed quantum systems:

- controllability [29],
- bounds on the controllability time [25],
- STIRAP processes [98],
- simultaneous control [71],
- optimal control ([67], [38], [50]),
- numerical simulations [77].

Several of these results use suitable transformations or approximations (for instance the so-called rotating wave) to reformulate the finite-dimensional Schrödinger equation as a sub-Riemannian system. Open systems have also been the object of an intensive research activity (see, for instance, [30], [68], [89], [44]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [34], [45], [63], [35]). The controllability issue is less understood than in the finite dimensional setting, but several advances should be mentioned. First of all, it is known that one cannot expect exact controllability on the whole Hilbert sphere [97]. Moreover, it has been shown that a relevant model, the quantum oscillator, is not even approximately controllable [90], [80]. These negative results have been more recently completed by positive ones. In [36], [37] Beauchard and Coron obtained the first positive controllability result for a quantum particle in a 1D potential well. The result is highly nontrivial and is based on Coron's return method (see [52]). Exact controllability is proven to hold among regular enough wave functions. In particular, exact controllability among eigenfunctions of the uncontrolled Schrödinger operator can be achieved. Other important approximate controllability results have then been proved using Lyapunov methods [79], [84], [64]. While [79] studies a controlled Schrödinger equation in \mathbb{R} for which the uncontrolled Schrödinger operator has mixed spectrum, [84], [64] deal mainly with general discrete-spectrum Schrödinger operators.

In all the positive results recalled in the previous paragraph, the quantum system is steered by a single external field. Different techniques can be applied in the case of two or more external fields, leading to additional controllability results [55], [41].

The picture is even less clear for nonlinear models, such as Gross–Pitaevski and Hartree–Fock equations. The obstructions to exact controllability, similar to the ones mentioned in the linear case, have been discussed in [61]. Optimal control approaches have also been considered [33], [46]. A comprehensive controllability analysis of such models is probably a long way away.

4.2. Neurophysiology

At the interface between neurosciences, mathematics, automatics and humanoid robotics, an entire new approach to neurophysiology is emerging. It arouses a strong interest in the four communities and its development requires a joint effort and the sharing of complementary tools.

A family of extremely interesting problems concerns the understanding of the mechanisms supervising some sensorial reactions or biomechanics actions such as image reconstruction by the primary visual cortex, eyes movement and body motion.

In order to study these phenomena, a promising approach consists in identifying the motion planning problems undertaken by the brain, through the analysis of the strategies that it applies when challenged by external inputs. The role of control is that of a language allowing to read and model neurological phenomena. The control algorithms would shed new light on the brain's geometric perception (the so-called neurogeometry [86]) and on the functional organization of the motor pathways.

- A challenging problem is that of the understanding of the mechanisms which are responsible for the process of image reconstruction in the primary visual cortex V1.

The visual cortex areas composing V1 are notable for their complex spatial organization and their functional diversity. Understanding and describing their architecture requires sophisticated modeling tools. At the same time, the structure of the natural and artificial images used in visual psychophysics can be fully disclosed only using rather deep geometric concepts. The word "geometry" refers here to the internal geometry of the functional architecture of visual cortex areas (not to the geometry of the Euclidean external space). Differential geometry and analysis both play a fundamental role in the description of the structural characteristics of visual perception.

A model of human perception based on a simplified description of the visual cortex V1, involving geometric objects typical of control theory and sub-Riemannian geometry, has been first proposed by Petitot ([87]) and then modified by Citti and Sarti ([51]). The model is based on experimental observations, and in particular on the fundamental work by Hubel and Wiesel [60] who received the Nobel prize in 1981.

In this model, neurons of V1 are grouped into orientation columns, each of them being sensitive to visual stimuli arriving at a given point of the retina and oriented along a given direction. The retina is modeled by the real plane, while the directions at a given point are modeled by the projective line. The fiber bundle having as base the real plane and as fiber the projective line is called the *bundle of directions of the plane*.

From the neurological point of view, orientation columns are in turn grouped into hypercolumns, each of them sensitive to stimuli arriving at a given point, oriented along any direction. In the same hypercolumn, relative to a point of the plane, we also find neurons that are sensitive to other stimuli properties, such as colors. Therefore, in this model the visual cortex treats an image not as a planar object, but as a set of points in the bundle of directions of the plane. The reconstruction is then realized by minimizing the energy necessary to activate orientation columns among those which are not activated directly by the image. This gives rise to a sub-Riemannian problem on the bundle of directions of the plane.

- Another class of challenging problems concern the functional organization of the motor pathways.

The interest in establishing a model of the motor pathways, at the same time mathematically rigorous and biologically plausible, comes from the possible spillovers in neurophysiology. It could help to design better control strategies for robots and artificial limbs, rendering them capable to move more progressively and smoothly and also to react to exterior perturbations in a flexible way. An underlying relevant societal goal (clearly beyond our domain of expertise) is to clarify the mechanisms of certain debilitating troubles such as cerebellar disease, chorea and Parkinson's disease.

A key issue in order to establish a model of the motor pathways is to determine the criteria underlying the brain's choices. For instance, for the problem of human locomotion (see [32]), identifying such criteria would be crucial to understand the neural pathways implicated in the generation of locomotion trajectories.

A nowadays widely accepted paradigm is that, among all possible movements, the accomplished ones satisfy suitable optimality criteria (see [96] for a review). One is then led to study an inverse optimal control problem: starting from a database of experimentally recorded movements, identify a cost function such that the corresponding optimal solutions are compatible with the observed behaviors.

Different methods have been taken into account in the literature to tackle this kind of problems, for instance in the linear quadratic case [66] or for Markov processes [85]. However all these methods have been conceived for very specific systems and they are not suitable in the general case. Two approaches are possible to overcome this difficulty. The direct approach consists in choosing a cost function among a class of functions naturally adapted to the dynamics (such as energy functions) and to compare the solutions of the corresponding optimal control problem to the experimental data. In particular one needs to compute, numerically or analytically, the optimal trajectories and to choose suitable criteria (quantitative and qualitative) for the comparison with observed trajectories. The inverse approach consists in deriving the cost function from the qualitative analysis of the data.

4.3. Switched systems

Switched systems form a subclass of hybrid systems, which themselves constitute a key growth area in automation and communication technologies with a broad range of applications. Existing and emerging areas include automotive and transportation industry, energy management and factory automation. The notion of hybrid systems provides a framework adapted to the description of the heterogeneous aspects related to the interaction of continuous dynamics (physical system) and discrete/logical components.

The characterizing feature of switched systems is the collective aspect of the dynamics. A typical question is that of stability, in which one wants to determine whether a dynamical system whose evolution is influenced by a time-dependent signal is uniformly stable with respect to all signals in a fixed class ([73]).

The theory of finite-dimensional hybrid and switched systems has been the subject of intensive research in the last decade and a large number of diverse and challenging problems such as stabilizability, observability, optimal control and synchronization have been investigated (see for instance [94], [74]).

The question of stability, in particular, because of its relevance for applications, has spurred a rich literature. Important contributions concern the notion of common Lyapunov function: when there exists a Lyapunov function that decays along all possible modes of the system (that is, for every possible constant value of the signal), then the system is uniformly asymptotically stable. Conversely, if the system is stable uniformly with respect to all signals switching in an arbitrary way, then a common Lyapunov function exists [75]. In the *linear* finite-dimensional case, the existence of a common Lyapunov function is actually equivalent to the global uniform exponential stability of the system [81] and, provided that the admissible modes are finitely many, the Lyapunov function can be taken polyhedral or polynomial [39], [40], [53]. A special role in the switched control literature has been played by common quadratic Lyapunov functions, since their existence can be tested rather efficiently (see [54] and references therein). Algebraic approaches to prove the stability of switched systems under arbitrary switching, not relying on Lyapunov techniques, have been proposed in [72], [26].

Other interesting issues concerning the stability of switched systems arise when, instead of considering arbitrary switching, one restricts the class of admissible signals, by imposing, for instance, a dwell time constraint [59].

Another rich area of research concerns discrete-time switched systems, where new intriguing phenomena appear, preventing the algebraic characterization of stability even for small dimensions of the state space [69]. It is known that, in this context, stability cannot be tested on periodic signals alone [42].

Finally, let us mention that little is known about infinite-dimensional switched system, with the exception of some results on uniform asymptotic stability ([78], [91], [92]) and some recent papers on optimal control ([58], [99]).

GEOSTAT Project-Team

4. Application Domains

4.1. Application Domains

In GEOSTAT, the development of nonlinear methods for the study of complex systems and signals is conducted on four broad types of complex signals:

- Ocean dynamics and ocean/atmosphere interactions: generation of high-resolution maps from cascading properties and the determination of optimal wavelets[9], geostrophic or non-geostrophic-complex oceanic dynamics, mixing phenomena.
- Speech signal (analysis, recognition, classification)[1].
- Optimal wavelets for phase reconstruction in adaptive optics[2].
- Heartbeat signals, in cooperation with IHU LIRYC and Professor M. Haissaguerre (INSERM EA 2668 Electrophysiology and Cardiac Stimulation)[5] [35].

I4S Team

4. Application Domains

4.1. Introduction

In this section, the problems we are faced with vibration-based monitoring and within our two major application domains are briefly described.

4.2. Vibrations-based monitoring

Detecting and localizing damages for monitoring the integrity of structural and mechanical systems is a topic of growing interest, due to the aging of many engineering constructions and machines and to increased safety norms. Many current approaches still rely on visual inspections or *local* non destructive evaluations performed manually. This includes acoustic, ultrasonic, radiographic or eddy-current methods; magnet or thermal field techniques, These experimental approaches assume an *a priori* knowledge and the accessibility of a neighborhood of the damage location. Automatic *global* vibration-based monitoring techniques have been recognized to be useful alternatives to those local evaluations [33]. However this has led to actual damage monitoring systems only in the field of rotating machines.

A common feature of the structures to be monitored (e.g. civil engineering structures subject to hurricanes or earthquakes, but also swell, wind and rain; aircrafts subject to strength and turbulences, ...) is the following. These systems are subject to both fast and unmeasured variations in their environment and small slow variations in their vibrating characteristics. The available data (measurements from e.g. strain gauges or accelerometers) do not separate the effects of the external forces from the effect of the structure. The external forces vary more rapidly than the structure itself (fortunately !), damages or fatigues on the structure are of interest, while any change in the excitation is meaningless. Expert systems based on a human-like exploitation of recorded spectra can hardly work in such a case : the changes of interest (1% in eigenfrequencies) are visible neither on the signals nor on their spectra. A global health monitoring method must rather rely on a model that will help in discriminating between the two mixed causes of the changes that are contained in the measurements.

Classical modal analysis and vibration monitoring methods basically process data registered either on test beds or under specific excitation or rotation speed conditions. However there is a need for vibration monitoring algorithms devoted to the processing of data recorded in-operation, namely during the actual functioning of the considered structure or machine, without artificial excitation, speeding down or stopping.

Health monitoring techniques based on processing vibration measurements basically handle two types of characteristics: the *structural parameters* (mass, stiffness, flexibility, damping) and the *modal parameters* (modal frequencies, and associated damping values and mode-shapes); see [37] and references therein. A central question for monitoring is to compute *changes* in those characteristics and to assess their *significance*. For the *frequencies*, crucial issues are then: how to compute the changes, to assess that the changes are significant, to handle *correlations* among individual changes. A related issue is how to compare the changes in the frequencies obtained from experimental data with the sensitivity of modal parameters obtained from an analytical model. Furthermore, it has been widely acknowledged that, whereas changes in frequencies bear useful information for damage *detection*, information on changes in (the curvature of) mode-shapes is mandatory for performing damage *localization*. Then, similar issues arise for the computation and the significance of the changes. In particular, assessing the significance of (usually small) changes in the mode-shapes, and handling the (usually high) correlations among individual mode-shape changes are still considered as open questions [37], [33].

Controlling the computational complexity of the processing of the collected data is another standard monitoring requirement, which includes a limited use of an analytical model of the structure. Moreover, the reduction from the analytical model to the experimental model (truncated modal space) is known to play a key role in the success of model-based damage detection and localization.

The approach which we have been developing, based on the foundations in modules 3.2 –3.5 , aims at addressing all the issues and overcoming the limitations above.

4.3. Civil engineering

Civil engineering is a currently renewing scientific research area, which can no longer be restricted to the single mechanical domain, with numerical codes as its central focus. Recent and significant advances in physics and physical chemistry have improved the understanding of the detailed mechanisms of the constitution and the behavior of various materials (see e.g. the multi-disciplinary general agreement CNRS-Lafarge). Moreover, because of major economical and societal issues, such as durability and safety of infrastructures, buildings and networks, civil engineering is evolving towards a multi-disciplinary field, involving in particular information sciences and technologies and environmental sciences.

These last ten years, monitoring the integrity of the civil infrastructure has been an active research topic, including in connected areas such as automatic control, for mastering either the aging of the bridges, as in America (US, Canada) and Great Britain, or the resistance to seismic events and the protection of the cultural heritage, as in Italy and Greece. The research effort in France seems to be more recent, maybe because a tendency of long term design without fatigue oriented inspections, as opposite to less severe design with planned mid-term inspections. One of the current thematic priorities of the Réseau de Génie Civil et Urbain (RGCU) is devoted to constructions monitoring and diagnostics. The picture in Asia (Japan, and also China) is somewhat different, in that the demand for automatic data processing for global SHM systems is much higher, because recent or currently built bridges are equipped with hundreds if not thousands of sensors, in particular the Hong Kong-Shenzen Western Corridor and Stonecutter Bridge projects.

Among the challenges for vibration-based bridges health monitoring, two major issues are the different kinds of (non measured) excitation sources and the environmental effects [38]. Typically the traffic on *and* under the bridge, the wind and also the rain, contribute to excite the structure, and influence the measured dynamics. Moreover, the temperature is also known to affect the eigenfrequencies and mode-shapes, to an extent which is significant w.r.t. the deviations to be monitored.

4.4. Aeronautics

The aging of aerospace structures is a major current concern of civilian and military aircraft operators. Another key driving factor for SHM is to increase the operation and support efficiency of an air vehicle fleet. A SHM system is viewed as a component of a global integrated vehicle health management (IVHM) system. An overview of the users needs can be found in [30].

Improved safety and performance and reduced aircraft development and operating costs are other major concerns. One of the critical design objectives is to clear the aircraft from unstable aero-elastic vibrations (flutter) in all flight conditions. This requires a careful exploration of the dynamical behavior of the structure subject to vibration and aero-servo-elastic forces. This is achieved via a combination of ground vibration tests and in-flight tests. For both types of tests, various sensors data are recorded, and modal analyses are performed. Important challenges of the in-flight modal analyses are the limited choices for measured excitation inputs, and the presence of unmeasured natural excitation inputs (turbulence). A better exploitation of flight test data can be achieved by using output-only system identification methods, which exploits data recorded under natural excitation conditions (e.g., turbulent), without resorting to artificial control surface excitation and other types of excitation inputs [10].

A crucial issue is to ensure that the newly designed airplane is stable throughout its operating range. A critical instability phenomenon, known under the name of “*aero-elastic flutter, involves the unfavorable interaction of aerodynamic, elastic, and inertia forces on structures to produce an unstable oscillation that often results in structural failure*” [35]. For preventing from this phenomenon, the airplane is submitted to a flight flutter testing procedure, with incrementally increasing altitude and airspeed. The problem of predicting the speed at which flutter can occur is usually addressed with the aid of identification methods achieving modal analysis from the in-flight data recorded during these tests. The rationale is that the damping coefficient reflects the rate of increase or decrease in energy in the aero-servo-elastic system, and thus is a relevant measure of stability. Therefore, while frequencies and mode-shapes are usually the most important parameters in structural analysis, the most critical ones in flutter analysis are the damping factors, for some critical modes. The mode-shapes are usually not estimated for flutter testing.

Until the late nineties, most approaches to flutter clearance have led to *data-based* methods, processing different types of data. A *combined data-based and model-based* method has been introduced recently under the name of flutterometer. Based on an aero-elastic state-space model and on frequency-domain transfer functions extracted from sensor data under controlled excitation, the flutterometer computes on-line a robust flutter margin using the μ -method for analyzing the worst case effects of model uncertainty. In recent comparative evaluations using simulated and real data [32], [36], several data-based methods are shown to fail in accurately predicting flutter when using data from low speed tests, whereas the flutterometer turns out not to converge to the true flutter speed during envelope expansion, due to inherent conservative predictions.

Algorithms achieving the *on-line in-flight* exploitation of flight test data are expected to allow a more direct exploration of the flight domain, with improved confidence and reduced costs. Among other challenges, one important issue to be addressed on-line is the flight flutter monitoring problem, stated as the problem of monitoring some specific damping coefficients. On the other hand, it is known, e.g. from Cramer-Rao bounds, that damping factors are difficult to estimate accurately. For improving the estimation of damping factors, and moreover for achieving this in real-time during flight tests, one possible although unexpected route is to rely on detection algorithms able to decide whether some damping factor decreases below some critical value or not. The rationale is that detection algorithms usually have a much shorter response time than identification algorithms.

IPSO Project-Team

4. Application Domains

4.1. Laser physics

Laser physics considers the propagation over long space (or time) scales of high frequency waves. Typically, one has to deal with the propagation of a wave having a wavelength of the order of $10^{-6}m$, over distances of the order $10^{-2}m$ to 10^4m . In these situations, the propagation produces both a short-scale oscillation and exhibits a long term trend (drift, dispersion, nonlinear interaction with the medium, or so), which contains the physically important feature. For this reason, one needs to develop ways of filtering the irrelevant high-oscillations, and to build up models and/or numerical schemes that do give information on the long-term behavior. In other terms, one needs to develop high-frequency models and/or high-frequency schemes.

Generally speaking, the demand in developing such models or schemes in the context of laser physics, or laser/matter interaction, is large. It involves both modeling and numerics (description of oscillations, structure preserving algorithms to capture the long-time behaviour, etc).

In a very similar spirit, but at a different level of modelling, one would like to understand the very coupling between a laser propagating in, say, a fiber, and the atoms that build up the fiber itself.

The standard, quantum, model in this direction is called the Bloch model: it is a Schrödinger like equation that describes the evolution of the atoms, when coupled to the laser field. Here the laser field induces a potential that acts directly on the atom, and the link between this potential and the laser itself is given by the so-called dipolar matrix, a matrix made up of physical coefficients that describe the polarization of the atom under the applied field.

The scientific objective here is twofold. First, one wishes to obtain tractable asymptotic models that average out the high oscillations of the atomic system and of the laser field. A typical phenomenon here is the *resonance* between the field and the energy levels of the atomic system. Second, one wishes to obtain good numerical schemes in order to solve the Bloch equation, beyond the oscillatory phenomena entailed by this model.

4.2. Molecular Dynamics

In classical molecular dynamics, the equations describe the evolution of atoms or molecules under the action of forces deriving from several interaction potentials. These potentials may be short-range or long-range and are treated differently in most molecular simulation codes. In fact, long-range potentials are computed at only a fraction of the number of steps. By doing so, one replaces the vector field by an approximate one and alternates steps with the exact field and steps with the approximate one. Although such methods have been known and used with success for years, very little is known on how the "space" approximation (of the vector field) and the time discretization should be combined in order to optimize the convergence. Also, the fraction of steps where the exact field is used for the computation is mainly determined by heuristic reasons and a more precise analysis seems necessary. Finally, let us mention that similar questions arise when dealing with constrained differential equations, which are a by-product of many simplified models in molecular dynamics (this is the case for instance if one replaces the highly-oscillatory components by constraints).

4.3. Plasma physics

The development of efficient numerical methods is essential for the simulation of plasmas and beams at the kinetic level of description (Vlasov type equations). It is well known that plasmas or beams give rise to small scales (Debye length, Larmor radius, gyroperiod, mean free path...) which make numerical simulations challenging. Instead of solving the limit or averaged models by considering these small scales equal to zero, our aim is to explore a different strategy, which consists in using the original kinetic equation. Specific numerical scheme called "Asymptotic Preserving" scheme is then built to discretize the original kinetic

equation. Such a scheme allows to pass to the limit with no stability problems, and provide in the limit a consistent approximation of the limit or average model. A systematic and robust way to design such a scheme is the micro-macro decomposition in which the solution of the original model is decomposed into an averaged part and a remainder.

MATHRISK Team

3. Application Domains

3.1. Application Domains

- Utility maximization in incomplete markets
- option pricing
- quantitative risk management
- systemic risk
- limit order books
- credit risk
- liquidity risk
- computational finance
- calibration

MAXPLUS Project-Team

4. Application Domains

4.1. Systèmes à événements discrets (productique, réseaux)/Discrete event systems (manufacturing systems, networks)

Une partie importante des applications de l'algèbre max-plus provient des systèmes dynamiques à événements discrets [6]. Les systèmes linéaires max-plus, et plus généralement les systèmes dynamiques monotones contractants, fournissent des modèles naturels dont les résultats analytiques peuvent être appliqués aux problèmes d'évaluation de performance. Relèvent de l'approche max-plus, tout au moins sous forme simplifiée : des problèmes de calcul de temps de cycle pour des circuits digitaux [80], des problèmes de calcul de débit pour des ateliers [126], pour des réseaux ferroviaires [79] ou routiers, et l'évaluation de performance des réseaux de communication [71]. L'approche max-plus a été appliquée à l'analyse du comportement temporel de systèmes concurrents, et en particulier à l'analyse de "high level sequence message charts" [75], [134]. Le projet Maxplus collabore avec le projet Metalau, qui étudie particulièrement les applications des modèles max-plus à la modélisation microscopique du trafic routier [142], [139], [104].

English version

One important part of applications of max-plus algebra comes from discrete event dynamical systems [6]. Max-plus linear systems, and more generally, monotone nonexpansive dynamical systems, provide natural models for which many analytical results can be applied to performance evaluation problems. For instance, problems like computing the cycle time of asynchronous digital circuits [80], or computing the throughput of a workshop [126] or of a transportation network, and performance evaluation problems for communication networks, are often amenable to max-plus algebra, at least in some simplified form, see in particular [79] and [71]. The max-plus approach has been applied to the analysis of the time behaviour of concurrent systems, and in particular, to the analysis of high level sequence message charts [75], [134]. The Maxplus team collaborates with the Metalau team, working particularly on the applications of max-plus models to the microscopic modelling of road traffic [142], [139], [104].

4.2. Commande optimale et jeux/Optimal control and games

La commande optimale et la théorie des jeux ont de nombreuses applications bien répertoriées: économie, finance, gestion de stock, optimisation des réseaux, aide à la décision, etc. En particulier, le projet Mathfi travaille sur les applications à des problèmes de mathématiques financières. Il existe une tradition de collaborations entre les chercheurs des projets Mathfi et Maxplus sur ces questions, voir par exemple [5] qui comprend un résultat exploitant des idées de théorie spectrale non-linéaire, présentées dans [3].

English version

Optimal control and game theory have numerous well established applications fields: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In particular, the Mathfi team works on applications in mathematical finance. There is a tradition of collaboration between researchers of the Maxplus team and of the Mathfi team on these questions, see as an illustration [5] where ideas from the spectral theory of monotone homogeneous maps [3] are applied.

4.3. Recherche opérationnelle/Operations research

L'algèbre max-plus intervient de plusieurs manières en Recherche opérationnelle. Premièrement, il existe des liens profonds entre l'algèbre max-plus et les problèmes d'optimisation discrète, voir [81]. Ces liens conduisent parfois à de nouveaux algorithmes pour les problèmes de recherche opérationnelle classiques,

comme le problème de circuit de poids moyen maximum [88]. Certains problèmes combinatoires, comme des problèmes de programmation disjonctive, peuvent être décomposés par des méthodes de type max-plus [177]. Ensuite, le rôle de l’algèbre max-plus dans les problèmes d’ordonnement est bien connu depuis les années 60, les dates de complétion pouvant souvent être calculées à partir d’équations linéaires max-plus. Plus récemment, des représentations de problèmes d’ordonnement ont pu être obtenues à partir de semi-groupes de matrices max-plus : une première représentation a été obtenue dans [113] pour le cas du “jobshop”, une représentation plus simple a été obtenue dans [136] dans le cas du “flowshop”. Ce point de vue algébrique a été très utile dans le cas du “flowshop” : il permet de retrouver des résultats anciens de dominance et d’obtenir ainsi de nouvelles bornes [136]. Finalement, en regardant l’algèbre max-plus comme une limite de l’algèbre classique, on peut utiliser des outils algébriques en optimisation combinatoire [132].

English version

Max-plus algebra arise in several ways in Operations Research. First, there are intimate relations between max-plus algebra and discrete optimisation problems, see [81]. Sometimes, these relations lead to new algorithms for classical Operations Research problems, like the maximal circuit mean [88]. There are also special combinatorial problems, like certain problems of disjunctive programming, which can be decomposed by max-plus type methods [177]. Next, the role of max-plus algebra in scheduling problems has been known since the sixties: completion dates can often be computed by max-plus linear equations. Recently, representations of certain scheduling problems using max-plus matrix semigroups have appeared, a first representation was given in [113] for the jobshop case, a simpler representation was given in [136] in the flowshop case. This algebraic point of view turned out to be particularly fruitful in the flowshop case: it allows one to recover old dominance results and to obtain new bounds [136]. Finally, viewing max-plus algebra as a limit of classical algebra allows to use algebraic tools in combinatorial optimisation [132].

4.4. Analyse statique de programmes/Static analysis of computer programs

L’interprétation abstraite est une technique, introduite par P. et R. Cousot [92], qui permet de déterminer des invariants de programmes en calculant des points fixes minimaux d’applications monotones définies sur certains treillis. On associe en effet à chaque point de contrôle du programme un élément du treillis, qui représente une sur-approximation valide de l’ensemble des valeurs pouvant être prises par les variables du programme en ce point. Le treillis le plus simple exprimant des propriétés numériques est celui des produits Cartésiens d’intervalles. Des treillis plus riches permettent de mieux tenir compte de relations entre variables, en particulier, des classes particulières de polyèdres sont souvent employées.

Voici, en guise d’illustration, un petit exemple de programme, avec le système de point fixe associé, pour le treillis des intervalles:

<code>void main() {</code>	<code>int x=0;</code>	<code>// 1</code>	$x_1 =$	$[0, 0]$
<code>while (x<100) {</code>	<code> x=x+1;</code>	<code>// 3</code>	$x_2 =$	$] - \infty, 99] \cap (x_1 \cup x_3)$
<code> }</code>	<code> }</code>	<code>// 4</code>	$x_3 =$	$x_2 + [1, 1]$
<code> }</code>	<code>}</code>		$x_4 =$	$[100, +\infty[\cap (x_1 \cup x_3)$

Si l’on s’intéresse par exemple aux valeurs maximales prise par la variable x au point de contrôle 2, soit $x_2^+ := \max x_2$, après une élimination, on parvient au problème de point fixe:

$$x_2^+ = \min(99, \max(0, x_2^+ + 1)) , \quad (2)$$

qui a pour plus petite solution $x_2^+ = 99$, ce qui prouve que x est majoré par 99 au point 2.

On reconnaît ici un opérateur de point fixe associé à un problème de jeux à deux joueurs et somme nulle. Cette analogie est en fait générale, dans le cadre d’une collaboration que l’équipe entretient depuis plusieurs années avec l’équipe MeASI d’Eric Goubault (CEA et LIX), spécialiste d’analyse statique, nous avons en effet mis progressivement en évidence une correspondance [91], [110], entre les problèmes de jeux à somme nulle et les problèmes d’analyse statique, qui peut se résumer par le dictionnaire suivant:

Jeux		Interprétation abstraite
système dynamique		programme
opérateur de Shapley		fonctionnelle
espace d’état	(# points de contrôle) × (# degrés de liberté du treillis)	exécution de n pas
problème en horizon n		invariant optimal (borne)
limite du problème en horizon fini		itération de Kleene
itération sur les valeurs		

Pour que le nombre d’états du jeu soit fini, il est nécessaire de se limiter à des treillis d’ensembles ayant un nombre fini de degrés de liberté, ce qui est le cas de domaines communément utilisés (intervalles, ensembles définis par des contraintes de potentiel de type $x_i - x_j \leq c$, mais aussi, les “templates” qui sont des sous-classes de polyèdres introduits récemment par Sankaranarayanan, Sipma et Manna [166]). L’ensemble des actions est alors fini si on se limite à une arithmétique affine. Signalons cependant qu’en toute généralité, on aboutit à des jeux avec un taux d’escompte négatif, ce qui pose des difficultés inédites. Cette correspondance entre jeux et analyse statique est non intuitive, au sens où les actions du minimiseur consistent à sélectionner des points extrêmes de certains polyèdres obtenus par un mécanisme de dualité.

Une pathologie bien répertoriée en analyse statique est la lenteur des algorithmes de point fixe, qui peuvent effectuer un nombre d’itérations considérable (99 itérations pour obtenir le plus petit point fixe de (8)). Celle-ci est usuellement traitée par des méthodes d’accélération de convergence dites d’élargissement et rétrécissement [93], qui ont cependant l’inconvénient de conduire à une perte de précision des invariants obtenus. Nous avons exploité la correspondance entre analyse statique et jeux pour développer des algorithmes d’une nature très différente, s’inspirant de nos travaux antérieurs sur l’itération sur les politiques pour les jeux répétés [111], [86], [87],[7]. Une version assez générale de cet algorithme, adaptée au domaine des templates, est décrite dans [110] et a fait l’objet d’une implémentation prototype. Chaque itération combine de la programmation linéaire et des algorithmes de graphes. Des résultats expérimentaux ont montré le caractère effectif de la méthode, avec souvent un gain en précision par rapport aux approches classiques, par exemple pour des programmes comprenant des boucles imbriquées.

Ce domaine se trouve être en pleine évolution, un enjeu actuel étant de traiter d’une manière qui passe à l’échelle des invariants plus précis, y compris dans des situations où l’arithmétique n’est plus affine.

English version

The abstract interpretation method introduced by P. and R. Cousot [92], allows one to determine automatically invariants of programs by computing the minimal fixed point of an order preserving map defined on a complete lattice. To every breakpoint of the program is associated an element of the lattice, which yields a valid overapproximation of the set of reachable values of the vectors of variables of the program, at this breakpoint. The simplest lattice expressing numerical invariants consists of Cartesian products of intervals. More sophisticated lattices, taking into account relations between variables, consisting in particular of subclasses of polyhedra, are often used.

As an illustration, we gave before Eqn (8) a simple example of program, together with the associated fixed-point equation. In this example, the value of the variable x at the breakpoint 2 is bounded by the smallest solution x_2^+ of the fixed point problem (8), which is equal to 99.

The fixed point equation (8) is similar to the one arising in the theory of zero-sum repeated games. This analogy turns out to be general. Un a series of joint works of our team with the MeASI team of Eric Goubault (CEA and LIX), we brought progressively to light a correspondence [91], [110], between the zero-sum game problems and the static analysis problems, which can be summarized by the following dictionary:

Games	Abstract interpretation
dynamical system	program
Shapley operator	functional
state space	(# breakpoints) \times (# degrees of freedom)
horizon n problem	execution of n logical steps
limit of the value in horizon n	optimal invariant (bound)
value iteration	Kleene iteration

For the game to have a finite state space, we must restrict our attention to lattices of sets with a finite number of degrees of freedom, which is the case of the domains commonly used in static analysis (intervals, sets defined by potentials constraints of the form $x_i - x_j \leq \text{cst}$, and also the subclasses of polyhedra called “templates”, introduced recently by Sankaranarayanan, Sipma and Manna [166]). Then, the action space is finite if the arithmetics of the program is affine. However, in full generality, the games we end up with have a negative discount rate, which raises difficulties which are unfamiliar from the game theory point of view. This correspondence between games and static analysis turns out to be non intuitive, in that the action of the minimizer consist of selecting an extreme point of a polyhedron arising from a certain duality construction.

A well known pathology in static analysis is the fact that the standard Kleene fixed point algorithm may have a very slow behavior (99 iterations are needed to get the smallest fixed point of (8)). This is usually solved by using some accelerations of convergence, called widening and narrowing [93], which however lead to a loss of precision. We exploited the correspondence between static analysis and games to develop algorithms of a very different nature, inspired by our earlier work on policy iteration for games [111], [86], [87],[7]. A rather general version of this policy iteration algorithm, adapted to the domain of templates, is described in [110], together with a prototype implementation. Every iteration combines linear programming and combinatorial algorithms. Some experimental results indicate that the method often leads to invariants which are more accurate than the ones obtained by alternative methods, in particular for some programs with nested loops.

This topic of research is currently evolving, a question of current interest being to find accurate invariants, in a scalable way, in situations in which the arithmetics is not affine.

4.5. Autres applications/Other applications

L’algèbre max-plus apparaît de manière naturelle dans le calcul de scores de similitudes dans la comparaison de séquences génétiques. Voir par exemple [90].

English version

Max-plus algebra arises naturally in the computation of similarity scores, in biological sequence comparison. See for instance [90].

MC2 Project-Team

4. Application Domains

4.1. Introduction

We now present our contribution to these above challenges concerning interface problem for complex fluids, direct simulations and analysis, flow control and optimization. From the technical point of view, many productions are common to the different parts of the project. For example, level-set methods, fast-marching procedure are used for shape optimization and for microfluidics, penalization methods are used for high Reynolds flows and for tumor growth. This leads to a strong politic of development of numerical modules.

4.2. Multi-fluid flows

- computation of bifluid flows : see the thesis of S. Tancogne ([83]) and P. Vigneaux ([86]). Stability of an interface, shape of droplets, formation of a jet. Study of the Plateau-Rayleigh instability. Behaviour of diphasic fluids evolving in square microchannels.
- mixing in micro-channel : see the thesis of J. Dambrine [67]. Passive mixing strategies involving boundary conditions. Enhanced oil recovery (study of mixing oil-water-polymer in a microchannel).
- emulsions and foam : see the thesis of S. Benito [52]. Applications in biology : behaviour of tissues, of tumor,....
- polymer nanotube conglomerate wire : it was the subject of a talk in the following conference "WCCM8-ECCOMAS2008" and of the talk [65].

4.3. Cancer modeling

- specific models : investigation of two particular cancer : gliomas (brain tumors), colorectal cancers lung and lever metastasis, brest cancer. This is one part of the PhD works of J.B. Lagaert and D. Lombardi.
- modelling of electrochemotherapy : see ARC C3MB (<http://www.math.u-bordeaux1.fr/ArcC3MB/>)
- parameter estimations with the help of low order models : see the PhD of J.B. Lagaert and D. Lombardi
- patient-specific simulations
- optimal shape design : the goal is to recover the vascularization of a model tumor from the knowledge of its shape evolution.

4.4. Newtonian fluid flows simulations and their analysis

- simulation of a synthetic or pulsed jet. This is an ongoing project with Renault and PSA inside a PREDIT project.
- vortex dynamics : see [69].
- simulation of compressible flows on cartesian grids : see the thesis of Gabriele Ottino's Thesis [79], who underwent his doctoral studies in conjunction in the MC2 team and at the Politecnico di Torino, and defended in April 2009. He had a grant of the French-Italian university.
- 3D turbulent flows through DESGRIVRE contract with AIRBUS. Thesis of C. Wervaecke [88]. The goal is to use Detached-Eddy Simulation to model turbulent flows around iced bodies.
- porous media : Numerical study of coupling between Richards and transport-diffusion equations in permeable sediment affected by tidal oscillation. See the thesis of R. Chassagne [63]

4.5. Flow control and shape optimization

- passive control : the idea is to put a porous interface between the solid body and the fluid. See the D. Depeyras thesis [68] and Yong-Liang Xiang [90] and CH Bruneau and Iraj Mortazavi) [55]. See also project [60] founded by the European Community.
- active control : see the three PhD thesis: M. Buffoni, J. Weller [87], E. Lombardi and FFAST project funded by EU and led by the University of Bristol and AIRBUS UK.
- shape optimization for turbo-machines : See [84].
- reduced order models : it consists in designing a non-linear observer that estimates the state of the flow field from a limited number of measurements in the field. The challenge is to reduce as much as possible the information required and to take it from the boundary. See J. Weller [87] and E. Lombardi.
- passive control of flows with porous media : see [57], [54], [53], [78], [58].
- inverse problems in imagery : see [62].

MCTAO Team

4. Application Domains

4.1. Space engineering, satellites, low thrust control

Space engineering is very demanding in terms of safe and high-performance (for instance in terms of fuel consumption, because only a finite amount of fuel is onboard a satellite for all its “life”) control laws. It is therefore prone to real industrial collaborations.

We are especially interested in trajectory control of space vehicles using their own propulsion devices, outside the atmosphere. Here we discuss “non-local” control problems (in the sense of section 3.1 point 1): in the geocentric case, orbit transfer rather than station keeping; also we do not discuss attitude control.

A space vehicle is subject to

- gravitational forces, from one or more central bodies (the corresponding acceleration is denoted by F_{grav} below),
- other forces of small amplitude (the corresponding acceleration is denoted by F_2 below),
- a thrust, the control, produced by a propelling device; it is the $G u$ term below; assume for simplicity that control in all directions is allowed, *i.e.* G is an invertible matrix.

In position-velocity coordinates, its dynamics can be written as

$$\ddot{x} = F_{\text{grav.}}(x, t) \left[+ F_2(x, \dot{x}, t) \right] + G(x, \dot{x}) u, \quad \|u\| \leq u_{\text{max}}. \quad (3)$$

The second term is often neglected in the design of the control. Time-dependence reflects the movement of attracting celestial bodies if there is more than one (see below).

4.1.1. Low thrust

means that u_{max} is small, or more precisely that the maximum magnitude of $G u$ is small with respect to the one of $F_{\text{grav.}}$ (but is usually large compared to F_2). Hence the influence of the control is very weak instantaneously, and trajectories can only be significantly modified by accumulating the effect of this low thrust on a long time. Obviously this is possible only because the free system is somehow conservative. This was “abstracted” in section 3.5.

Why low thrust ? The common principle to all propulsion devices is to eject particles, with some relative speed with respect to the vehicle; conservation of momentum then induces, from the point of view of the vehicle alone, an external force, the “thrust” (and a mass decrease). Ejecting the same mass of particles with a higher relative speed results in a proportionally higher thrust; this relative speed (specific impulse, I_{sp}) is a characteristic of the engine; the higher the I_{sp} , the smaller the mass of particles needed for the same change in the vehicle momentum. Engines with a higher I_{sp} are highly desirable because, for the same maneuvers, they reduce the mass of “fuel” to be taken on-board the satellite, hence leaving more room (mass) for the payload. “Classical” chemical engines use combustion to eject particles, at a somehow limited speed even with very efficient fuel; the more recent electric engines use a magnetic field to accelerate particles and eject them at a considerably higher speed; however electrical power is limited (solar cells), and only a small amount of particles can be accelerated per unit of time, inducing the limitation on thrust magnitude.

Electric engines theoretically allow many more maneuvers with the same amount of particles, with the drawback that the instant force is very small; sophisticated control design is necessary to circumvent this drawback. High thrust engines allow simpler control procedures because they almost allow instant maneuvers (strategies consist in a few burns at precise instants).

4.1.2. Typical problems

Let us mention two.

- *Orbit transfer or rendez-vous.* It is the classical problem of bringing a satellite to its operating position from the orbit where it is delivered by the launcher; for instance from a GTO orbit to the geostationary orbit at a prescribed longitude (one says rendez-vous when the longitude, or the position on the orbit, is prescribed, and transfer if it is free). In equation (1) for the dynamics, F_{grav} is the Newtonian gravitation force of the earth (it then does not depend on time); F_2 contains all the terms coming either from the perturbations to the Newtonian potential or from external forces like radiation pressure, and the control is usually allowed in all directions, or with some restrictions to be made precise.
- *Three body problem.* This is about missions in the solar system leaving the region where the attraction of the earth, or another single body, is preponderant. We are then no longer in the situation of a single central body, F_{grav} contains the attraction of different planets and the sun. In regions where two central bodies have an influence, say the earth and the moon, or the sun and a planet, the term F_{grav} in (1) is the one of the restricted three body problem and dependence on time reflects the movement of the two “big” attracting bodies.

In the 2003 mission SMART-1, the project was to send a small observation vehicle from the Earth to the Moon using low-thrust propulsion. The vehicle was launched and reached the moon: the goal was achieved; precise reports on the control used can be found in [73]. There was no attempt to minimize fuel consumption, or transfer time, and it is not a surprise that the implemented solution is far from optimal with respect to these criteria. In a recent work in the Dijon team, and in collaboration with J. Gergaud from APO team at IRIT- ENSEEIHT (Toulouse) we have computed optimal trajectories for the Earth-Moon transfer according to the energy minimization problem, the time minimal transfer or the propellant minimization consumption. These results combine geometric optimal control and numerical simulations with adapted numerical codes. The contributions are described in G. Picot Phd thesis, (Dijon November 2010), B. Daoud (Phd thesis defended at Dijon in October 2011) and the numerical codes are developed by O. Cots (Phd thesis to be defended at Dijon in June 2012). Our previous work [29] gives a *feedback* solution for this problem, divided in three phases, the design being based on a two-body model in the first and last phase, where the effect of the primaries is preponderant and the second phase is in the neighborhood of the L2 Lagrange point. This opens perspectives in trajectory optimization; see the recent work [32]. For a state of the art, the reader may refer to [62] or [28].

An issue for future experimental missions in the solar system is interplanetary flight planning with gravitational assistance. Tackling this global problem, that even contains some combinatorial problems (itinerary), goes beyond the methodology developed here, but the above considerations are a brick in this puzzle.

4.1.3. Properties of the control system.

If there are no restrictions on the thrust direction, i.e., in equation (1), if the control u has dimension 3 with an invertible matrix G , then the control system is “static feedback linearizable”, and a fortiori flat, see section 3.2. However, implementing the static feedback transformation would consist in using the control to “cancel” the gravitation; this is obviously impossible since the available thrust is very small. As mentioned in section 3.1, point 3, the problem remains fully nonlinear in spite of this “linearizable” structure².

4.1.4. Context for these applications

The geographic proximity of Thales Alenia Space, in conjunction with the “Pole de compétitivité” PEGASE in PACA region is an asset for a long term collaboration between Inria - Sophia Antipolis and Thales Alenia

²However, the linear approximation around *any* feasible trajectory is controllable (a periodic time-varying linear system); optimal control problems will have no singular or abnormal trajectories.

Space (Thales Alenia Space site located in Cannes hosts one of the very few European facilities for assembly, integration and tests of satellites).

B. Bonnard and J.-B. Caillau in Dijon have had a strong activity in optimal control for space, in collaboration with the APO Team from IRIT at ENSEEIHT (Toulouse), and sometimes with EADS, for development of geometric methods in numerical algorithms.

4.2. Quantum Control

The activity of B. Bonnard in quantum control started as a collaboration with D. Sugny (a physicist from ICB) in the ANR project Comoc, localized mainly at the University of Dijon; the problem was the control of the orientation of a molecule using a laser field, with a model that does take into account the dissipation due to the interaction with the environment, molecular collisions for instance. The model is a dissipative generalization of the finite dimensional Schrödinger equation, known as Lindblad equation. In particular we have computed the minimum time control and the minimum energy control for the orientation of a two-level system, using geometric optimal control and adapted numerical methods (shooting and numeric continuation) [36], [35]. The model is a 3-dimensional system depending upon 3 parameters, yielding a very complicated optimal control problem that we have solved for prescribed boundary conditions.

More recently, based on this project, we have reoriented our control activity towards Nuclear Magnetic Resonance (MNR). In MNR medical imaging, the contrast problem is the one of designing a variation of the magnetic field with respect to time that maximizes the difference, on the resulting image, between two different chemical species; this research is conducted with Prof. S. Glaser (TU-München); it was evidenced experimentally that the current contrast of the image is significantly improved by using “our” exact optimal control methods. The model is the Bloch equation for spin $\frac{1}{2}$ particles, that can be interpreted as a sub-case of Lindblad equation for a two-level system; the control problem to solve amounts to driving in minimum time the magnetization vector of the spin to zero (for parameters of the system corresponding to one of the species), and generalizations where such spin $\frac{1}{2}$ particles are coupled: double spin inversion for instance. This research project is supported on the french side by a [PEPS INSIS \(Control-Image\)](#).

4.3. Applications of optimal transport

Optimal Transportation in general has many applications. Image processing, biology, fluid mechanics, mathematical physics, game theory, traffic planning, financial mathematics, economics are among the most popular fields of application of the general theory of optimal transport. Many developments have been made in all these fields recently. Two more specific fields:

- In image processing, since a grey-scale image may be viewed as a measure, optimal transportation has been used because it gives a distance between measures corresponding to the optimal cost of moving densities from one to the other, see e.g. the work of J.-M. Morel and co-workers [66].
- In representation and approximation of geometric shapes, say by point-cloud sampling, it is also interesting to associate a measure, rather than just a geometric locus, to a distribution of points (this gives a small importance to exceptional “outlier” mistaken points); this was developed in Q. Mérigot’s PhD [69] in the GEOMETRICA project-team. The relevant distance between measures is again the one coming from optimal transportation. Such approach, combined with evolution of densities mentioned above, may help to perform robust stabilization of set of particles.

McTAO is not directly involved in applications of transport. A starting point may be Alice Erlinger’s PhD, in co-supervision with Ludovic Rifford and Robert McCann from the University of Toronto, that will comprehend applications of optimal transportation to the modeling of markets in economy; it is starting in september, 2012.

Applications that would be *specific to the type of costs that we consider*, i.e. these coming from optimal control are concerned with evolutions of densities under state or velocity constraints. A fluid motion or a crowd movement can be seen as the evolution of a density in a given space. If constraints are given on the directions in which these densities can evolve, we are in the framework of non-holonomic transport problems. Such approach could be useful to stabilize a large set of particles.

4.4. Applications to some domains of mathematics.

Control theory (in particular thinking in terms of inputs and reachable set) has brought novel ideas and progresses to mathematics. For instance, some problems from classical calculus of variations have been revisited in terms of optimal control and Pontryagin's Maximum Principle [55]; also, closed geodesics for perturbed Riemannian metrics were constructed in [58], [59] with control techniques.

The work in progress [45] is definitely in this line, applying techniques from control to construct some perturbations under constraints of Hamiltonian systems to solve longstanding open questions in dynamical systems. Other work by Rifford and Ruggiero [13] applied successfully geometric control techniques to obtain genericity properties for Hamiltonian systems.

MICMAC Project-Team

4. Application Domains

4.1. Electronic structure of large systems

As the size of the systems one wants to study increases, more efficient numerical techniques need to be resorted to. In computational chemistry, the typical scaling law for the complexity of computations with respect to the size of the system under study is N^3 , N being for instance the number of electrons. The Holy Grail in this respect is to reach a linear scaling, so as to make possible simulations of systems of practical interest in biology or material science. Efforts in this direction must address a large variety of questions such as

- how can one improve the nonlinear iterations that are the basis of any *ab initio* models for computational chemistry?
- how can one more efficiently solve the inner loop which most often consists in the solution procedure for the linear problem (with frozen nonlinearity)?
- how can one design a sufficiently small variational space, whose dimension is kept limited while the size of the system increases?

An alternative strategy to reduce the complexity of *ab initio* computations is to try to couple different models at different scales. Such a mixed strategy can be either a sequential one or a parallel one, in the sense that

- in the former, the results of the model at the lower scale are simply used to evaluate some parameters that are inserted in the model for the larger scale: one example is the parameterized classical molecular dynamics, which makes use of force fields that are fitted to calculations at the quantum level;
- while in the latter, the model at the lower scale is concurrently coupled to the model at the larger scale: an instance of such a strategy is the so called QM/MM coupling (standing for Quantum Mechanics/Molecular Mechanics coupling) where some part of the system (typically the reactive site of a protein) is modeled with quantum models, that therefore accounts for the change in the electronic structure and for the modification of chemical bonds, while the rest of the system (typically the inert part of a protein) is coarse grained and more crudely modeled by classical mechanics.

The coupling of different scales can even go up to the macroscopic scale, with methods that couple a microscopic description of matter, or at least a mesoscopic one, with the equations of continuum mechanics at the macroscopic level.

4.2. Computational Statistical Mechanics

The orders of magnitude used in the microscopic description of matter are far from the orders of magnitude of the macroscopic quantities we are used to: The number of particles under consideration in a macroscopic sample of material is of the order of the Avogadro number $N_A \sim 10^{23}$, the typical distances are expressed in Å (10^{-10} m), the energies are of the order of $k_B T \simeq 4 \times 10^{-21}$ J at room temperature, and the typical times are of the order of 10^{-15} s when the proton mass is the reference mass.

To give some insight into such a large number of particles contained in a macroscopic sample, it is helpful to compute the number of moles of water on earth. Recall that one mole of water corresponds to 18 mL, so that a standard glass of water contains roughly 10 moles, and a typical bathtub contains 10^5 mol. On the other hand, there are approximately 1.3×10^{18} m³ of water in the oceans, *i.e.* 7.2×10^{22} mol, a number comparable to the Avogadro number. This means that inferring the macroscopic behavior of physical systems described at the microscopic level by the dynamics of several millions of particles only is like inferring the ocean's dynamics from hydrodynamics in a bathtub...

For practical numerical computations of matter at the microscopic level, following the dynamics of every atom would require simulating N_A atoms and performing $O(10^{15})$ time integration steps, which is of course impossible! These numbers should be compared with the current orders of magnitude of the problems that can be tackled with classical molecular simulation, where several millions of atoms only can be followed over time scales of the order of $0.1 \mu\text{s}$.

Describing the macroscopic behavior of matter knowing its microscopic description therefore seems out of reach. Statistical physics allows us to bridge the gap between microscopic and macroscopic descriptions of matter, at least on a conceptual level. The question is whether the estimated quantities for a system of N particles correctly approximate the macroscopic property, formally obtained in the thermodynamic limit $N \rightarrow +\infty$ (the density being kept fixed). In some cases, in particular for simple homogeneous systems, the macroscopic behavior is well approximated from small-scale simulations. However, the convergence of the estimated quantities as a function of the number of particles involved in the simulation should be checked in all cases.

Despite its intrinsic limitations on spatial and timescales, molecular simulation has been used and developed over the past 50 years, and its number of users keeps increasing. As we understand it, it has two major aims nowadays.

First, it can be used as a *numerical microscope*, which allows us to perform “computer” experiments. This was the initial motivation for simulations at the microscopic level: physical theories were tested on computers. This use of molecular simulation is particularly clear in its historic development, which was triggered and sustained by the physics of simple liquids. Indeed, there was no good analytical theory for these systems, and the observation of computer trajectories was very helpful to guide the physicists’ intuition about what was happening in the system, for instance the mechanisms leading to molecular diffusion. In particular, the pioneering works on Monte-Carlo methods by Metropolis et al, and the first molecular dynamics simulation of Alder and Wainwright were performed because of such motivations. Today, understanding the behavior of matter at the microscopic level can still be difficult from an experimental viewpoint (because of the high resolution required, both in time and in space), or because we simply do not know what to look for! Numerical simulations are then a valuable tool to test some ideas or obtain some data to process and analyze in order to help assessing experimental setups. This is particularly true for current nanoscale systems.

Another major aim of molecular simulation, maybe even more important than the previous one, is to compute macroscopic quantities or thermodynamic properties, typically through averages of some functionals of the system. In this case, molecular simulation is a way to obtain *quantitative* information on a system, instead of resorting to approximate theories, constructed for simplified models, and giving only qualitative answers. Sometimes, these properties are accessible through experiments, but in some cases only numerical computations are possible since experiments may be unfeasible or too costly (for instance, when high pressure or large temperature regimes are considered, or when studying materials not yet synthesized). More generally, molecular simulation is a tool to explore the links between the microscopic and macroscopic properties of a material, allowing one to address modelling questions such as “Which microscopic ingredients are necessary (and which are not) to observe a given macroscopic behavior?”

4.3. Homogenization and related problems

Over the years, the project-team has developed an increasing expertise on how to couple models written at the atomistic scale, with more macroscopic models, and, more generally, an expertise in multiscale modelling for materials science.

The following observation motivates the idea of coupling atomistic and continuum description of materials. In many situations of interest (crack propagation, presence of defects in the atomistic lattice, ...), using a model based on continuum mechanics is difficult. Indeed, such a model is based on a macroscopic constitutive law, the derivation of which requires a deep qualitative and quantitative understanding of the physical and mechanical properties of the solid under consideration. For many solids, reaching such an understanding is a challenge, as loads they are submitted to become larger and more diverse, and as experimental observations

helping designing such models are not always possible (think of materials used in the nuclear industry). Using an atomistic model in the whole domain is not possible either, due to its prohibitive computational cost. Recall indeed that a macroscopic sample of matter contains a number of atoms on the order of 10^{23} . However, it turns out that, in many situations of interest, the deformation that we are after is not smooth in *only a small part* of the solid. So, a natural idea is to try to take advantage of both models, the continuum mechanics one and the atomistic one, and to couple them, in a domain decomposition spirit. In most of the domain, the deformation is expected to be smooth, and reliable continuum mechanics models are then available. In the rest of the domain, the expected deformation is singular, one needs an atomistic model to describe it properly, the cost of which remains however limited as this region is small.

From a mathematical viewpoint, the question is to couple a discrete model with a model described by PDEs. This raises many questions, both from the theoretical and numerical viewpoints:

- first, one needs to derive, from an atomistic model, continuum mechanics models, under some regularity assumptions that encode the fact that the situation is smooth enough for such a macroscopic model to be a good description of the materials;
- second, couple these two models, e.g. in a domain decomposition spirit, with the specificity that models in both domains are written in a different language, that there is no natural way to write boundary conditions coupling these two models, and that one would like the decomposition to be self-adaptive.

More generally, the presence of numerous length-scales in material science problems represents a challenge for numerical simulation, especially when some *randomness* is assumed on the materials. It can take various forms, and includes defects in crystals, thermal fluctuations, and impurities or heterogeneities in continuous media. Standard methods available in the literature to handle such problems often lead to very costly computations. Our goal is to develop numerical methods that are more affordable. Because we cannot embrace all difficulties at once, we focus on a simple case, where the fine scale and the coarse-scale models can be written similarly, in the form of a simple elliptic partial differential equation in divergence form. The fine scale model includes heterogeneities at a small scale, a situation which is formalized by the fact that the coefficients in the fine scale model vary on a small length scale. After homogenization, this model yields an effective, macroscopic model, which includes no small scale. In many cases, a sound theoretical groundwork exists for such homogenization results. We consider mostly the setting of stochastic homogenization of linear, scalar, second order elliptic PDEs, where analytical formulas for the effective properties are known. The difficulty stems from the fact that they generally lead to prohibitively costly computations. For such a case, simple from the theoretical viewpoint, our aim is to focus on different practical computational approaches to speed-up the computations. One possibility, among others, is to look for specific random materials, relevant from the practical viewpoint, and for which a dedicated approach can be proposed, that is less expensive than the general approach.

MISTIS Project-Team

4. Application Domains

4.1. Image Analysis

Participants: Christine Bakhous, Lotfi Chaari, Senan James Doyle, Thomas Vincent, Florence Forbes, Ludovic Leau-Mercier, Huu Giao Nguyen, Stéphane Girard, Kai Qin, Darren Wraith.

As regards applications, several areas of image analysis can be covered using the tools developed in the team. More specifically, in collaboration with team Perception, we address various issues in computer vision involving Bayesian modelling and probabilistic clustering techniques. Other applications in medical imaging are natural. We work more specifically on MRI data, in collaboration with the Grenoble Institute of Neuroscience (GIN) and LNAO from the NeuroSpin center of CEA Saclay. We also consider other statistical 2D fields coming from other domains such as remote sensing, in collaboration with Laboratoire de Planétologie de Grenoble. In the context of the ANR MDCO project Vahine, we work on hyperspectral multi-angle images. In the context of the "pole de competitivite" project I-VP, we work of images of PC Boards.

4.2. Biology, Environment and Medicine

Participants: Thomas Vincent, Christine Bakhous, Lotfi Chaari, Senan James Doyle, Florence Forbes, Stéphane Girard, Jonathan El-Methni, Gildas Mazo, Angelika Studeny.

A second domain of applications concerns biology and medicine. We consider the use of missing data models in epidemiology. We also investigated statistical tools for the analysis of bacterial genomes beyond gene detection. Applications in population genetics and neurosciences are also considered. Finally, in the context of the ANR VMC project Medup, we study the uncertainties on the forecasting and climate projection for Mediterranean high-impact weather events.

4.3. Reliability

Participants: Jean-Baptiste Durand, Stéphane Girard.

Reliability and industrial lifetime analysis are applications developed through collaborations with the EDF research department and the LCFR laboratory (Laboratoire de Conduite et Fiabilité des Réacteurs) of CEA / Cadarache. We also consider failure detection in print infrastructure [16] through collaboration with Xerox, Meylan.

MODAL Project-Team

4. Application Domains

4.1. Application Domains

Potential application areas of statistical modelling for heterogeneous data are extensive but some particular areas are identified. For historical reasons and considering the background of the team members, MODAL is mainly focused on *biological applications* where new challenges in high throughput technologies are opened. In addition, other secondary applications areas are considered in *industry, retail, credit scoring* and astronomy. Several contacts and collaborations are already established with some partners in these application areas and are described in Sections [7.](#) and [8.1](#) .

NACHOS Project-Team

4. Application Domains

4.1. Computational electromagnetics

Electromagnetism has found and continues to find applications in a wide range of areas, encompassing both industrial and societal purposes. Applications of current interest include those related to communications (e.g. transmission through optical fiber lines), to biomedical devices and health (e.g. tomography, power-line safety, etc.), to circuit or magnetic storage design (electromagnetic compatibility, hard disc operation), to geophysical prospecting, and to non-destructive evaluation (e.g. crack detection), to name but just a few. Although the principles of electromagnetics are well understood, their application to practical configurations of current interest is significantly complicated and far beyond manual calculation in all but the simplest cases. These complications typically arise from the geometrical characteristics of the propagation medium (irregular shapes, geometrical singularities), the physical characteristics of the propagation medium (heterogeneity, physical dispersion and dissipation) and the characteristics of the sources (wires, etc.). The significant advances in computer technology that have taken place over the last two decades have been such that numerical modeling and computer simulation is nowadays ubiquitous in the study of electromagnetic interactions. The team is actively contributing to the design of advanced numerical methodologies for the solution of the PDE models of electromagnetism with a focus on problems relevant to computational bioelectromagnetics on one hand, and computational nanophotonics on the other hand. In the first case, the objective is to develop numerical methodologies for the simulation of the interaction of electromagnetic waves with biological tissues at microwave frequencies (see Fig. 1). Potential applications are concerned with the evaluation of potential sanitary effects of human exposure to electromagnetic waves, or with the design of biomedical devices and systems (i.e. imaging systems, implantable antennas, etc.). In the second case, the target simulations involve the interaction of electromagnetic wave with nanoparticles at optical frequencies. Indeed, nanophotonics is the branch of nanotechnology that deals with the study and behaviour of light and optics at nanometer scale. It directly deals with the optics and is widely used in optic engineering. It includes all the phenomena that are used in optical sciences for the development of optical devices. Therefore, nanophotonics finds numerous applications such as in optical microscopy, the design of optical switches and electromagnetic chips circuits, transistor filaments, etc.

4.2. Computational geoseismics

Computational challenges in geoseismics span a wide range of disciplines and have significant scientific and societal implications. Two important topics are mitigation of seismic hazards and discovery of economically recoverable petroleum resources. The team is before all considering the first of these topics. Indeed, to understand the basic science of earthquakes and to help engineers better prepare for such an event, scientists want to identify which regions are likely to experience the most intense shaking, particularly in populated sediment-filled basins. This understanding can be used to improve building codes in high risk areas and to help engineers design safer structures, potentially saving lives and property. In the absence of deterministic earthquake prediction, forecasting of earthquake ground motion based on simulation of scenarios is one of the most promising tools to mitigate earthquake related hazard. This requires intense modeling that meets the spatial and temporal resolution scales of the continuously increasing density and resolution of the seismic instrumentation, which record dynamic shaking at the surface, as well as of the basin models. Another important issue is to improve our physical understanding of the earthquake rupture processes and seismicity. Large-scale simulations of earthquake rupture dynamics, and of fault interactions, are currently the only means to investigate these multi-scale physics together with data assimilation and inversion. High resolution models are also required to develop and assess fast operational analysis tools for real time seismology and early warning systems. Modeling and forecasting earthquake ground motion in large basins is a challenging and

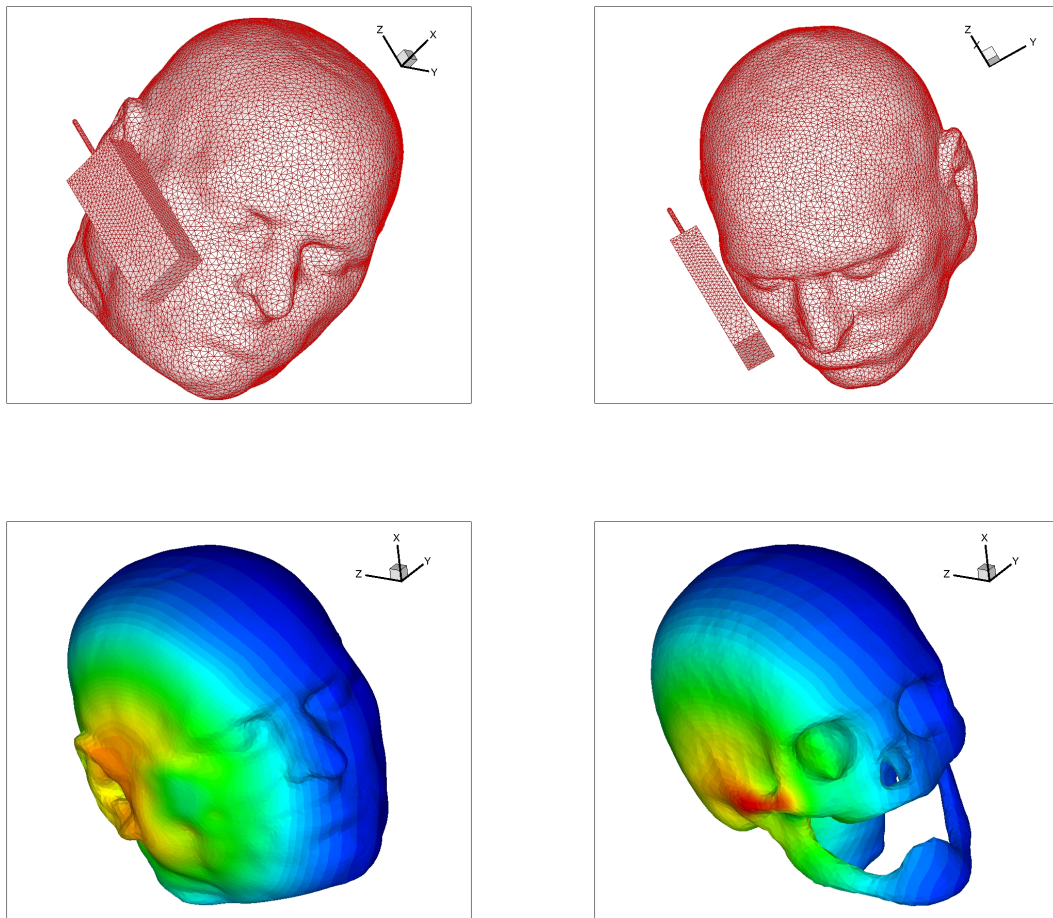


Figure 1. Exposure of head tissues to an electromagnetic wave emitted by a localized source. Top figures: surface triangulations of the skin and the skull. Bottom figures: contour lines of the amplitude of the electric field.

complex task. The complexity arises from several sources. First, multiple scales characterize the earthquake source and basin response: the shortest wavelengths are measured in tens of meters, whereas the longest measure in kilometers; basin dimensions are on the order of tens of kilometers, and earthquake sources up to hundreds of kilometers. Second, temporal scales vary from the hundredth of a second necessary to resolve the highest frequencies of the earthquake source up to as much as several minutes of shaking within the basin. Third, many basins have a highly irregular geometry. Fourth, the soil's material properties are highly heterogeneous. And fifth, geology and source parameters are observable only indirectly and thus introduce uncertainty in the modeling process. In this context, the team undertakes research and development activities aiming at the design of numerical modeling strategies for accurately and efficiently handling the interaction of seismic waves generated by an earthquake source with complex geological media. These activities are conducted in the framework of a collaboration with CETE Méditerranée <http://www.cete-mediterranee.fr/gb> which is a regional technical and engineering centre whose activities are concerned with seismic risk assessment studies.

NANO-D Team

4. Application Domains

4.1. Overview

NANO-D is *a priori* concerned with all applications domains involving atomistic representations, including chemistry, physics, electronics, material science, biology, etc.

Historically, though, our first applications have been in biology, as the next two sections detail. Thanks to the development of algorithms to efficiently simulate reactive force fields, as well as to perform interactive quantum mechanical calculations, however, we now have the possibility to address problems in chemistry, and physics.

4.2. Structural Biology

Structural biology is a branch of molecular biology, biochemistry, and biophysics concerned with the molecular structure of biological macromolecules, especially proteins and nucleic acids. Structural biology studies how these macromolecules acquire the structures they have, and how alterations in their structures affect their function. The methods that structural biologists use to determine the structure typically involve measurements on vast numbers of identical molecules at the same time, such as X-Ray crystallography, NMR, cryo-electron microscopy, etc. In many cases these methods do not directly provide the structural answer, therefore new combinations of methods and modeling techniques are often required to advance further.

We develop a set of tools that help biologists to model structural features and motifs not resolved experimentally and to understand the function of different structural fragments.

- Symmetry is a frequent structural trait in molecular systems. For example, most of the water-soluble and membrane proteins found in living cells are composed of symmetrical subunits, and nearly all structural proteins form long oligomeric chains of identical subunits. Only a limited number of symmetry groups is allowed in crystallography, and thus, in many cases the native macromolecular conformation is not present on high-resolution X-ray structures. Therefore, to understand the realistic macromolecular packing, modeling techniques are required.
- Many biological experiments are rather costly and time-demanding. For instance, the complexity of mutagenesis experiments grows exponentially with the number of mutations tried simultaneously. In other experiments, many candidates are tried to obtain a desired function. For example, about 250,000 candidates were tested for the recently discovered antibiotic Platensimycin. Therefore, there is a vast need in advance modeling techniques that can predict interactions and foresee the function of new structures.
- Structure of many macromolecules is still unknown. For other complexes, it is known only partially. Thus, software tools and new algorithms are needed by biologists to model missing structural fragments or predict the structure of those molecule, where there is no experimental structural information available.

4.3. Pharmaceuticals and Drug Design

Drug design is the inventive process of finding new medications based on the knowledge of the biological target. The drug is most commonly an organic small molecule which activates or inhibits the function of a biomolecule such as a protein, which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves design of small molecules that are complementary in shape and charge to the biomolecular target to which they interact and therefore will bind to it. Drug design frequently relies on computer modeling techniques. This type of modeling is often referred to as computer-aided drug design.

Structure-based drug design attempts to use the structure of proteins as a basis for designing new ligands by applying accepted principles of molecular recognition. The basic assumption underlying structure-based drug design is that a good ligand molecule should bind tightly to its target. Thus, one of the most important principles for designing or obtaining potential new ligands is to predict the binding affinity of a certain ligand to its target and use it as a criterion for selection.

We develop new methods to estimate the binding affinity using an approximation to the binding free energy. This approximation is assumed to depend on various structural characteristics of a representative set of native complexes with their structure solved to a high resolution. We study and verify different structural characteristics, such as radial distribution functions, and their affect on the binding free energy approximation.

4.4. Nano-engineering

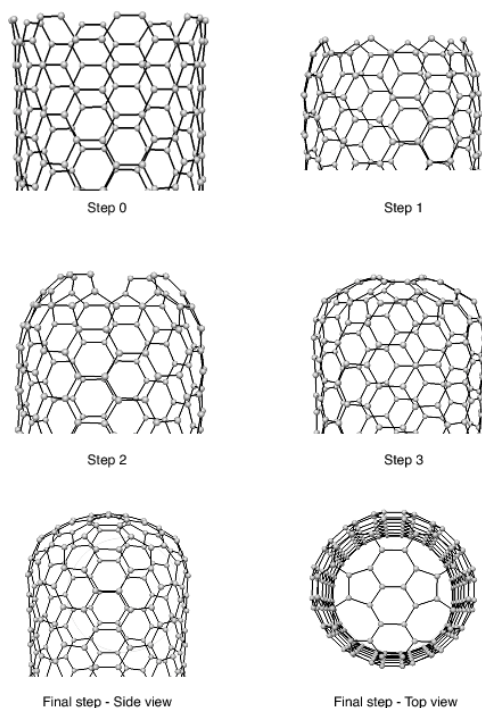


Figure 4. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.

The magazine Science has recently featured a paper demonstrating an example of DNA nanotechnology, where DNA strands are stacked together through programmable self-assembly. In February 2007, the cover of Nature Nanotechnology showed a “nano-wheel” composed of a few atoms only. Several nanosystems have already been demonstrated, including a wheelbarrow molecule, a nano-car and a Morse molecule, etc. Typically, these nanosystems are designed in part *via* quantum mechanics calculations, such as the semi-empirical ASED+ calculation technique.

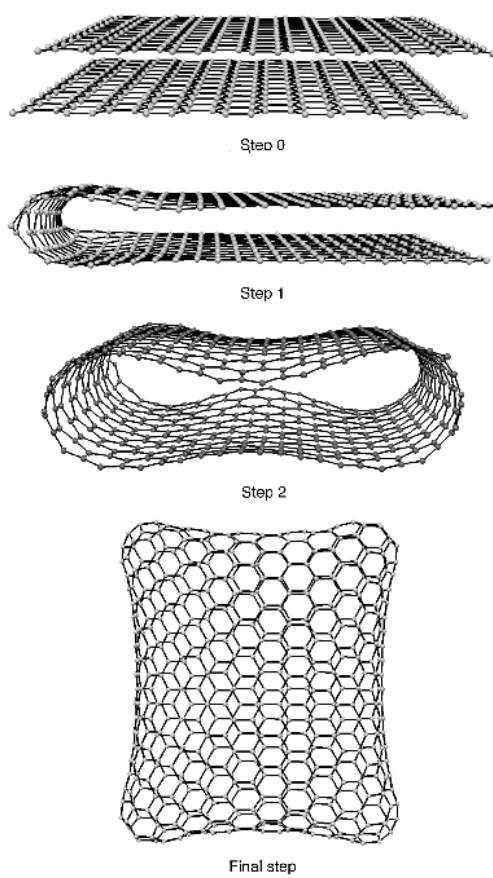


Figure 5. Different steps to prototype a “nano-pillow” with the adaptive interactive modeler.

Of course, not all small systems that currently fall under the label “nano” have mechanical, electronic, optical properties similar to the examples given above. Furthermore, current construction capabilities lack behind some of the theoretical designs which have been proposed. However, the trend is clearly for adding more and more functionality to nanosystems. While designing nanosystems is still very much an art mostly performed by physicists, chemists and biologists in labs throughout the world, there is absolutely no doubt that fundamental engineering practices will progressively emerge, and that these practices will be turned into quantitative rules and methods. Similar to what has happened with macroscopic engineering, powerful and generic software will then be employed to engineer complex nanosystems.

We have recently shown that our incremental and adaptive algorithms allow us to easily edit and model complex shapes, such as a nanotube (Fig. 4) and the “nano-pillow” below (Fig. 5).

NECS Project-Team

4. Application Domains

4.1. Application domains

Closing feedback loops around Wireless sensor networks offer new challenges and new opportunities for the area of control. Several new application areas can be enabled, or enhanced if systematic methods are developed for the design of NCS. Examples include:

- Intelligent transportation systems, where traffic flow or density can be measured using novel wireless technologies and used to determine control inputs such as on-ramp metering schemes and variable message signs.
- Intelligent buildings, where sensor information on CO_2 concentration, temperature, room occupancy, etc. can be used to control the heating, ventilation and air conditioning (HVAC) system under multi-objective considerations of comfort, air quality, and energy consumption.
- Disaster relief operations, where data collected by sensor networks can be used to guide the actions of rescue crews and operate automated rescue equipment.
- Surveillance using swarms of Uninhabited Aerial Vehicles (UAVs), where sensor information (from sensors on the ground and/or on-board the vehicles) can be used to guide the UAVs to accomplish their mission.
- Environmental monitoring and exploration using schools of Autonomous Underwater Vehicles (AUVs), where underwater sensors and communication are used to guide the AUVs.
- Infrastructure security and protection using smart camera networks, where the images collected are shared among the cameras and used to control the cameras themselves (pan-tilt-zoom) and ensure tracking of potential threat.

In particular, the team is already involved in the areas described in detail below.

4.2. Vehicular transportation systems

4.2.1. Car industry

Car industry has been already identified as a potential homeland application for NCS [67], as the evolution of micro-electronics paved the way for introducing distributed control in vehicles. In addition, automotive control systems are becoming the more complex and iterative, as more on-board sensors and actuators are made available through technology innovations. The increasing number of subsystems, coupled with overwhelming information made available through on-board and off-board sensors and communication systems, rises new and interesting challenges to achieve optimal performance while maintaining the safety and the robustness of the total system. Causes of such an increase of complexity/difficulties are diverse: interaction between several control sub-systems (ABS, TCS, ESP, etc.), loose of synchrony between sub-systems, limitations in the computation capabilities of each dedicate processor, etc. The team had several past collaborations with the car industry (Renault since 1992, and Ford). In addition, in the ANR project VOLHAND, in progress, the team works on developing a new generation of electrical power-assisted steering specifically designed for disabled and aged persons. More recently, a grant with IFP has been signed with the aim of studying the potential in terms of energy saving and traffic improvement of communicating vehicles.

4.2.2. Intelligent transportation systems

Throughout the world, roadways are notorious for their congestion, from dense urban network to large freeway systems. This situation tends to get worse over time due to the continuous increase of transportation demand whereas public investments are decreasing and space is lacking to build new infrastructures. The most obvious impact of traffic congestion for citizens is the increase of travel times and fuel consumption. Another critical effect is that infrastructures are not operated at their capacity during congestion, implying that fewer vehicles are served than the amount they were designed for. Using macroscopic fluid-like models, the NeCS team has initiated new researches to develop innovative traffic management policies able to improve the infrastructure operations. This activity is currently focused on automatic model calibration and traffic prediction, two important items to implement efficient Intelligent Transportation Systems (ITS) such as traffic responsive ramp metering and varying speed limit as well as producing relevant user information. The team is currently setting up a consortium with local authorities involved in traffic management to build to a demonstrator called GTL (Grenoble Traffic Lab). One target of this activity is to transfer part of the developed technology to a start-up named Karrus.

4.3. Underwater systems

Underwater systems, as presently used or intended by the offshore industry and marine research, are subject to severe technological constraints. In AUVs, the on-board power is limited and calls for both control and computing optimization. The links between the master and slave nodes use acoustic devices, which have a very low bandwidth and are subject to frequent transient loss, thus calling for sharing the decisional process among the nodes and for a robust implementation of the distributed control, taking into account the communication network features. These constraints together with the potential cost of failures make these systems good candidates for safe and flexible control, communication and computing co-design. The team already got a significant experience in this domain with a past collaboration with IFREMER and other EU projects. The projects CONNECT and FeedNetBack dealt with this type of problems (see 8.3.1.1).

4.4. Systems on chip

Achieving a good compromise between computing power and energy consumption is one of the challenge in embedded architecture of the future. This management is especially difficult for 45nm or 32nm known to be at the limit of the scalability. Automatic control loops have therefore to be designed in order to make the performance fit the requirement in order to minimize the energy loss in a context of highly unknown performance of the chip. The main objective is to control the computing power and the consumption using the voltage and frequency automatically according to the requirements of the OS. For this, appropriate sensors must be implemented on the chip and a high-performance repartition between hardware and software implementation must be made.

NON-A Project-Team

4. Application Domains

4.1. Application domains

Unlike the traditional methods, the estimators we defined are "non-asymptotic": solutions are provided by an explicit formulae. They result in relatively simple and fast algorithms. In this sense, rather than being a project linked to a specific domain of application, we can say that the project Non-A is method-driven. However, one must not forget that applicability remains a guideline in all our researches. As it has been noted, estimation is a huge area, which touches a variety of possible application fields, which our new methods address. Figure 3 illustrates the connections between our techniques and the possible applications.

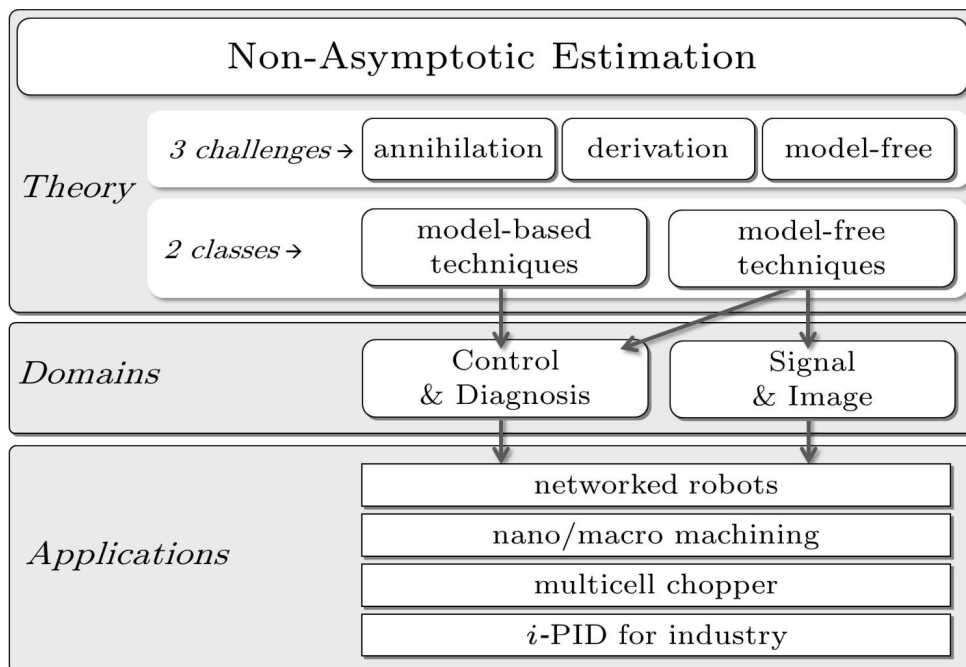


Figure 3. Non-A is a method-driven project, centered around non-asymptotic estimation techniques (i.e., providing estimates in finite-time), and connected to applications.

During these few years, our techniques have already generated 3 patents [98], [100], [99]. It shows their efficiency in various industrial domains, including (see the previous reports):

- Vehicle control (engine throttle [112], lateral and longitudinal velocities [96], stop-and-go [131], tire/road contact condition [134]) with PSA, APEDGE, Mines-ParisTech, Inria IMARA, Universidad Carlos III (Madrid), Université Paris Sud, spark ignition engine control with GM [28];
- Hydroelectric power plants [111], [110] with EDF-CIH (patent pending FR0858532);
- Shape memory actuators [107] with Université de Bretagne Occidentale and ANR MAFESMA;
- Magnetic actuators with Univ. des Saarlandes;

- Power Electronics [121] with Univ. du Québec à Trois-Rivières;
- Aircraft identification [130] with ONERA DCSD;
- Secured communications (chaos-based cryptography [128], [136], [135], CPM demodulation [122]) with CINVESTAV Mexico, Math.Dept. Tlemcen Univ. Algeria and PRISME ENSI-Bourges.
- Image and video processing (denoising [102], edge detection [113]) with Inria QGAR, compression [103], compressive sensing [132], [133] with CINVESTAV Mexico and Whuan Univ., China;
- Financial engineering [97] with MEREOR Investment Management and Advisory SAS;
- Neural signal processing [35], [83];
- Stabilization of a ball & beam system through saturated control [91], [81]
- Control of a new class of power converters [48], [49];
- Power management of a hydrogen fuel cell system combined with super capacitors [69];
- Sensorless parameter identification for permanent magnet stepper motors [60], [58], [59];
- A flexible robot joint control [78];
- Localization of mobile robots [82];
- Teleoperation [46], [44], [45];
- Oscillatory failure case (OFC) detection for aircraft [38].

OPALE Project-Team

4. Application Domains

4.1. Aeronautics and space

The demand of the aeronautical industry remains very strong in aerodynamics, as much for conventional aircraft, whose performance must be enhanced to meet new societal requirements in terms of economy, noise (particularly during landing), vortex production near runways, etc., as for high-capacity or supersonic aircraft of the future. Our implication concerns shape optimization of wings or simplified configurations.

Our current involvement with Space applications relates to software platforms for code coupling.

4.2. Mechanical industry

A new application domain related to the parameter and shape optimization of mechanical structures is under active development. The mechanical models range from linear elasticity of 2D or 3D structures, or thin shells, to nonlinear elastoplasticity and structural dynamics. The criteria under consideration are multiple: formability, stiffness, rupture, fatigue, crash, and so on. The design variables are the thickness and shape, and possibly the topology, of the structures. The applications are performed in collaboration with world-leading industrials, and involve the optimization of the stamping process (Blank Force, Die and Tools shapes) of High Performance steel structures as well as the optimal design of structures used for packaging purposes (cans and sprays under high pressure). Our main contribution relies on providing original and efficient algorithms to capture Pareto fronts, using smart meta-modelling, and to apply game theory approaches and algorithms to propose stable compromise solutions (e.g. Nash equilibria).

4.3. Electromagnetics

In the context of shape optimization of antennas, we can split the existing results in two parts: the two-dimensional modeling concerning only the specific transverse mode TE or TM, and treatments of the real physical 3-D propagation accounting for no particular symmetry, whose objective is to optimize and identify real objects such as antennas.

Most of the numerical literature in shape optimization in electromagnetics belongs to the first part and makes intensive use of the 2-D solvers based on the specific 2-D Green kernels. The 2-D approach for the optimization of *directivity* led recently to serious errors due to the modeling defect. There is definitely little hope for extending the 2-D algorithms to real situations. Our approach relies on a full analysis in unbounded domains of shape sensitivity analysis for the Maxwell equations (in the time-dependent or harmonic formulation), in particular, by using the integral formulation and the variations of the Colton and Kreiss isomorphism. The use of the France Telecom software SR3D enables us to directly implement our shape sensitivity analysis in the harmonic approach. This technique makes it possible, with an adequate interpolation, to retrieve the shape derivatives from the physical vector fields in the time evolution processes involving initial impulses, such as radar or tomography devices, etc. Our approach is complementary to the “automatic differentiation codes” which are also very powerful in many areas of computational sciences. In Electromagnetics, the analysis of hyperbolic equations requires a sound treatment and a clear understanding of the influence of space approximation.

4.4. Biology and medicine

A particular effort is made to apply our expertise in solid and fluid mechanics, shape and topology design, multidisciplinary optimization by game strategies to biology and medicine. Two selected applications are privileged: solid tumors and wound healing.

Opale's objective is to push further the investigation of these applications, from a mathematical-theoretical viewpoint and from a computational and software development viewpoint as well. These studies are led in collaboration with biologists, as well as image processing specialists.

4.5. Traffic flow

The modeling and analysis of traffic phenomena can be performed at a macroscopic scale by using partial differential equations derived from fluid dynamics. Such models give a description of collective dynamics in terms of the spatial density $\rho(t, x)$ and average velocity $v(t, x)$. Continuum models have shown to be in good agreement with empirical data. Moreover, they are suitable for analytical investigations and very efficient from the numerical point of view. Finally, they contain only few variables and parameters and they can be very versatile in order to describe different situations encountered in practice.

Opale's research focuses on the study of macroscopic models of vehicular and pedestrian traffic, and how optimal control approaches can be used in traffic management. The project opens new perspectives of interdisciplinary collaborations on urban planning and crowd dynamics analysis.

4.6. Multidisciplinary couplings

Our expertise in theoretical and numerical modeling, in particular in relation to approximation schemes, and multilevel, multi-scale computational algorithms, allows us to envisage to contribute to integrated projects focused on disciplines other than, or coupled with fluid dynamics, such as structural mechanics, electromagnetics, biology and virtual reality, image processing, etc in collaboration with specialists of these fields. Part of this research is conducted in collaboration with ONERA.

POEMS Project-Team

4. Application Domains

4.1. Introduction

We are concerned with all application domains where linear wave problems arise: acoustics and elastodynamics (including fluid-structure interactions), electromagnetism and optics, and gravity water waves. We give in the sequel some details on each domain, pointing out our main motivations and collaborations.

4.2. Acoustics

As the acoustic propagation in a fluid at rest can be described by a scalar equation, it is generally considered by applied mathematicians as a simple preliminary step for more complicated (vectorial) models. However, several difficult questions concerning coupling problems have occupied our attention recently. Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, is for our team a new and very challenging topic, which gives rise to a lot of open questions, from the modeling (Euler equations, Galbrun equations) to the numerical approximation of such models (which poses new difficulties). Our works in this area are partially supported by EADS and Airbus. The typical objective is to reduce the noise radiated by Airbus planes. Vibroacoustics, which concerns the interaction between sound propagation and vibrations of thin structures, also raises up a lot of relevant research subjects.

Both applications (Aeroacoustics and Vibroacoustics) led us in particular to develop an academic research between volumic methods and integral equations in time domain.

Finally, A particularly attractive application concerns the simulation of musical instruments, whose objectives are both a better understanding of the behavior of existing instruments and an aid for the manufacturing of new instruments. The modeling and simulation of the timpani and of the guitar have been carried out in collaboration with A. Chaigne of ENSTA. We are currently on the piano.

4.3. Electromagnetism

This is a particularly important domain, first because of the very important technological applications but also because the treatment of Maxwell's equations is much more technically involved from the mathematical point of view than the scalar wave equation. Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology, electromagnetic compatibility, design of optoelectronic micro-components or smart materials. Stealth technology relies in particular on the conception and simulation of new absorbing materials (anisotropic, chiral, non-linear...). The simulation of antennas raises delicate questions related to the complexity of the geometry (in particular the presence of edges and corners). In optics, the development of the micro and nano optics has made recently fantastic progress and the thematic of metamaterials (with negative index of refraction) opens new amazing applications. For all these reasons, we are developing an intense research in the following areas

- Highly accurate and hybrid numerical methods in collaboration with CEA (Gramat) and ONERA (Toulouse).
- Electromagnetic wave propagation in periodic media.
- Development of simplified approximate models by asymptotic analysis for various applications : boundary layers, thin coatings, thin domains, thin wires and cables, ...
- Mathematical and numerical questions linked to the modeling of metamaterials.

4.4. Elastodynamics

Wave propagation in solids is with no doubt, among the three fundamental domains that are acoustics, electromagnetism and elastodynamics, the one that poses the most significant difficulties from mathematical and numerical points of view.

Our activity on this topic began with applications in geophysics, which unfortunately has been forced to slow down in the middle of the 90's due to the disengagement of French oil companies in matter of research. However it has seen a most welcomed rebound through new academic problems (in particular surface waves, perfectly matched layers techniques, inverse problems in wave guides) and industrial contacts, more precisely with CEA-LIST with which we have developed a long term collaboration in the domain of non destructive testing by ultrasounds. The most recent problems we have been dealing with in this domain concern elastic wave propagation in plates, the modeling of piezoelectric devices or elastic wave propagation in highly heterogeneous media.

REALOPT Project-Team

4. Application Domains

4.1. Introduction

Our group has tackled applications in logistics, transportation and routing [57], [56], [53], [55], in production planning [70] and inventory control [53], [55], in network design and traffic routing [39], [46], [51], [75], [36], [47], [61], [66], in cutting and placement problems [60], [67], [68], [69], [71], and in scheduling [7], [62], [34].

4.2. Network Design and Routing Problems

We are actively working on problems arising in network topology design, implementing a survivability condition of the form “at least two paths link each pair of terminals”. We have extended polyhedral approaches to problem variants with bounded length requirements and re-routing restrictions [46]. Associated to network design is the question of traffic routing in the network: one needs to check that the network capacity suffices to carry the demand for traffic. The assignment of traffic also implies the installation of specific hardware at transient or terminal nodes.

To accommodate the increase of traffic in telecommunication networks, today’s optical networks use grooming and wavelength division multiplexing technologies. Packing multiple requests together in the same optical stream requires to convert the signal in the electrical domain at each aggregation of disaggregation of traffic at an origin, a destination or a bifurcation node. Traffic grooming and routing decisions along with wavelength assignments must be optimized to reduce opto-electronic system installation cost. We developed and compared several decomposition approaches [77], [76], [75] to deal with backbone optical network with relatively few nodes (around 20) but thousands of requests for which traditional multi-commodity network flow approaches are completely overwhelmed. We also studied the impact of imposing a restriction on the number of optical hops in any request route [74]. We also developed a branch-and-cut approach to a problem that consists in placing sensors on the links of a network for a minimum cost [51], [52].

We studied several time dependent formulations for the unit demand vehicle routing problem [41], [40]. We gave new bounding flow inequalities for a single commodity flow formulation of the problem. We described their impact by projecting them on some other sets of variables, such as variables issued of the Picard and Queyranne formulation or the natural set of design variables. Some inequalities obtained by projection are facet defining for the polytope associated with the problem. We are now running more numerical experiments in order to validate in practice the efficiency of our theoretical results.

We also worked on the p-median problem, applying the matching theory to develop an efficient algorithm in Y-free graphs and to provide a simple polyhedral characterization of the problem and therefore a simple linear formulation [65] simplifying results from Baiou and Barahona.

We considered the multi-commodity transportation problem. Applications of this problem arise in, for example, rail freight service design, “less than truckload” trucking, where goods should be delivered between different locations in a transportation network using various kinds of vehicles of large capacity. A particularity here is that, to be profitable, transportation of goods should be consolidated. This means that goods are not delivered directly from the origin to the destination, but transferred from one vehicle to another in intermediate locations. We proposed an original Mixed Integer Programming formulation for this problem which is suitable for resolution by a Branch-and-Price algorithm and intelligent primal heuristics based on it.

For the problem of routing freight railcars, we proposed two algorithmes based on the column generation approach. These algorithmes have been testes on a set of real-life instances coming from a Russian freight real transportation company. Our algorithmes have been faster on these instances than the current solution approach being used by the company.

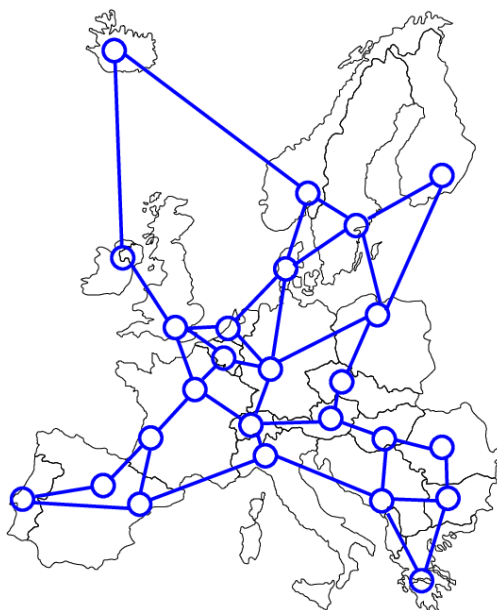


Figure 1. Design of a SDH/SONET european network where demands are multiplexed.

4.3. Packing and Covering Problems

We developed a branch-and-price algorithm for the Bin Packing Problem with Conflicts which improves on other approaches available in the literature [20]. The algorithm uses our methodological advances like the generic branching rule for the branch-and-price and the column based heuristic. One of the ingredients which contributes to the success of our method are fast algorithms we developed for solving the subproblem which is the Knapsack Problem with Conflicts. Two variants of the subproblem have been considered: with interval and arbitrary conflict graphs. The paper which presents this work is being finalized.

We have designed a new algorithm for vertex packing (equivalently stable set) in claw-free graphs [58]. Previously the best known algorithm for this problem had a running time of $O(n^6)$ (with n the number of vertices in the graph) while our new algorithm runs in $O(n^4)$.

We studied a variant of the knapsack problem encountered in inventory routing problem [55]: we faced a multiple-class integer knapsack problem with setups [54] (items are partitioned into classes whose use implies a setup cost and associated capacity consumption). We showed the extent to which classical results for the knapsack problem can be generalized to this variant with setups and we developed a specialized branch-and-bound algorithm.

We studied the orthogonal knapsack problem, with the help of graph theory [50], [48], [12], [11]. Fekete and Schepers proposed to model multi-dimensional orthogonal placement problems by using an efficient representation of all geometrically symmetric solutions by a so called *packing class* involving one *interval graph* for each dimension. Though Fekete & Schepers' framework is very efficient, we have however identified several weaknesses in their algorithms: the most obvious one is that they do not take advantage of the different possibilities to represent interval graphs. We propose to represent these graphs by matrices with consecutive ones on each row. We proposed a branch-and-bound algorithm for the 2d knapsack problem that uses our 2D packing feasibility check.

4.4. Planning, Scheduling, and Logistic Problems

Inventory routing problems combine the optimization of product deliveries (or pickups) with inventory control at customer sites. We considered an industrial application where one must construct the planning of single product pickups over time; each site accumulates stock at a deterministic rate; the stock is emptied on each visit. We have developed a truncated branch-and-price algorithm: periodic plans are generated for vehicles by solving a multiple choice knapsack subproblem; the global planning of customer visits is generated by solving a master program. Confronted with the issue of symmetry in time, we used a state-space relaxation idea. Our algorithm provides solutions with reasonable deviation from optimality for large scale problems (260 customer sites, 60 time periods, 10 vehicles) coming from industry [13]. We previously developed approximate solutions to a related problem combining vehicle routing and planning over a fixed time horizon (solving instances involving up to 6000 pick-ups and deliveries to plan over a twenty day time horizon with specific requirements on the frequency of visits to customers [57]).

Together with our partner company GAPSO from the associate team SAMBA, we are working on the equipment routing task scheduling problem [28] arising during port operations. In this problem, a set of tasks needs to be performed using equipments of different types with the objective of maximizing the weighted sum of performed tasks.

We participated to the project on an airborne radar scheduling. For this problem, we developed fast heuristics [45] and exact algorithms [34]. A substantial research has been done on machine scheduling problems. A new compact MIP formulation was proposed for a large class of these problems [33]. An exact decomposition algorithm was developed for the NP-hard maximizing the weighted number of late jobs problem on a single machine [62]. A dominant class of schedules for malleable parallel jobs was discovered in the NP-hard problem to minimize the total weighted completion time [64]. We proved that a special case of the scheduling problem at cross docking terminals to minimize the storage cost is polynomially solvable [19], [63].

Another application area in which we have successfully developed MIP approaches is in the area of tactical production and supply chain planning. In [14], we proposed a simple heuristic for challenging multi-echelon problems that makes effective use of a standard MIP solver. [14] contains a detailed investigation of what makes solving the MIP formulations of such problems challenging; it provides a survey of the known methods for strengthening formulations for these applications, and it also pinpoints the specific substructure that seems to cause the bottleneck in solving these models. Finally, the results of [35] provide demonstrably stronger formulations for some problem classes than any previously proposed.

We have been developing **robust optimization** models and methods to deal with a number of applications like the above in which uncertainty is involved. In [43], [42], we analyzed fundamental MIP models that incorporate uncertainty and we have exploited the structure of the stochastic formulation of the problems in order to derive algorithms and strong formulations for these and related problems. These results appear to be the first of their kind for structured stochastic MIP models. In addition, we have engaged in successful research to apply concepts such as these to health care logistics [37]. We considered train timetabling problems and their re-optimization after a perturbation in the network [44]. The question of formulation is central. Models of the literature are not satisfactory: continuous time formulations have poor quality due to the presence of discrete decision (re-sequencing or re-routing); arc flow in time-space graph blow-up in size (they can only handle a single line timetabling problem). We have developed a discrete time formulation that strikes a compromise between these two previous models. Based on various time and network aggregation strategies, we develop a 2-stage approach, solving the contiguous time model having fixed the precedence based on a solution to the discrete time model.

REGULARITY Project-Team

4. Application Domains

4.1. Uncertainties management

Our theoretical works are motivated by and find natural applications to real-world problems in a general frame generally referred to as uncertainty management, that we describe now.

Since a few decades, modeling has gained an increasing part in complex systems design in various fields of industry such as automobile, aeronautics, energy, etc. Industrial design involves several levels of modeling: from behavioural models in preliminary design to finite-elements models aiming at representing sharply physical phenomena. Nowadays, the fundamental challenge of numerical simulation is in designing physical systems while saving the experimentation steps.

As an example, at the early stage of conception in aeronautics, numerical simulation aims at exploring the design parameters space and setting the global variables such that target performances are satisfied. This iterative procedure needs fast multiphysical models. These simplified models are usually calibrated using high-fidelity models or experiments. At each of these levels, modeling requires control of uncertainties due to simplifications of models, numerical errors, data imprecisions, variability of surrounding conditions, etc.

One dilemma in the design by numerical simulation is that many crucial choices are made very early, and thus when uncertainties are maximum, and that these choices have a fundamental impact on the final performances.

Classically, coping with this variability is achieved through *model registration* by experimenting and adding fixed *margins* to the model response. In view of technical and economical performance, it appears judicious to replace these fixed margins by a rigorous analysis and control of risk. This may be achieved through a probabilistic approach to uncertainties, that provides decision criteria adapted to the management of unpredictability inherent to design issues.

From the particular case of aircraft design emerge several general aspects of management of uncertainties in simulation. Probabilistic decision criteria, that translate decision making into mathematical/probabilistic terms, require the following three steps to be considered [58]:

1. build a probabilistic description of the fluctuations of the model's parameters (*Quantification of uncertainty sources*),
2. deduce the implication of these distribution laws on the model's response (*Propagation of uncertainties*),
3. and determine the specific influence of each uncertainty source on the model's response variability (*Sensitivity Analysis*).

The previous analysis now constitutes the framework of a general study of uncertainties. It is used in industrial contexts where uncertainties can be represented by *random variables* (unknown temperature of an external surface, physical quantities of a given material, ... at a given *fixed time*). However, in order for the numerical models to describe with high fidelity a phenomenon, the relevant uncertainties must generally depend on time or space variables. Consequently, one has to tackle the following issues:

- *How to capture the distribution law of time (or space) dependent parameters, without directly accessible data?* The distribution of probability of the continuous time (or space) uncertainty sources must describe the links between variations at neighbor times (or points). The local and global regularity are important parameters of these laws, since it describes how the fluctuations at some time (or point) induce fluctuations at close times (or points). The continuous equations representing the studied phenomena should help to *propose models for the law of the random fields*. Let us notice that interactions between various levels of modeling might also be used to derive distributions of probability at the lowest one.

- The navigation between the various natures of models needs a kind of *metric* which could *mathematically describe the notion of granularity or fineness* of the models. Of course, the local regularity will not be totally absent of this mathematical definition.
- All the various levels of conception, preliminary design or high-fidelity modelling, require *registrations by experimentation* to reduce model errors. This *calibration* issue has been present in this frame since a long time, especially in a deterministic optimization context. The random modeling of uncertainty requires the definition of a systematic approach. The difficulty in this specific context is: statistical estimation with few data and estimation of a function with continuous variables using only discrete setting of values.

Moreover, a multi-physical context must be added to these questions. The complex system design is most often located at the interface between several disciplines. In that case, modeling relies on a coupling between several models for the various phenomena and design becomes a *multidisciplinary optimization* problem. In this uncertainty context, the real challenge turns robust optimization to manage technical and economical risks (risk for non-satisfaction of technical specifications, cost control).

We participate in the uncertainties community through several collaborative research projects (ANR and Pôle SYSTEM@TIC), and also through our involvement in the MASCOT-NUM research group (GDR of CNRS). In addition, we are considering probabilistic models as phenomenological models to cope with uncertainties in the DIGITEO ANIFRAC project. As explained above, we focus on essentially irregular phenomena, for which irregularity is a relevant quantity to capture the variability (e.g. certain biomedical signals, terrain modeling, financial data, etc.). These will be modeled through stochastic processes with prescribed regularity.

4.2. Design of complex systems

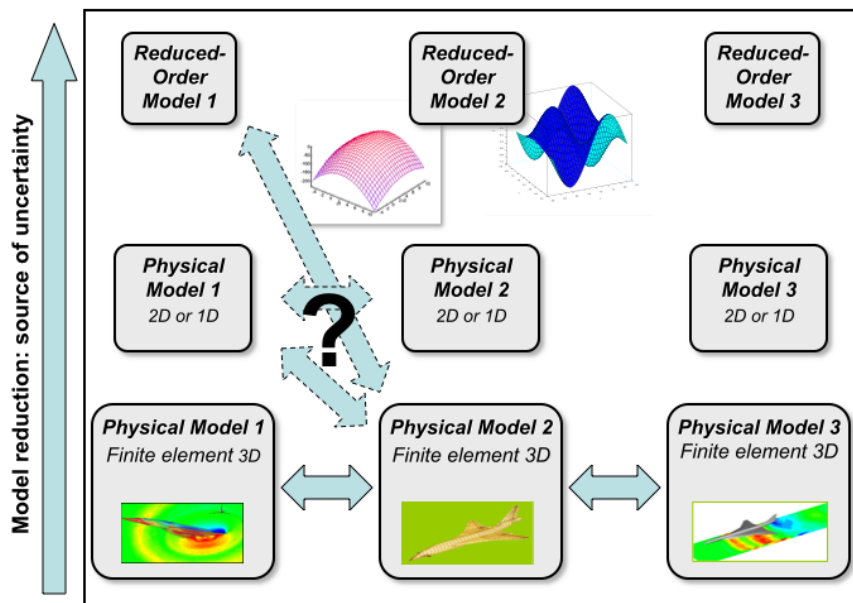


Figure 3. Coupling uncertainty between heterogeneous models

The design of a complex (mechanical) system such as aircraft, automobile or nuclear plant involves numerical simulation of several interacting physical phenomena: CFD and structural dynamics, thermal evolution of a fluid circulation, ... For instance, they can represent the resolution of coupled partial differential equations using finite element method. In the framework of uncertainty treatment, the studied "phenomenological model" is a chaining of different models representing the various involved physical phenomena. As an example, the pressure field on an aircraft wing is the result of both aerodynamic and structural mechanical phenomena. Let us consider the particular case of two models of partial differential equations coupled by limit conditions. The direct propagation of uncertainties is impossible since it requires an exploration and then, many calls to costly models. As a solution, engineers use to build reduced-order models: the complex high-fidelity model is substituted with a CPU less costly model. The uncertainty propagation is then realized through the simplified model, taking into account the approximation error (see [52]).

Interactions between the various models are usually explicitated at the finest level (cf. Fig. 3). How may this coupling be formulated when the fine structures of exchange have disappeared during model reduction? How can be expressed the interactions between models at different levels (in a multi-level modeling)? The ultimate question would be: how to choose the right level of modeling with respect to performance requirements?

In the multi-physical numerical simulation, two kinds of uncertainties then coexist: the uncertainty due to substitution of high-fidelity models with approximated reduced-order models, and the uncertainty due to the new coupling structure between reduced-order models.

According to the previous discussion, the uncertainty treatment in a multi-physical and multi-level modeling implies a large range of issues, for instance numerical resolutions of PDE (which do not enter into the research topics of *Regularity*). Our goal is to contribute to the theoretical arsenal that allows to fly among the different levels of modeling (and then, among the existing numerical simulations). We will focus on the following three axes:

- In the case of a phenomenon represented by two coupled partial differential equations whose resolution is represented by reduced-order models, how to define a probabilistic model of the coupling errors? In connection with our theoretical development, we plan to characterize the regularity of this error in order to quantify its distribution. This research axis is supported by an ANR grant (OPUS project).
- The multi-level modeling assumes the ability to choose the right level of details for the models in adequacy to the goals of the study. In order to do that, a rigorous mathematical definition of the notion of *model fineness/granularity* would be very helpful. Again, a precise analysis of the fine regularity of stochastic models is expected to give elements toward a precise definition of granularity. This research axis is supported by a Pôle SYSTEM@TIC grant (EHPOC project), and also by a collaboration with EADS.
- Some fine characteristics of the phenomenological model may be used to define the probabilistic behaviour of its variability. The action of modeling a phenomena can be seen as an interpolation issue between given observations. This interpolation can be driven by physical evolution equations or fine analytical description of the physical quantities. We are convinced that Hölder regularity is an essential parameter in that context, since it captures how variations at a given point induce variations at its neighbors. Stochastic processes with prescribed regularity (see section 3.3) have already been used to represent various fluctuating phenomena: Internet traffic, financial data, ocean floor. We believe that these models should be relevant to describe solutions of PDE perturbed by uncertain (random) coefficients or limit conditions. This research axis is supported by a Pôle SYSTEM@TIC grant (CSDL project).

The preliminary design of complex systems can be described as an exploration process of a so-called design space, generated by the global parameters. An interactive exploration, with a decisional visualization goal, needs reduced-order models of the involved physical phenomena. We are convinced that the local regularity of phenomena is a relevant quantity to drive these approximated models. Roughly speaking, in order to be representative, a model needs more informations where the fluctuations are the more important (and consequently, where irregularity is the more important).

In collaboration with Dassault Aviation, EDF and EADS, we study how the local regularity can provide a good quantification of the concept of *granularity* of a model, in order to select the good level of fidelity adapted to the required precision.

Our works in that field can be expressed into:

- The definition and the study of stochastic partial differential equations driven by processes with prescribed regularity (that do not enter into the classical theory of stochastic integration).
- The study of the evolution of the local regularity inside stochastic partial differential equations (SPDE). The stochastic 2-microlocal analysis should provide informations about the local regularity of the solutions, in function of the coefficients of the equations. The knowledge of the fine behaviour of the solution of the SPDE will provide important informations in the view of numerical simulations.

4.3. Biomedical Applications

ECG analysis and modelling

ECG and signals derived from them are an important source of information in the detection of various pathologies, including *e.g.* congestive heart failure, arrhythmia and sleep apnea. The fact that the irregularity of ECG bears some information on the condition of the heart is well documented (see *e.g.* the web resource <http://www.physionet.org>). The regularity parameters that have been studied so far are mainly the box and regularization dimensions, the local Hölder exponent and the multifractal spectrum [61], [63]. These have been found to correlate well with certain pathologies in some situations. From a general point of view, we participate in this research area in two ways.

- First, we use refined regularity characterizations, such as the regularization dimension, 2-microlocal analysis and advanced multifractal spectra for a more precise analysis of ECG data. This requires in particular to test current estimation procedures and to develop new ones.
- Second, we build stochastic processes that mimic in a faithful way some features of the dynamics of ECG. For instance, the local regularity of RR intervals, estimated in a parametric way based on a modelling by an mBm, displays correlations with the amplitude of the signal, a feature that seems to have remained unobserved so far [3]. In other words, RR intervals behave as SRP. We believe that modeling in a simplified way some aspects of the interplay between the sympathetic and parasympathetic systems might lead to an SRP, and to explain both this self-regulating property and the reasons behind the observed multifractality of records. This will open the way to understanding how these properties evolve under abnormal behaviour.

Pharmacodynamics and patient drug compliance

Poor adherence to treatment is a worldwide problem that threatens efficacy of therapy, particularly in the case of chronic diseases. Compliance to pharmacotherapy can range from 5% to 90%. This fact renders clinical tested therapies less effective in ambulatory settings. Increasing the effectiveness of adherence interventions has been placed by the World Health Organization at the top list of the most urgent needs for the health system. A large number of studies have appeared on this new topic in recent years [77], [76]. In collaboration with the pharmacy faculty of Montréal university, we consider the problem of compliance within the context of multiple dosing. Analysis of multiple dosing drug concentrations, with common deterministic models, is usually based on patient full compliance assumption, *i.e.*, drugs are administered at a fixed dosage. However, the drug concentration-time curve is often influenced by the random drug input generated by patient poor adherence behaviour, inducing erratic therapeutic outcomes. Following work already started in Montréal [70], [71], we consider stochastic processes induced by taking into account the random drug intake induced by various compliance patterns. Such studies have been made possible by technological progress, such as the “medication event monitoring system”, which allows to obtain data describing the behaviour of patients.

We use different approaches to study this problem: statistical methods where enough data are available, model-based ones in presence of qualitative description of the patient behaviour. In this latter case, piecewise deterministic Markov processes (PDP) seem a promising path. PDP are non-diffusion processes whose evolution follows a deterministic trajectory governed by a flow between random time instants, where it undergoes a jump according to some probability measure [56]. There is a well-developed theory for PDP, which studies stochastic properties such as extended generator, Dynkin formula, long time behaviour. It is easy to cast a simplified model of non-compliance in terms of PDP. This has allowed us already to obtain certain properties of interest of the random concentration of drug [44]. In the simplest case of a Poisson distribution, we have obtained rather precise results that also point to a surprising connection with infinite Bernoulli convolutions [44], [11], [10]. Statistical aspects remain to be investigated in the general case.

SCIPORT Team

4. Application Domains

4.1. Automatic Differentiation

Automatic Differentiation of programs gives sensitivities or gradients, that are useful for many types of applications:

- optimum shape design under constraints, multidisciplinary optimization, and more generally any algorithm based on local linearization,
- inverse problems, such as parameter estimation and in particular 4Dvar data assimilation in climate sciences (meteorology, oceanography),
- first-order linearization of complex systems, or higher-order simulations, yielding reduced models for simulation of complex systems around a given state,
- mesh adaptation and mesh optimization with gradients or adjoints,
- equation solving with the Newton method,
- sensitivity analysis, propagation of truncation errors.

4.2. Multidisciplinary optimization

A CFD program computes the flow around a shape, starting from a number of inputs that define the shape and other parameters. From this flow one can define optimization criteria e.g. the lift of an aircraft. To optimize a criterion by a gradient descent, one needs the gradient of the output criterion with respect to all the inputs, and possibly additional gradients when there are constraints. Adjoint-mode AD is a promising way to compute these gradients.

4.3. Inverse problems and Data Assimilation

Inverse problems aim at estimating the value of hidden parameters from other measurable values, that depend on the hidden parameters through a system of equations. For example, the hidden parameter might be the shape of the ocean floor, and the measurable values the altitude and speed of the surface.

One particular case of inverse problems is *data assimilation* [33] in weather forecasting or in oceanography. The quality of the initial state of the simulation conditions the quality of the prediction. But this initial state is largely unknown. Only some measures at arbitrary places and times are available. A good initial state is found by solving a least squares problem between the measures and a guessed initial state which itself must verify the equations of meteorology. This boils down to solving an adjoint problem, which can be done through AD [36]. Figure 1 shows an example of a data assimilation exercise using the oceanography code OPA [34] and its AD adjoint produced by TAPENADE.

The special case of 4Dvar data assimilation is particularly challenging. The 4th dimension in “4D” is time, as available measures are distributed over a given assimilation period. Therefore the least squares mechanism must be applied to a simulation over time that follows the time evolution model. This process gives a much better estimation of the initial state, because both position and time of measurements are taken into account. On the other hand, the adjoint problem involved grows in complexity, because it must run (backwards) over many time steps. This demanding application of AD justifies our efforts in reducing the runtime and memory costs of AD adjoint codes.

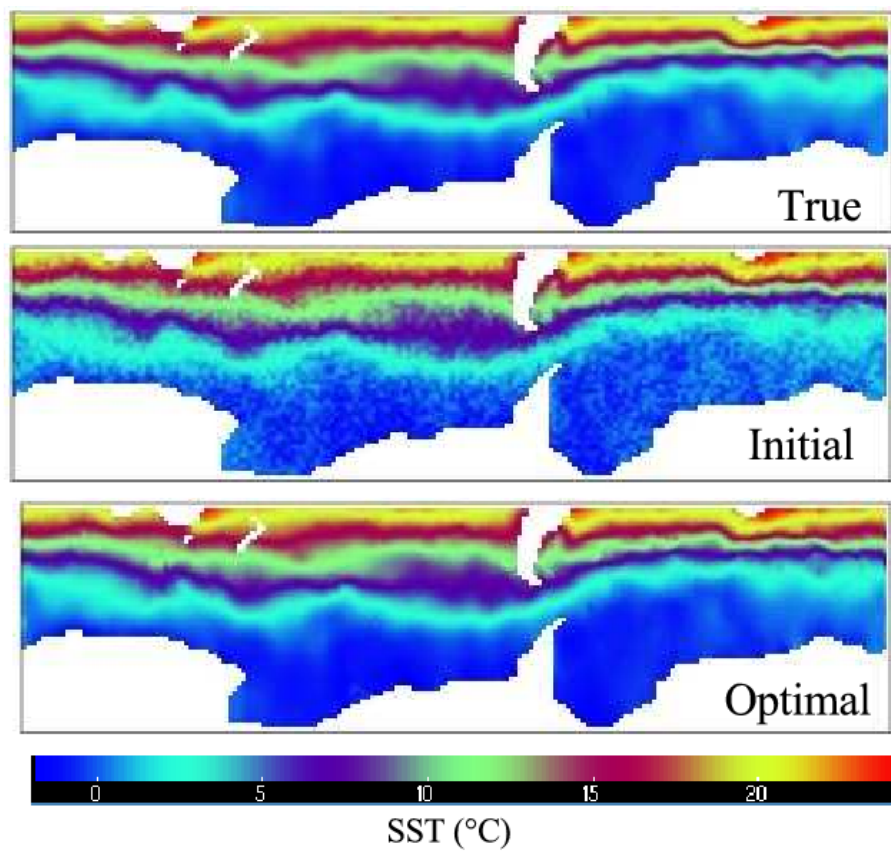


Figure 1. Twin experiment using the adjoint of OPA. We add random noise to a simulation of the ocean state around the Antarctic, and we remove this noise by minimizing the discrepancy with the physical model

4.4. Linearization

Simulating a complex system often requires solving a system of Partial Differential Equations. This is sometimes too expensive, in particular in the context of real time. When one wants to simulate the reaction of this complex system to small perturbations around a fixed set of parameters, there is a very efficient approximate solution: just suppose that the system is linear in a small neighborhood of the current set of parameters. The reaction of the system is thus approximated by a simple product of the variation of the parameters with the Jacobian matrix of the system. This Jacobian matrix can be obtained by AD. This is especially cheap when the Jacobian matrix is sparse. The simulation can be improved further by introducing higher-order derivatives, such as Taylor expansions, which can also be computed through AD. The result is often called a *reduced model*.

4.5. Mesh adaptation

Some approximation errors can be expressed by an adjoint state. Mesh adaptation can benefit from this. The classical optimization step can give an optimization direction not only for the control parameters, but also for the approximation parameters, and in particular the mesh geometry. The ultimate goal is to obtain optimal control parameters up to a precision prescribed in advance.

SELECT Project-Team

4. Application Domains

4.1. Introduction

A key goal of SELECT is to produce methodological contributions in statistics. For this reason, the SELECT team works with applications that serve as an important source of interesting practical problems and require innovative methodologies to address them. Most of our applications involve contracts with industrial partners, e.g. in reliability, although we also have several more academic collaborations, e.g. genomics, genetics, neuroimaging and ancient material imaging.

4.2. Curves classification

The field of classification for complex data as curves, functions, spectra and time series is important. Standard data analysis questions are being revisited to define new strategies that take the functional nature of the data into account. Functional data analysis addresses a variety of applied problems, including longitudinal studies, analysis of fMRI data and spectral calibration.

We are focusing on unsupervised classification. In addition to standard questions as the choice of the number of clusters, the norm for measuring the distance between two observations, and the vectors for representing clusters, we must also address a major computational problem. The functional nature of the data needs to be design efficient anytime algorithms.

4.3. Computer Experiments and Reliability

Since several years, SELECT has collaborations with EDF-DER *Maintenance des Risques Industriels* group. An important theme concerns the resolution of inverse problems using simulation tools to analyze uncertainty in highly complex physical systems. A collaboration on an analogous topic is developed with Dassault Aviation.

The other major theme concerns probabilistic modeling in fatigue analysis in the context of a research collaboration with SAFRAN an high-technology group (Aerospace propulsion, Aircraft equipment, Defense Security, Communications).

4.4. Neuroimaging

Since 2007 SELECT participates to a working group with team Neurospin (CEA-INSERM-Inria) on Classification, Statistics and fMRI (functional Magnetic Resonance Imaging) analysis. In this framework two theses have been co-supervised by SELECT and Neurospin researchers (Merlin Keller 2006-2009 and Vincent Michel 2007-2010). The aim of this research is to determine which parts of the brain are activated by different types of stimuli. A model selection approach is useful to avoid "false-positive" detections.

4.5. Analysis of genomic data

For the past few years SELECT has collaborated with Marie-Laure Martin-Magniette (URGV) for the analysis of genomic data. An important theme of this collaboration is using statistically sound model-based clustering methods to discover groups of co-expressed genes from microarray and high-throughput sequencing data. In particular, identifying biological entities that share similar profiles across several treatment conditions, such as co-expressed genes, may help identify groups of genes that are involved in the same biological processes.

4.6. Environment

A study has been achieved by Jean-Michel Poggi, Michel Misiti, Yves Misiti and Bruno Portier (INSA de Rouen), in the context of a collaboration between AirNormand, Orsay University and INSA of Rouen. Mixtures of linear regression models are used for the short-term statistical forecasting of the daily mean PM10 concentration in three cities in Haute-Normandie (France): Rouen, Le Havre and Dieppe. The Haute-Normandie region is located at northwest of Paris, near the south side of Manche sea and is heavily industrialized. Six monitoring stations reflecting the diversity of situations: urban background, traffic, rural and industrial stations are considered. This recent statistical method has been used and beyond the application, this study shed light on this method [35].

4.7. Analysis spectroscopic imaging of ancient materials

Ancient materials, encountered in archaeology, paleontology and cultural heritage, are often complex, heterogeneous and poorly characterised before their physico-chemical analysis. A technique of choice to gather as much physico-chemical information as possible is spectro-microscopy or spectral imaging where a full spectra, made of more than thousand samples, is measured for each pixel. The produced data is tensorial with two or three spatial dimensions and one or more spectral dimensions and it requires the combination of an «image» approach with «curve analysis» approach. Since 2010 SELECT collaborates with Serge Cohen (IPANEMA) on the development of conditional density estimation through GMM and non-asymptotic model selection to perform stochastic segmentation of such tensorial dataset. This technic enables the simultaneous accounting for spatial and spectral information while producing statistically sound information on morphological and physico-chemical aspects of the studied samples.

SEQUEL Project-Team

4. Application Domains

4.1. Introduction

SEQUEL aims at solving problems of prediction, as well as problems of optimal and adaptive control. As such, the application domains are very numerous.

The application domains have been organized as follows:

- adaptive control,
- signal processing and functional prediction,
- medical applications,
- web mining,
- computer games.

4.2. Adaptive Control

Adaptive control is an important application of the research being done in SEQUEL. Reinforcement learning (RL) precisely aims at controlling the behavior of systems and may be used in situations with more or less information available. Of course, the more information, the better, in which case methods of (approximate) dynamic programming may be used [69]. But, reinforcement learning may also handle situations where the dynamics of the system is unknown, situations where the system is partially observable, and non stationary situations. Indeed, in these cases, the behavior is learned by interacting with the environment and thus naturally adapts to the changes of the environment. Furthermore, the adaptive system may also take advantage of expert knowledge when available.

Clearly, the spectrum of potential applications is very wide: as far as an agent (a human, a robot, a virtual agent) has to take a decision, in particular in cases where he lacks some information to take the decision, this enters the scope of our activities. To exemplify the potential applications, let us cite:

- game softwares: in the 1990's, RL has been the basis of a very successful Backgammon program, TD-Gammon [75] that learned to play at an expert level by basically playing a very large amount of games against itself. Today, various games are studied with RL techniques.
- many optimization problems that are closely related to operation research, but taking into account the uncertainty, and the stochasticity of the environment: see the job-shop scheduling, or the cellular phone frequency allocation problems, resource allocation in general [69]
- we can also foresee that some progress may be made by using RL to design adaptive conversational agents, or system-level as well as application-level operating systems that adapt to their users habits.

More generally, these ideas fall into what adaptive control may bring to human beings, in making their life simpler, by being embedded in an environment that is made to help them, an idea phrased as "ambient intelligence".

- The sensor management problem consists in determining the best way to task several sensors when each sensor has many modes and search patterns. In the detection/tracking applications, the tasks assigned to a sensor management system are for instance:
 - detect targets,
 - track the targets in the case of a moving target and/or a smart target (a smart target can change its behavior when it detects that it is under analysis),
 - combine all the detections in order to track each moving target,
 - dynamically allocate the sensors in order to achieve the previous three tasks in an optimal way. The allocation of sensors, and their modes, thus defines the action space of the underlying Markov decision problem.

In the more general situation, some sensors may be localized at the same place while others are dispatched over a given volume. Tasking a sensor may include, at each moment, such choices as where to point and/or what mode to use. Tasking a group of sensors includes the tasking of each individual sensor but also the choice of collaborating sensors subgroups. Of course, the sensor management problem is related to an objective. In general, sensors must balance complex trade-offs between achieving mission goals such as detecting new targets, tracking existing targets, and identifying existing targets. The word “target” is used here in its most general meaning, and the potential applications are not restricted to military applications. Whatever the underlying application, the sensor management problem consists in choosing at each time an action within the set of available actions.

- sequential decision processes are also very well-known in economy. They may be used as a decision aid tool, to help in the design of social helps, or the implementation of plants (see [73], [72] for such applications).

4.3. Signal Processing

Applications of sequential learning in the field of signal processing are also very numerous. A signal is naturally sequential as it flows. It usually comes from the recording of the output of sensors but the recording of any sequence of numbers may be considered as a signal like the stock-exchange rates evolution with respect to time and/or place, the number of consumers at a mall entrance or the number of connections to a web site. Signal processing has several objectives: predict, estimate, remove noise, characterize or classify. The signal is often considered as sequential: we want to predict, estimate or classify a value (or a feature) at time t knowing the past values of the parameter of interest or past values of data related to this parameter. This is typically the case in estimation processes arising in dynamical systems.

Signals may be processed in several ways. One of the best-known way is the time-frequency analysis in which the frequencies of each signal are analyzed with respect to time. This concept has been generalized to the time-scale analysis obtained by a wavelet transform. Both analysis are based on the projection of the original signal onto a well-chosen function basis. Signal processing is also closely related to the probability field as the uncertainty inherent to many signals leads to consider them as stochastic processes: the Bayesian framework is actually one of the main frameworks within which signals are processed for many purposes. It is worth noting that Bayesian analysis can be used jointly with a time-frequency or a wavelet analysis. However, alternatives like belief functions came up these last years. Belief functions were introduced by Dempster few decades ago and have been successfully used in the few past years in fields where probability had, during many years, no alternatives like in classification. Belief functions can be viewed as a generalization of probabilities which can capture both imprecision and uncertainty. Belief functions are also closely related to data fusion.

4.4. Medical Applications

One of the initial motivations of the multi-arm bandit theory stems from clinical trials when one researches the effects of different treatments while maximizing the improvement of the patients' health states.

Medical health-care and in particular patient-management is up today one of the most important applications of the sequential decision making. This is because the treatment of the more complex health problems is typically sequential: A physician repeatedly observes the current state of the patient and makes the decision in order to improve the health condition as measured for example by *qualys* (quality adjusted life years).

Moreover, machine learning methods may be used for at least two means in neuroscience:

1. as in any other (experimental) scientific domain, the machine learning methods relying heavily on statistics, they may be used to analyse experimental data,
2. dealing with induction learning, that is the ability to generalize from facts which is an ability that is considered to be one of the basic components of “intelligence”, machine learning may be considered as a model of learning in living beings. In particular, the temporal difference methods for reinforcement learning has strong ties with various concepts of psychology (Thorndike's law of effect, and the Rescorla-Wagner law to name the two most well-known).

4.5. Web Mining

We work on the news/ad recommendation. These online learning algorithms reached a critical importance over the last few years due to these major applications. After designing a new algorithm, it is critical to be able to evaluate it without having to plug it into the real application in order to protect user experiences or/and the company's revenue. To do this, people used to build simulators of user behaviors and try to achieve good performances against it. However designing such a simulator is probably much more difficult than designing the algorithm itself! An other common way to evaluate is to not consider the exploration/exploitation dilemma (also known as "Cold Start" for recommender systems). Lately data-driven methods have been developed. We are working on building automatic replay methodology with some theoretical guarantees. This work also exhibits strong link with the choice of the number of contexts to use with recommender systems wrt your audience.

An other point is that web sites must forecast Web page views in order to plan computer resource allocation and estimate upcoming revenue and advertising growth. In this work, we focus on extracting trends and seasonal patterns from page view series. We investigate Holt-Winters/ARIMA like procedures and some regularized models for making short-term prediction (3-6 weeks) wrt to logged data of several big media websites. We work on some news event related webpages and we feel that kind of time series deserves a particular attention. Self-similarity is found to exist at multiple time scales of network traffic, and can be exploited for prediction. In particular, it is found that Web page views exhibit strong impulsive changes occasionally. The impulses cause large prediction errors long after their occurrences and can sometime be predicted (e.g., elections, sport events, editorial changes, holidays) in order to improve accuracies. It also seems that some promising model could arise from using global trends shift in the population.

4.6. Games

The problem of artificial intelligence in games consists in choosing actions of players in order to produce artificial opponents. Most games can be formalized as Markov decision problems, so they can be approached with reinforcement learning.

In particular, SEQUEL was a pioneer of Monte Carlo Tree Search, a technique that obtained spectacular successes in the game of Go. Other application domains include the game of poker and the Japanese card game of hanafuda.

SIERRA Project-Team

4. Application Domains

4.1. Application Domains

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: object recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing. In collaboration with Telecom Paris-tech.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG). In collaboration with the Parietal project-team.

SIMPAF Project-Team

4. Application Domains

4.1. Fluid mechanics

4.1.1. Numerical methods for viscous flows

We are concerned with systems of PDEs describing the evolution of mixture flows. The fluid is described by its density, its velocity and its pressure. These quantities obey mass and momentum conservation. On the one hand, when we deal with the 2D variable density incompressible Navier-Stokes equations, we aim at studying some instabilities phenomena such as the Raileigh-Taylor instability.

Furthermore, diffuse interface models have gained renewed interest in the last few years in fluid mechanics applications. From a physical viewpoint, they allow to describe phase transition phenomena. In this case, a specific stress tensor, introduced originally by Korteweg, must be added to the momentum equation.

If we use in addition the Fick law to relate the divergence of the velocity to derivatives of the density, one obtains the so-called KS model. In this case, the density of the mixture is naturally highly heterogeneous, and may model powder-snow avalanches or specific pollutants. Similar models also appear in combustion theory.

4.1.2. Control

Flow control strategies using passive or active devices are crucial tools in order to save energy in transports (especially for cars, trucks or planes), or to avoid the fatigue of some materials arising in a vast amount of applications. Nowadays, shape optimization needs to be completed by other original means, such as porous media located on the profiles, as well as vortex generator jets in order to drive active control.

4.2. Rubber elasticity

At the continuum level, rubber is modelled by an energy E defined as the integral over a domain D of \mathbb{R}^d of some energy density W depending only locally on the gradient of the deformation u : $E(u) = \int_D W(\nabla u(x)) dx$. At the microscopic level (say 100nm), rubber is a network of cross-linked and entangled polymer chains (each chain is made of a sequence of monomers). At this scale the physics of polymer chains is well-understood in terms of statistical mechanics: monomers thermally fluctuate according to the Boltzmann distribution [61]. The associated Hamiltonian of a network is typically given by a contribution of the polymer chains (using self-avoiding random bridges) and a contribution due to steric effects (rubber is packed and monomers are surrounded by an excluded volume). The main challenge is to understand how this statistical physics picture yields rubber elasticity, and how this can be taken advantage of in order to devise physically-based constitutive laws for rubber. A rigorous derivation can be performed using tools of stochastic homogenization. The challenge is now to devise a numerical method to pass from polymer physics to an explicit continuum constitutive law, and to apply this approach to physically relevant models.

4.3. Maxwell equations

The a posteriori error estimators developed for the Maxwell equations are very useful tools. They are implemented in the software "Carmel-3D" (see the softwares section). This numerical code is used in order to study some original applications, like electrical machines or specific actuators. It is also devoted to nondestructive control by the use of Foucault currents, to the simulation of devices using magnetic fluids or of induced currents in human bodies.

4.4. Corrosion models

The application domains related to corrosion models are the following.

4.4.1. Corrosion modelling of iron based alloy in nuclear waste repository

The concept for long term storage of high-level radioactive waste in France is based on an underground repository. The waste shall be confined in a glass matrix and then placed into cylindrical steel canisters. These containers shall be placed into micro-tunnels in the highly impermeable Callovo-Oxfordian claystone layer at a depth of several hundred meters. At the request of the French nuclear waste management agency ANDRA, investigations are conducted to optimize and finalize this repository concept with the aim to ensure its long-term safety and its reversibility. The long-term safety assessment of the geological repository has to take into account the degradation of the carbon steel used for the waste overpacks and the cell disposal liners, which are in contact with the claystone formation. This degradation is mainly caused by generalized corrosion processes which form a passive layer on the metal surface consisting of a dense oxide inner layer and a porous hydroxide outer layer in contact with the groundwater in the pore space of the claystones. The processes take place under anaerobic conditions, since the groundwater is anoxic.

As a tool to investigate the corrosion processes at the surface of the carbon steel canisters, the Diffusion Poisson Coupled Model (DPCM) for corrosion has been developed by Bataillon *et al.* [47]. The numerical approximation of this corrosion model by accurate and efficient methods is challenging.

4.4.2. Corrosion modelling of Ni-base alloys in Pressurized Water Reactor primary water

The understanding of the oxidation behavior of Ni-base alloys in PWR primary water is of major importance due to the cations released due to corrosion of the steam generators which is a source of the radioactivity of the primary circuit. Moreover, the oxidation process is the reason of the initiation of intergranular stress corrosion cracking in some alloys. A numerical model, called EKINOX (Estimation KINetics OXYdation), has been developed at CEA [48] in order to simulate the oxide scale growth. This model should be able to calculate the evolutions of concentration profiles of the species and of their point defects in the oxide and in the substrate. Numerical experiments have shown the limits of this existing numerical model, especially the need of very small time steps for the computations ; a macroscopic model should be developed and numerical methods proposed for its simulation.

4.5. Plasmas

In the last decade, satellite-positioning devices have become one of the most powerful navigation assistance systems. Recently, Europe launched the first satellites of Galileo, a competitor to the constellation of satellites Navstar GPS American system. In order to guarantee the integrity service level, it is fundamental to take into account the various physical phenomena that can affect the mission and to identify all the potential sources of system unavailability. One of the main sources of data unavailability that has been identified is the phenomenon of ionospheric scintillations. Indeed scintillation causes radio frequency signal amplitude to fade and phase variations as satellite signals pass through the ionosphere. Such effects may induce loss of lock or cycle slips on ranging signals broadcast by Galileo satellites making them totally useless for accurate integrity information determination.

4.6. Optical fibers

When designing optical fibers, the question of the envelope dynamics of laser impulses is crucial because its answer decides how much information the fiber will actually be able to carry. Recent progresses in the optical fiber design now allow to choose specific physical parameters very precisely, and the main limitation for being able to design new and more efficient optical fibers is the understanding of the effect of specific conjunctions of parameter values. In other terms, a main issue is the understanding of the influence of several parameters on specific solutions of nonlinear dispersive equations of Schrödinger type.

TAO Project-Team

4. Application Domains

4.1. Energy Management

Energy management, our priority application field, involves sequential decision making with:

- stochastic uncertainties (typically weather);
- both high scale combinatorial problems (as induced by nuclear power plants) and non-linear effects;
- high dimension (including hundreds of hydroelectric stocks);
- multiple time scales:
 - minutes (dispatching, ensuring the stability of the grid), essentially beyond the scope of our work, but introducing constraints for our time scales;
 - days (unit commitment, taking care of compromises between various power plants);
 - years, for evaluating marginal costs of long term stocks (typically hydroelectric stocks);
 - tenths of years, for investments.

Nice challenges also include:

- spatial distribution of problems; due to capacity limits we can not consider a power grid like Europe + North Africa as a single “production = demand” constraint; with extra connections we can equilibrate excess production by renewables for remote areas, but not in an unlimited manner.
- other uncertainties, which might be modeled by adversarial or stochastic frameworks (e.g. technological breakthroughs, decisions about ecological penalization).

We have several related projects (Citines, a European (FP7) project; in the near future we should start the Post project (Ademe); IOMCA, a ANR project). We have a collaboration with a company, Artelys, working on optimization in general, and in particular on energy management.

Technical challenges: Our work focuses on the combination of reinforcement learning tools, with their anytime behavior and asymptotic guarantees, with existing fast approximate algorithms; see 6.4 . Our goal is to extend the state of the art by taking into account non-linearities which are often neglected in power systems due to the huge computational cost.

Related Activities:

- We are in the process of creating a Franco-Taiwanese company (maybe a taiwanese company using French software) for energy optimization in Taiwan.
- We have a joint team with Taiwan, namely the Indema associate team (see Section 8.4.1.1).
- We have a “Ilab” in progress with Artelys (see Section 5.1) for industrialization of our work.
- We organized various forums and meetings around Energy Management.

4.2. Air Traffic Control

Air Traffic Control has been an application field of Marc Schoenauer’s work in the late 90s (PhD theses of F. Médioni in 98 and S. Oussedik in 2000). It was revived recently with Gaëtan Marceau-Caron’s CIFRE PhD together with Thalès Air Systems (Areskin HAdjaz) and Thalès TRT (Pierre Savéant), around global optimization of the traffic in order to increase the capacity of the airspace without overloading the controllers. First results concern the on-line prediction of individual plane trajectories [42]: these results demonstrate that there is no hope to ever predict the global behavior of the traffic by considering planes individually. A new formulation of the problem is hence considered, where plane flows are modeled with Bayesian Networks [42]. On-going work is concerned with the simulation and optimization of air traffic based on this model.

TOSCA Project-Team

4. Application Domains

4.1. Application Domains

TOSCA is interested in developing stochastic models and probabilistic numerical methods. Our present motivations come from Finance, Neuroscience and Biology, Fluid Mechanics and Meteorology, Chemical Kinetics, Diffusions in random media, Transverse problems, Software and Numerical experiments.

Finance For a long time now TOSCA has collaborated with researchers and practitioners in various financial institutions and insurance companies. We are particularly interested in calibration problems, risk analysis (especially model risk analysis), optimal portfolio management, Monte Carlo methods for option pricing and risk analysis, asset and liabilities management. We also work on the partial differential equations related to financial issues, for example the stochastic control Hamilton–Jacobi–Bellman equations. We study existence, uniqueness, qualitative properties and appropriate deterministic or probabilistic numerical methods. At the moment we pay special attention to the financial consequences induced by modelling errors and calibration errors on hedging strategies and portfolio management strategies.

Neuroscience and Biology The interest of TOSCA in biology is developing in three main directions: neuroscience, molecular dynamics and population dynamics. In neuroscience, stochastic methods are developed to analyze stochastic resonance effects, to solve inverse problems and to investigate mean-field/McKean-Vlasov equations. For example, we are studying probabilistic interpretations and Monte Carlo methods for divergence form second-order differential operators with discontinuous coefficients, motivated by the 3D MEG inverse problem. Our research in molecular dynamics focuses on the development of Monte Carlo methods for the Poisson-Boltzmann equation which also involves a divergence form operator, and of original algorithms to construct improved simulation techniques for protein folding or interaction. Finally, our interest in population dynamics comes from ecology, evolution and genetics. For example, we are studying the emergence of diversity through the phenomenon of evolutionary branching in adaptive dynamics. Some collaborations in biostatistics on cancer problems are also being initiated.

Fluid Mechanics and Meteorology In Fluid Mechanics we develop probabilistic methods to solve vanishing vorticity problems and to study the behavior of complex flows at the boundary, and their interaction with the boundary. We elaborate and analyze stochastic particle algorithms. Our studies concern the convergence analysis of these methods on theoretical test cases and the design of original schemes for applicative cases. A first example concerns the micro-macro model of polymeric fluids (the FENE model). A second example concerns Pope’s Lagrangian modelling of turbulent flows, motivated by the problem of modelling and computing characteristic properties of the local wind activity in areas where windmills are built. Our goal is to estimate local energy resources which are subject to meteorological randomness by combining large scale wind models and small scale Monte Carlo techniques, and to simulate management strategies of wind resources.

Chemical Kinetics The TOSCA team is studying coagulation and fragmentation models, that have numerous areas of applications (polymerization, aerosols, cement industry, copper industry, population dynamics...). Our current motivation comes from the industrial copper crushers in Chile. We aim to model and calibrate the process of fragmentation of brass particles of copper in industrial crushers, in order to improve their efficiency at a low cost.

Diffusions in random media A *random medium* is a material with a lot of heterogeneity which can be described only statistically. Typical examples are fissured porous media within rocks of different types, turbulent fluids or unknown or deficient materials in which polymers evolve or waves propagate. For the last few years, the TOSCA team has been collaborating with the Geophysics

community on problems related to underground diffusions, especially those which concern waste transport or oil extraction. We are extending our previous results on the simulation of diffusion processes generated by divergence form operators with discontinuous coefficients. Such an operator appears for example in the Darcy law for the behavior of a fluid in a porous media. We are also developing another class of Monte Carlo methods to simulate diffusion phenomena in discontinuous media.

Transverse problems Several of the topics of interest of TOSCA do not only concern a single area of application. This is the case in particular for long time simulation methods of nonlinear McKean-Vlasov PDEs, the problem of simulation of multivalued models, variance reduction techniques or stochastic partial differential equations. For example, multivalued processes have applications in random mechanics or neuroscience, and variance reduction techniques have applications in any situation where Monte Carlo methods are applicable.

Software, numerical experiments TOSCA is interested in designing algorithms of resolution of specific equations in accordance with the needs of practitioners. We benefit from our strong experience of the programming of probabilistic algorithms of various architectures including intensive computation architectures. In particular, our activity will concern the development of grid computing techniques to solve large dimensional problems in Finance. We are also interested in intensively comparing various Monte Carlo methods for PDEs and in the development of open source libraries for our numerical methods in Fluid Mechanics, Meteorology, MEG or Chemical Kinetics.

ABS Project-Team (section vide)

AMIB Project-Team (section vide)

ASCLEPIOS Project-Team (section vide)

ATHENA Project-Team

4. Application Domains

4.1. Applications of Diffusion MRI

Various examples of CNS diseases as Alzheimer's and Parkinson's diseases and others like multiple sclerosis, traumatic brain injury and schizophrenia have characteristic abnormalities in the micro-structure of brain tissues that are not apparent and cannot be revealed reliably by standard imaging techniques. Diffusion MRI can make visible these co-lateral damages to the fibers of the CNS white matter that connect different brain regions. This is why in our research, Diffusion MRI is the major anatomical imaging modality that will be considered to recover the CNS connectivity.

Clinical domain: Diagnosis of neurological disorder

- *Parkinson's and Alzheimer's diseases* are among the most important CNS diseases. Six million patients (among which 850.000 in France) are suffering from Alzheimer's, making it the most important neurodegenerative disease in Europe. Over 85 years of age, 1 woman in 4 and 1 man in 5 are affected in Europe. In France, the number of Alzheimer's patients is expected to reach at least 2 million in 2025 and will probably double in 2050, with the increasing age of the population. Parkinson's disease is the second most important neurodegenerative disease. There are six and a half million patients in the world and roughly 150.000 patients in France, among which 10% are under 40 and 50% over 58. Together with our partners from NeuroSpin (Saclay), Inserm U678 and CENIR (CHUPS, Paris), we are involved in the ANR project NucleiPark which is about high field MRI of the brainstem, the deep nuclei and their connections in the Parkinsonian syndromes.
- *Spinal Cord Injury* (SCI) has a significant impact on the quality of life since it can lead to motor deficits (paralysis) and sensory deficits. In the world, about 2.5 million people live with SCI (<http://www.campaignforcure.org>). To date, there is no consensus for full rehabilitative cure in SCI, although many therapeutic approaches have shown benefits [69], [71]. It is thus of great importance to develop tools that will improve the characterization of spinal lesions as well as the integrity of remaining spinal tracts to eventually establish better prognosis after spinal injury. We have already started to be active in this domain with our collaborators at Inserm U678 (H. Benali) and CRSN/Faculté de médecine Université de Montréal (Pr. S. Rossignol).

4.2. Applications of M/EEG

Applications of EEG and MEG cover:

- **Clinical domain:** diagnosis of neurological disorders such as
 - Diagnosis of neurological disorders such as epilepsy, schizophrenia, tinnitus, ...
 - Presurgical planning of brain surgery.
- **Cognitive research** aims at better understanding the brain spatio-temporal organisation.
- **Brain Computer Interfaces** look at allowing a direct control of the world using brain signal such as EEG signals. Those can be considered both as an application of EEG processing techniques and as a tool for fundamental and applied research as it opens the way for more dynamical and active brain cognitive protocols.

The dream of all M/EEG researchers is to alleviate the need for invasive recordings (electrocorticograms or intracerebral electrodes), which are often necessary prior to brain surgery, in order to precisely locate both pathological and vital functional areas. We are involved in this quest, particularly through our collaboration with the La Timone hospital in Marseille. M/EEG are also used in **cognitive research**, and we collaborate with the *Laboratory for Neurobiology of Cognition* in order to develop methods that suit their needs for sophisticated data analysis.

BAMBOO Project-Team

4. Application Domains

4.1. Biology with a focus on symbiosis

The main area of application of BAMBOO is biology, with a special focus on symbiosis (ERC project) and on intracellular interactions.

BANG Project-Team

4. Application Domains

4.1. Biology and medicine

The team is mostly involved in applications to biology and medicine. More precisely it aims at understanding biophysical mechanisms that sustain cell proliferation or malfunction. The main examples are biopolymers size repartition, cell self-organisation, tissue growth and cancer development or treatment.

4.2. Geophysical flows and environment

The team will split and give rise to another team ANGE specialised in complex geophysical flows in interaction with environment. Free surface flows as tsunamis, flows in river and coastal areas and their ecological consequences are typical examples of applications developed in the team based on algorithms for the free-surface Navier-Stokes equations.

BEAGLE Team

4. Application Domains

4.1. Application Domains

- Molecular and cellular biology
- Genome evolution

BIGS Project-Team

4. Application Domains

4.1. Data analysis and local regression

Our expertise in data analysis and advanced statistics methods has given rise to a wide number of interdisciplinary collaborations. Among those, here are the most challenging at a scientific level:

(i) *Health inequalities*: We have recently developed a statistical procedure in order to create a neighborhood socioeconomic index and investigate its influence on health inequalities. The study setting is composed with 3 major French metropolitan areas (Lille, Lyon and Marseille), and we collaborate for this project with a medical team at EHESP (Ecole des Hautes Etudes en Santé Publique) lead by D. Zmirou (see [19] for further details).

(ii) *Fetal pathology*: An ongoing work concerning local regression techniques is related to Fetal Biometry, an investigation line suggested by a collaboration between our team and the *Centre de Placentologie et Foetopathologie de la Maternité Régionale de Nancy*, under the direction of Professor Bernard Foliguet. The methods involved in Fetal Biometry are usually based on the comparison of some measured values with the predicted values derived from reference charts or equations in a normal population. However, it happens that maternal and pregnancy characteristics have a significant influence on in-utero Fetal Biometry. We will thus produce some models allowing to construct customized fetal biometric size charts. In order to evaluate them, classical and polynomial regression can be used, but they are not the most appropriate to the kind of data we have to handle. Hence, we plan to use local regression estimation in order to perform such an evaluation.

(iii) *Cohorts analysis*: Some medical teams in Nancy are faced with an overwhelming amount of data, for which a serious statistical assessment is needed. Among those let us mention the INSERM team of Pr. Jean-Louis Guéant and the Inria team Orpailleur (particularly with Marie-Dominique Desvignes and Malika Smail). The goal of this collaboration is to extract biological markers for different diseases (cognitive decline; inflammatory intestinal diseases; liver cancer). To this aim, the INSERM team provides us with several data cohorts with a high number of variables and subjects. As in many instances in Biostatistics, one is then faced with a very high dimensional data, from which we hope to extract a reduced number of significant variables allowing to predict the cardiovascular risk accurately. Moreover, these characters should be meaningful to practitioners. The objective for us is thus to design an appropriate variable selection, plus a classification procedure in this demanding context. Let us highlight an original feature of this collaboration: it combines our own data analysis techniques with those developed by the Orpailleur team, based on symbolic tools. We hope that this experience will enrich both points of view and give rise to new methods of data analysis.

4.2. Estimation for complex and biological systems

Our main application for this line of investigation is the photodynamic therapy developed by T. Bastogne. We shall also focus on bacteriophage therapies and subdiffusion within molecules.

(i) *Photodynamic therapy*. One of the main applications we have in mind for our identification problems is to model photodynamic therapy. This promising cancer treatment involves selective uptake and retention of a photosensitive drug in a tumor, followed by irradiation with light at an appropriate wavelength. Photosensitizers are photoactive compounds such as for instance porphyrins and chlorins. The activated photosensitizer is thought to produce singlet oxygen at high doses and thereby to initiate apoptotic and necrotic death of tumor. Due to the lack of response reproducibility, the complexity of interactions between physical, chemical and biological aspects and the high cost of experiments, there is a real demand in good mathematical and physical models which might help to better control and understand PDT responses. We are particularly concerned with modeling the drug uptake into cancer cells, the photoreactions induced by light exposition and tumor growth kinetics.

(ii) *Bacteriophage systems.* A collaboration between our team, the Mathematics and the Genetics and Microbiology Departments at the *Universitat Autònoma de Barcelona* (UAB) is being set up, focusing on probabilistic aspects of bacteriophage therapies for animal diseases like hemorrhagic septicemia in cattle or atrophic rhinitis in swine. This kind of therapy consists in inoculating a (benign) virus to animals in order to kill the bacteria known to be responsible of the disease. It was in use in the Soviet Union until the 80s, and is now re-emerging, still at an experimental level, due to the progressive slowdown in antibiotic efficiency.

Within this context, our analysis of a noisy predator-prey competition modeling the treatment helps to calibrate and to understand better the behavior of the system in terms of fluctuations around an equilibrium. Note that our preliminary contacts with the Genetics and Microbiology Departments at UAB also open the way to a particle model in order to represent the couple bacteria/virus living on a surface.

(iii) *Subdiffusion into molecules.* Our purpose here is a better understanding of the phenomena observed in nanoscale Biophysics, as explained in the series of papers [52]. The technological advances in nanoscale technologies allow the observation of single molecules, and thus the description of newly observed phenomenon. A typical example of this new kind of observation is given by the fluctuations in the folding of a protein-enzyme compound called *Fre*, which is involved in the DNA synthesis of the (canonical) bacterium *E. Coli*.

More specifically, the paper [52] advocates for modeling this folding fluctuations by means of a Volterra type equation driven by a fractional Brownian motion. This convincing model is based on some experimental and physical evidences, and have also been observed in a wide number of recent biological experiments. However, the model exhibited in [52] also raises some unsolved questions: some stochastic equations appearing in the models are not properly defined and their long time behavior is still mysterious. The lack of a method in order to simulate and estimate coefficients of these equations on a solid mathematical ground should also be mentioned. This is the kind of topic we wish to address, for which a preliminary contact with S. Kou and N. Pillai (Princeton University, USA) has been established.

(iv) *Osteoporosis.* During the year 2011-2012, C. Lacaux has been visiting the MAP 5 (Paris Descartes University) laboratory and joined the ANR Project MATAIM (Modèles Anisotropes de Textures. Applications à l'Imagerie Médicale). This project, which involves both mathematicians and practitioners, is in particular interested in the osteoporosis diagnostic. The paper [34] is a first step in the direction of modeling trabecular bone x-ray images by some operator scaling fields. Actually the estimation of the matrix, which characterizes the anisotropy of the model, is crucial for practical purposes. Hermine Biermé (Paris Descartes University) and Céline Lacaux are working on this problem using quadratic variations. Once the problem of estimation is solved, they plan a comparison of the theoretical model with real data provided by our Biologist colleagues of the MATAIM project. If the model corresponds to real data (as suggested in [34]), this approach may help for the diagnostic of osteoporosis: a numerical study has to be performed in order to find the parameter value which characterizes osteoporosis.

BIOCORE Project-Team

4. Application Domains

4.1. Bioenergy

Finding sources of renewable energy is a key challenge for our society. We contribute to this topic through two main domains for which a strong and acknowledged expertise has been acquired over the years. First, we consider anaerobic digesters, the field of expertise of the LBE members of the team, for the production of methane and/or biohydrogen from organic wastes. The main difficulty is to make these processes more reliable and exploit more efficiently the produced biogas by regulating both its quality and quantity despite high variability in the influent wastes. One of the specific applications that needs to be tackled is the production of biogas in a plant when the incoming organic waste results from the mixing of a finite number of substrates. The development of control laws that optimize the input mix of the substrates as a function of the actual state of the system is a key challenge for the viability of this industry.

The second topic consists in growing microalgae, the field of expertise of the LOV members of the team, to produce biofuel. These microorganisms can synthesize lipids with a much higher productivity than terrestrial oleaginous species. The difficulty is to better understand the involved processes, which are mainly transient, to stimulate and optimize them on the basis of modeling and control strategies. Predicting and optimizing the productivity reached by these promising systems in conditions where light received by each cell is strongly related to hydrodynamics, is a crucial challenge.

Finally, for the energy balance of the process, it is important to couple microalgae and anaerobic digestion to optimize the solar energy that can be recovered from microalgae, as was explored within the [ANR Symbiose](#) project (2009-2012) [72].

4.2. CO₂ fixation and fluxes

Phytoplanktonic species, which assimilate CO₂ during photosynthesis, have received a lot of attention in the last years. Microalgal based processes have been developed in order to mitigate industrial CO₂. As for biofuel productions, many problems arise when dealing with microalgae which are more complex than bacteria or yeasts. Several models have been developed within our team to predict the CO₂ uptake in conditions of variable light and nitrogen availability. The first modeling challenge in that context consists in taking temperature effects and light gradient into account.

The second challenge consists in exploiting the microalgal bioreactors which have been developed in the framework of the quantification of carbon fluxes between ocean and atmospheres. The SEMPO platform (simulator of variable environment computer controlled), developed within the LOV team, has been designed to reproduce natural conditions that can take place in the sea and to accurately measure the cells behavior. This platform, for which our team has developed models and control methods over the years, is an original and unique tool to develop relevant models which stay valid in dynamic conditions. It is worth noting that a better knowledge of the photosynthetic mechanisms and improved photosynthesis models will benefit both thematics: CO₂ mitigation and carbon fluxes predictions in the sea.

4.3. Biological control for plants and micro-plants production systems

This work concentrates on the protection of cultures of photosynthetic organisms against their pests or their competitors. The forms of cultures that we study are crop and micro-algae productions. In both cases, the devices are more or less open to the outside, depending on the application (greenhouse/field, photobioreactor/raceway) so that they may give access to invading species which can be harmful to the cultures; we opt for protecting the culture through the use of biocontrol agents which are, generically, natural enemies of these noxious populations [9].

In crop production, biological control is indeed a very promising alternative to pesticide usage; the use of predators, parasitoids or pathogens of crop pests in order to fight them has many advantages with respect to environmental protection, health of the consumers and the producers, the limited development of resistance (compared to chemicals),... It is however not widespread yet because it often lacks efficiency in real-life crop production systems (while its efficiency in the laboratory is much higher) and can fail to be economically competitive. Our objective is to propose models that would help to explain which factors are locks that prevent the smooth transition from the laboratory to the agricultural crop as well as develop new methods for the optimal deployment of the pests natural enemies.

Microalgae production is faced with exactly the same problems since predators of the produced microalgae (e.g. zooplankton) or simply other species of microalgae can invade the photobioreactors and outcompete or eradicate the one that we wish to produce. Methods need therefore to be proposed for fighting the invading species; this could be done by introducing predators of the pest and so keeping it under control, or by controlling the conditions of culture in order to reduce the possibility of invasion; the design of such methods could greatly take advantage of our knowledge developed in crop protection since the problems and models are related.

4.4. Biological depollution

These works will be carried out with the LBE , mainly on anaerobic treatment plants. This process, despite its strong advantages (methane production and reduced sludge production) can have several locally stable equilibria. In this sense, proposing reliable strategies to stabilize and optimise this process is a key issue. Because of the recent (re)development of anaerobic digestion, it is crucial to propose validated supervision algorithms for this technology. A problem of growing importance is to take benefit of various waste sources in order to adapt the substrate quality to the bacterial biomass activity and finally optimize the process. This generates new research topics for designing strategies to manage the fluxes of the various substrate sources meeting at the same time the depollution norms and providing a biogas of constant quality. In the past years, we have developed models of increasing complexity. However there is a key step that must be considered in the future: how to integrate the knowledge of the metabolisms in such models which represent the evolution of several hundreds bacterial species? How to improve the models integrating this two dimensional levels of complexity? With this perspective, we wish to better represent the competition between the bacterial species, and drive this competition in order to maintain, in the process, the species with the highest depollution capability. This approach, initiated in [103] must be extended from a theoretical point of view and validated experimentally.

4.5. Experimental Platforms

To test and validate our approach, we use experimental platforms developed by our partner teams; these are highly instrumented for accurately monitoring the state of biological species:

- A photobioreactor (SEMPO) for Lagrangian simulation of the dynamical environment of marine microalgae (LOV) with computer controlled automata for high frequency measurement and on-line control. This photobioreactor is managed by Amélie Talec.
- Several pilot anaerobic digesters that are highly instrumented and computerized and the algotron, that is the coupling of a digester and a photobioreactor for microalgae production (LBE). Eric Latrille is our main contact for this platform at LBE.
- Experimental greenhouses of various sizes (from laboratory to semi-industrial size) and small scale devices for insect behavior testing at ISA.

Moreover, we may use the data given by several experimental devices at EPI IBIS/ Hans Geiselmann Laboratory (University J. Fourier, Grenoble) for microbial genomics.

BONSAI Project-Team

4. Application Domains

4.1. Sequence processing for Next Generation Sequencing

As said in the introduction of this document, biological sequence analysis is a foundation subject for the team. In the last years, sequencing techniques experienced remarkable advances with NGS, that allows for fast and low-cost acquisition of huge amounts of sequence data, and outperforms conventional sequencing methods. These technologies can apply to genomics, with DNA sequencing, as well as to transcriptomics, with RNA sequencing allowing to gene expression analysis. They promise to address a broad range of applications including: Comparative genomics, individual genomics, high-throughput SNP detection, identifying small RNAs, identifying mutant genes in disease pathways, profiling transcriptomes for organisms where little information is available, researching lowly expressed genes, studying the biodiversity in metagenomics. From a computational point of view, NGS gives rise to new problems and gives new insight on old problems by revisiting them: Accurate and efficient remapping, pre-assembling, fast and accurate search of non exact but quality labelled reads, functional annotation of reads, ...

4.2. Noncoding RNA

Our expertise in sequence analysis also applies to noncodingRNA analysis. Noncoding RNA genes play a key role in many cellular processes. First examples were given by microRNAs (miRNAs) that were initially found to regulate development in *C. elegans*, or small nucleolar RNAs (snoRNAs) that guide chemical modifications of other RNAs in mammals. Hundreds of miRNAs are estimated to be present in the human genome, and computational analysis suggests that more than 20% of human genes are regulated by miRNAs. To go further in this direction, the 2007 ENCODE Pilot Project provides convincing evidence that the Human genome is pervasively transcribed, and that a large part of this transcriptional output does not appear to encode proteins. All those observations open a universe of “RNA dark matter” that must be explored. From a combinatorial point of view, noncoding RNAs are complex objects. They are single stranded nucleic acids sequences that can fold forming long-range base pairings. This implies that RNA structures are usually modelled by complex combinatorial objects, such as ordered labeled trees, graphs or arc-annotated sequences.

4.3. Genome structures

Our third application domain is concerned with the structural organization of genomes. Genome rearrangements are able to change genome architecture by modifying the order of genes or genomic fragments. The first studies were based onto linkage maps and mathematical models appeared fifteen years ago. But the usage of computational tools was still limited because of lack of data. The increasing availability of complete and partial genomes now offers an unprecedented opportunity to analyse genome rearrangements in a systematic way and gives rise to a wide spectrum of problems: Taking into account several kinds of evolutionary events, looking for evolutionary paths conserving common structure of genomes, dealing with duplicated content, being able to analyse large sets of genomes even at the intraspecific level, computing ancestral genomes and paths transforming these genomes into several descendant genomes.

4.4. Nonribosomal peptides

Lastly, the team has been developing for several year a tight collaboration with Probiogem lab on nonribosomal peptides, and has became a leader on that topic. Nonribosomal peptide synthesis produces small peptides not going through the central dogma. As the name suggests, this synthesis uses neither messenger RNA nor ribosome but huge enzymatic complexes called nonribosomal peptide synthetases (NRPSs). This alternative pathway is found typically in bacteria and fungi. It has been described for the first time in the 70's [21]. For the last decade, the interest in nonribosomal peptides and their synthetases has considerably increased, as witnessed by the growing number of publications in this field. These peptides are or can be used in many biotechnological and pharmaceutical applications (e.g. anti-tumors, antibiotics, immuno-modulators).

CARMEN Team

4. Application Domains

4.1. Scientific context: the Liryc

Our fields of application are naturally: electrophysiology and cardiac physiopathology at the tissue scale on one side; medical and clinical cardiology on the other side.

The team's research project is part of the IHU Liryc project, initiated by Pr. M. Haissaguerre. It is concerned by the major issues of modern electrocardiology: atrial arrhythmias, sudden death due to ventricular fibrillation and heart failure related to ventricular dyssynchrony.

We aim at bringing applied mathematics and scientific computing closer to biomedical research applied to cardiac rhythmology and clinical cardiology. It aims at enhancing our fundamental knowledge of the normal and abnormal cardiac electrical activity, of the patterns of the electrocardiogram; and we will develop new simulation tools for training, biological and clinical applications.

4.2. Basic experimental electrophysiology

Our modeling is carried out in coordination with the experimental teams from the Liryc. It will help to write new concepts concerning the multiscale organisation of the cardiac action potentials and will serve our understanding in many electrical pathologies:

- At the atrial level, we apply our models to understand the mechanisms of complex arrhythmias and the relation with the heterogeneities at the insertion of the pulmonary vein.
- At the ventricula level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles and (2) modeling the structural heterogeneities at the cellular scale, taking into account the complex organisation and disorganisation of the myocytes and fibroblasts. Point (1) is supposed to play a major role in sudden cardiac death and point (2) is important in the study of infarct scars for instance.

4.3. Cardiac electrical signals

The Liryc use, on a daily basis and in the clinical context, complex electrical imaging systems, like intracardiac catheters and the CardioInsight vest with 252 body surface electrodes.

The numerical models can guide the analysis of these signals and conversely, the models can be guided by the signals.

Other applied questions can be addressed by modeling, like the nature of the various electrical signals measured by catheters, that heavily depends on the nature and spatial localisation of the electrodes.

CLIME Project-Team

4. Application Domains

4.1. Introduction

The central application domain of the project-team is atmospheric chemistry. We develop and maintain the air quality modeling system Polyphemus, which includes several numerical models (Gaussian models, Lagrangian model, two 3D Eulerian models including Polair3D) and their adjoints, and different high level methods: ensemble forecast, sequential and variational data assimilation algorithms. Advanced data assimilation methods, network design, inverse modeling, ensemble forecast are studied in the context of air chemistry. Note that addressing these high level issues requires controlling the full software chain (models and data assimilation algorithms).

The activity on assimilation of satellite data is mainly carried out for meteorology and oceanography. This is addressed in cooperation with external partners who provide numerical models. Concerning oceanography, the aim is to improve the forecast of ocean circulation, by assimilation of fronts and vortices displayed on image data. Concerning meteorology, the focus is on correcting the model location of structures related to high-impact weather events (cyclones, convective storms, *etc.*) by assimilating images.

4.2. Air quality

Air quality modeling implies studying the interactions between meteorology and atmospheric chemistry in the various phases of matter, which leads to the development of highly complex models. The different usages of these models comprise operational forecasting, case studies, impact studies, *etc.*, with both societal (e.g., public information on pollution forecast) and economical impacts (e.g., impact studies for dangerous industrial sites). Models lack some appropriate data, for instance better emissions, to perform an accurate forecast and data assimilation techniques are recognized as a major key point for the improvement of forecast's quality.

In this context, Clime is interested in various problems, the following being the crucial ones:

- The development of ensemble forecast methods for estimating the quality of the prediction, in relation with the quality of the model and the observations. Sensitivity analysis with respect to the model's parameters so as to identify physical and chemical processes, whose modeling must be improved.
- The development of methodologies for sequential aggregation of ensemble simulations. What ensembles should be generated for that purpose, how spatialized forecasts can be generated with aggregation, how can the different approaches be coupled with data assimilation?
- The definition of second-order data assimilation methods for the design of optimal observation networks. Management of combinations of sensor types and deployment modes. Dynamic management of mobile sensors' trajectories.
- How to estimate the emission rate of an accidental release of a pollutant, using observations and a dispersion model (from the near-field to the continental scale)? How to optimally predict the evolution of a plume? Hence, how to help people in charge of risk evaluation for the population?
- The definition of non-Gaussian approaches for data assimilation.
- The assimilation of satellite measurements of troposphere chemistry.

The activities of Clime in air quality are supported by the development of the Polyphemus air quality modeling system. This system has a modular design, which makes it easier to manage high level applications such as inverse modeling, data assimilation and ensemble forecast.

4.3. Oceanography

The capacity of performing a high quality forecast of the state of the ocean, from the regional to the global scales, is of major interest. Such a forecast can only be obtained by systematically coupling numerical models and observations (*in situ* and satellite data). In this context, being able to assimilate image structures becomes a key point. Examples of such image structures are:

- apparent motion linked to surface velocity;
- trajectories, obtained either from tracking of features or from integration of the velocity field;
- spatial objects, such as fronts, eddies or filaments.

Image Models for these structures are developed and take into account the underlying physical processes. Image data are assimilated in Image Models to derive pseudo-observations of state variables, which are further assimilated in numerical ocean forecast models.

4.4. Meteorology

Meteorological forecasting constitutes a major applicative challenge for Image Assimilation. Although satellite data are operationally assimilated within models, this is mainly done on an independent pixel basis: the observed radiance is linked to the state variables via a radiative transfer model, that plays the role of an observation operator. Indeed, because of their limited spatial and temporal resolutions, numerical weather forecast models fail to exploit image structures, such as precursors of high impact weather:

- cyclogenesis related to the intrusion of dry stratospheric air in the troposphere (a precursor of cyclones),
- convective systems (supercells) leading to heavy winter time storms,
- low-level temperature inversion leading to fog and ice formation, *etc.*

To date, there is no available method for assimilating such data, which are characterized by a strong coherence in space and time. Meteorologists have developed qualitative Conceptual Models (CMs), for describing the high impact weathers and their signature on images, and tools to detect CMs on image data. The result of this detection is used for correcting the numerical models, for instance by modifying the initialization. The aim is therefore to develop a methodological framework allowing to assimilate the detected CMs within numerical forecast models. This is a challenging issue given the considerable impact of the related meteorological events.

CORTEX Project-Team

4. Application Domains

4.1. Overview

Our application domain is twofold:

On one hand, neuro-scientists are end-users of our researches. Data analysis is one issue, but the main outcomes concern modeling, namely the validation of biological assumptions either at a theoretical level or via numerical experiments and simulation of bio-processes. This includes algorithmic expertises and dedicated softwares.

On the other hand, science and technology of information processing is impacted. This concerns embedded systems such as in-silico implementations of bio-inspired processes, focusing on spatial and distributed computing. This also concerns embodied systems such as robotic implementation of sensori-motor loops, the bio-inspiration yielding such interesting properties as adaptivity and robustness.

DEMAR Project-Team

4. Application Domains

4.1. Objective quantification and understanding of movement disorders

One main advantage of developing a model based on a physical description of the system is that the parameters are meaningful. Therefore, these parameters when identified on a given individual (valid or deficient), give objective and quantitative data that characterize the system and thus can be used for diagnosis purposes.

Modelling provides a way to simulate movements for a given patient and therefore based on an identification procedure it becomes possible to analyse and then understand his pathology. In order to describe complex pathology such as spasticity that appears on paraplegic patients, you need not only to model the biomechanics parts - including muscles -, but also parts of the peripheral nervous system - including natural sensors - to assess reflex problems. One important application is then to explore deficiencies globally due to both muscles and peripheral neural nets disorders.

4.2. Palliative solutions for movement deficiencies

Functional electrical stimulation is one possibility to restore or control motor functions in an evolutive and reversible way. Pacemaker, cochlear implants, deep brain stimulation (DBS) are successful examples. DEMAR focuses on movement disorder restoration in paraplegic and quadriplegic patients, enhancements in hemiplegic patients, and some other motor disorders such as bladder and bowel control. Nevertheless, since some advances in neuroprosthetic devices can be exploited for the next generation of cochlear implants, the team also contributes to technological and scientific improvements in this domain.

The possibility to interface the sensory motor system, both activating neural structure with implanted FES, and sensing through implanted neural signal recordings open a wide application area:

- Restoring motor function such as grasping for quadriplegic patient, standing and walking for paraplegic patient, compensating foot drop for hemiplegic patients. These applications can be firstly used in a clinical environment to provide physiotherapist with a new efficient FES based therapy (using mainly surface electrodes) in the rehabilitation process. Secondly, with a more sophisticated technology such as implanted neuroprostheses, systems can be used at home by the patient himself without a clinical staff.
- Modulating motor function such as tremors in Parkinsonian patient using DBS. Techniques are very similar but for the moment, modelling is not achieved because it implies the central nervous system modelling in which we are not implied.
- Sensing the afferent pathways such as muscle's spindles, will be used to provide a closed loop control of FES through natural sensing and then a complete implanted solution. Sensing the neural system is a necessity in some complex motor controls such as the bladder control. Indeed, antagonist muscle's contractions, and sensory feedbacks interfere with FES when applied directly on the sacral root nerve concerned. Thus, enhanced activation waveforms and sensing feedback or feedforward signals are needed to perform a highly selective stimulation.

To achieve such objectives, experimentations in animals and humans are necessary. This research takes therefore a long time in order to go from theoretical results to real applications. This process is a key issue in biomedical research and is based on: i) design of complex experimental protocols and setups both for animals and humans, ii) ethical attitude both for humans and animals, with ethical committee approval for human experiments iii) volunteers and selected, both disabled and healthy, persons to perform experiments with the adequate medical staff.

DRACULA Project-Team

4. Application Domains

4.1. Normal hematopoiesis

4.1.1. Introduction

Modelling normal hematopoiesis will allow us to explore the dynamical appearance of the various cell types, originating from the stem cell compartment, through the bone marrow development up to the blood stream. The differentiated cell types will both fulfill physiological functions, and play a key role on the feedback control on homeostasis (balance of the system) in their own lineages. We will describe the hematopoiesis from three different points of view:

- The initial cell type, the hematopoietic stem cell (HSC);
- The lineage choice question;
- Three differentiated lineages that are responsible for specific function, namely oxygen transport, immune response and coagulation.

The basic mechanisms of our modelling approach are as follows:

- Any cell type can have two possibilities at each time step: to divide or to die.
- At any division step, the cell can either give rise to two daughter cells which are identical to the mother cell (self-renewal) or that are more advanced in their differentiation.

All these processes will be first modelled at the cellular level. In parallel, we will develop models of intra-cellular molecular networks (as some proteins controlling the cell cycle) influencing this decision making process, so as to be able to describe both micro-to-macro effects (molecules influencing the global cell behaviour) as well as macro-to-micro effects (like the global state of the cell population influencing the molecular behaviour).

4.1.2. Hematopoietic stem cells (HSC)

Although widely studied by biologists, HSC are still poorly understood and many questions remain open: How fast and how frequently do they divide? How many of them are in the bone marrow and where? How is their behaviour modified under stress conditions such as blood loss or transfusion?

Our modelling approach will be based on two methods: deterministic and stochastic differential equations with delays (discrete and distributed), on one hand, and the DPD method using the individual based modelling on the other hand. The differential equation models based on the work initiated by Mackey [38] will describe the HSC compartment in normal conditions and the behaviour of these cells under some stress. The DPD method, as a complementary approach, will emphasize the spatial regulation of stem cell behaviour, and we will focus our attention to give a possible answer regarding their location in the bone marrow and the roles of the niche, their number in the system, their possible role under stress (that is their reaction under the different feedback controls).

4.1.3. Lineage choice

Positive and negative feedbacks in intra-cellular regulatory networks create a bistable or multistable situation where different cell populations can co-exist. This allows the production of different blood cells beginning from stem cells. It is an important property of hematopoietic cell populations, which is not yet completely understood. We will focus on the erythroid/myelomonocytic choice, which is governed by a balance of lineage-affiliated transcription factors, such as GATA1 and PU.1. How the ratios of lineage-determining transcription factors stabilize progenitor cells and resolve their indeterminacy to commit them to discrete, mutually exclusive fates remains unexplained.

We will analyze the dynamics of a binary fate decision governed by a gene-circuit containing auto-stimulation and cross-inhibition, as embodied by the GATA1-PU.1 paradigm. We will use mathematical models based on ordinary and partial differential equations and individually based modelling to study fundamental properties of hematopoiesis and its quantitative characteristics. We will also explore the fate decision process from a stochastic point of view.

4.1.4. Blood cell functions

(i) O_2 transport: red lineage

O_2 transport is provided by red blood cells (RBC) also called erythrocytes. Many different stages of maturity (including progenitors, precursors, reticulocytes and erythrocytes) are necessary to achieve the complete formation of RBC. These latter are then released in the blood stream where they transport oxygen. The whole process is tightly dependent on a robust well-balanced equilibrium called homeostasis.

It has been shown in the 1990's that apoptosis is regulated by EPO, a growth factor released by the kidneys under hypoxia. But also, under severe stress (like an important blood loss) some other molecules known as glucocorticoids can be released leading to an increase of the self-renewing rate for each generation. This led to the formulation of a first model, demonstrating the role of self-renewal.

The study of the red blood cell lineage will involve different scale levels, from the molecular one, with the effects of the hormones on the surface and internal parts of the cell, the cell contacts in each stage of RBC formation, and the red branch population in its whole with all the interactions taken into account (see Figure 2) in normal and stress conditions.

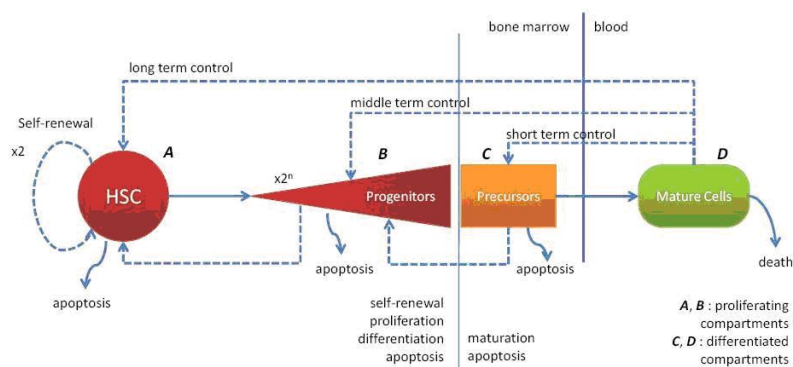


Figure 2. Scheme of Erythropoiesis Modelling. Without considering explicitly growth factor mediated regulation, all controls (proliferation, self-renewal, differentiation, apoptosis) are mediated by cell populations (dashed arrows). Mature cells can either regulate immature (HSC, progenitors) or almost mature (precursors) cells, precursors may act on progenitor dynamics, etc..

In order to couple the cellular behaviour to explicit molecular events, we will describe the events through a molecular network that is based upon the work of [42]. A first version of this model is shown in Figure 3 .

(ii) Immune response

We will focus on the production of T-cells during an immune response. This represents an important activity of the lymphoid branch, part of leucopoiesis (white blood cell production). Several models of the myeloid branch of leucopoiesis have been investigated in the frame of specific diseases (for instance cyclical neutropenia ([37], [34]), chronic myelogenous leukemia [39]).

Time evolution of T-cell counts during an infection is well known: following the antigen presentation, the number of cells quickly increases (expansion), then decreases more slowly (contraction) and stabilizes around a value higher than the initial value. Memory cells have been produced, and will allow a faster response when encountering the antigen for a second time. Mechanisms that regulate this behaviour are however not well known.

A recent collaboration just started with immunologists (J. Marvel, Ch. Arpin) from the INSERM U851 in Lyon, who provide experimental data that are essential to assess the significance of models, based on strongly nonlinear ordinary differential equations, that can be proposed for T-cell production (Figure 4). By considering molecular events leading to cell activation when encountering a virus, we will propose a multi-scale model of the immune response.

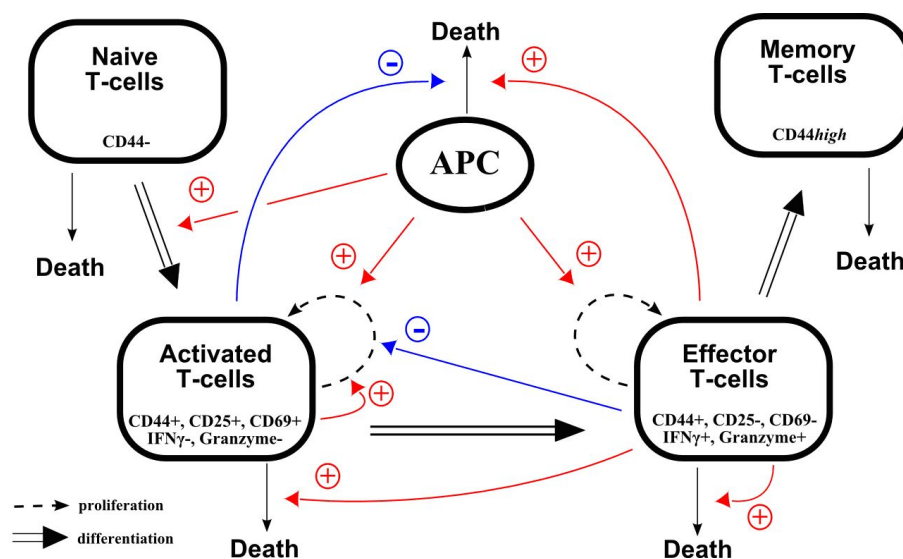


Figure 4. Model of the immune response resulting in the generation of CD8 memory T cells. The response starts with a viral infection resulting in the presentation of viral antigens through antigen presenting cells (APC) to naive T-cells. These latter, once activated, differentiate into activated cells which, under specific feedback loops will either die, differentiate into effector cells or self-renew. Differentiation of effector cells (killer cells) will result in the production of memory cells.

(iii) Coagulation: platelet lineage

Thrombopoiesis, the process of production and regulation of platelets, is similar to erythropoiesis although important differences are observed. These two processes have an immature progenitor (MEP) in common. Platelets are involved in blood coagulation, and can be the source of blood diseases (thrombopenia, thrombocytosis). Their production is mainly regulated by thrombopoietin (TPO), a growth factor similar to EPO.

It is important to mention that very few experimental data exist in the literature, and mathematical modelling of thrombopoiesis did not attract so much attention in the past 20 years. However, collaboration with some leading hematologists in this domain will allow us to get updated and new data regarding this process.

Deterministic models, in the form of structured transport partial differential equations, will be proposed to describe platelet dynamics, through the description of HSC, megakaryocytic progenitor and megakaryocyte (platelet precursor) compartments. Circulating TPO, regulated by platelets, will induce feedback loops in

thrombopoiesis, and we will investigate the dynamics of platelet production and emergence of platelet-related diseases.

4.2. Pathological hematopoiesis

The knowledge of hematopoiesis and related diseases has evolved to become a great deal in the past years, and Mackey's previous models (ref. [32]) do not allow us to correctly answer current questions that are clearly oriented toward the investigation of cell signalling pathways. These models nevertheless bring relevant ideas about the essential features of such modelling. It is also noteworthy that even though models of hematopoiesis have existed for quite a long time, their application to questions of explanation and prediction of hematopoiesis dynamics that are encountered in the clinic is still not sufficiently frequent, even though much progress has been achieved in the cooperation between hematologists and mathematicians [40]. This is in the optic of testable experimental predictions that the multi-scale model for pathological hematopoiesis will be developed. For instance, we will concentrate on myeloid leukemias (CML and AML) and their treatment.

4.2.1. Leukemia Modelling

(i) Chronic Myeloid Leukemia

The strong tyrosine kinase activity of the BCR-ABL protein is the basis for the main cell effects that are observed in CML: significant proliferation, anti-apoptotic effect, disruption of stroma adhesion properties, genomic instability. This explains the presence in CML blood of a very important number of cells belonging to the myeloid lineage, at all stages of maturation.

We will consider models based on ordinary differential equations for the action of the main intra- and extra-cellular proteins involved in CML (as BCR-ABL protein), and of transport equations (with or without delay, physiologically structured or not to represent healthy and leukemic cell populations, take into account many interactions between proteins (especially BCR-ABL), cells (anti-apoptotic effect, etc.), and their environment (disruption of stroma adhesion properties, for example). Transport pertains thus to cells from one compartment (or a group of compartments) to another compartment, with a determined speed of aging or maturation. These compartments may be detailed or not: the less mature are stem cells, then progenitor cells, etc.

(ii) Acute Myeloid Leukemia

The natural history of CML leads to its transformation ("blast crisis") in acute myeloid leukemia (AML), following supplementary genetic alterations that produce a maturation arrest (myeloid in 3/4 of cases, lymphoid in 1/4 of cases, confirming the insult to pluripotent stem cells), leading to an accumulation of immature cells in the bone marrow and in the general circulation, resulting in deep medullary impairment and fast fatal outcome, in spite of chemotherapy. This phenomenon is the same as the one observed in de novo AML, i.e., AML without a previous chronic phase.

The different modelling methods of AML will be similar to the ones described for CML, with some exceptions: the appearance of BCR-ABL mutations, which are not relevant in the case of AML, the appearance of a gene (*spi-1*) involved in the differentiation arrest, and constitutive activation of EPO receptor or Kit activating mutations promote proliferation and survival. This explains the accumulation of immature cells in the bone marrow and in the blood stream.

4.2.2. Treatment

As far as treatment of pathological hematopoiesis is concerned, two main strategies currently exist that aim at slowing down or eliminating damaged cell proliferation. The first of these strategies consists in launching the apoptotic process during the cell division cycle. This process is activated, for example when the cell is unable to repair damages, e.g., after exposure to cytostatic drugs. A typical example is apoptosis induced by chemotherapy-induced DNA damage: The damage is recognised by the cell, which then activates the sentinel protein p53 ("guardian of the genome") that arrests the cell cycle to allow, if possible, damage repair. If the latter is unrecoverable, then p53 activates the endogenous apoptotic processes.

The second strategy aims at pushing damaged cells toward the differentiation that has been stopped in the course of their genetic mutation. Since a few years back, a new approach has been developed around the strategy of differentiation therapy. This therapy relies on molecules (growth factors and specific cytokines) that are able to re-initialise the cell differentiation programs that have been modified during malignant transformation. The cancer that is most concerned by the development of this differentiation therapy is AML whose malignant cells present highly undifferentiated features and the ones that present a translocation responsible for the differentiation (PML/RAR of the promyelocytic form, AML1/ETO and CBFbeta/MyH11, involving Core Binding Factors alpha and beta).

Mathematical models based on ordinary differential equations will be developed to describe the action of drugs (in the two cases mentioned above). They will take into account interactions between drugs and their environment. Our goal will be the optimization of possible synergies between drugs acting on distinct cellular targets, and the control of resistances to these treatments as well as their toxicities.

Curative and palliative strategies must take into account the dynamics of healthy and leukemic hematopoietic cells at multiple scales. In time, from optimal scheduling of combination therapy (hours) to avoiding the development of resistances and relapse (months to years). In space, from the stem cell niche to circulating blood. In organization, from gene and signalling networks (JAK/STAT, BCR-ABL) to cell populations and cytokine regulation (EPO, CSFs). Several recent qualitative models have provided insight in the complex dynamics of the disease and the response to treatments. Many of these models focus on the control or regulation processes that promote homeostasis or oscillatory behavior in cell number. However, as A. Morley points out, "once the control-systems features of hematopoiesis are accepted, the ability to construct a model that shows oscillatory behavior, even if the model incorporates the latest advances in hematopoietic cell biology, really adds little new knowledge. Rather, the challenge to modellers would seem to be to provide detailed predictions for the input-output characteristics of the different parts of the various control systems so that these predictions can be tested by experimental hematologists and a truly quantitative description of hematopoiesis can emerge".

We propose for instance, to use models in the form of structured transport partial differential equations (with or without delay, physiologically structured or not) to represent the competition between target, resistant and healthy cell populations. The resulting models to describe the dynamic of these cell populations under the action of drugs are multi-scale systems of the form (Hyperbolic PDE)-ODE or DDE-ODE. For instance, we will develop mathematical models of chronotherapy and pharmacotherapy for CML and AML.

DYLISS Team

4. Application Domains

4.1. Formal models in molecular biology

As mentioned before, our main goal in biology is to characterize groups of genetic actors that control the response of living species capable of facing extreme environments. To focus our developments, applications and collaborations, we have identified three biological questions which deserve integrative studies. Each axis may be considered independently from the others although their combination, a mid-term challenge, will have the best impact in practice towards the long-term perspective of identifying proteins controlling the production of a metabolite of industrial interest. It is illustrated in our presentation for a major algae product: polyunsaturated fatty acids (PUFAs) and their derivatives.

4.2. Biological data integration

Axis 1 (data integration) aims at identifying **who** is involved in the specific response of a biological system to an environmental stress. Targeted actors will mainly consist in groups of genetic products or biological pathways. For instance, which pathways are implied in the specific production of PUFAs in brown algae? The main work is to represent in a system of logical constraints the full knowledge at hand concerning the genetic or metabolic actors, the available observations and the effects of the system dynamics. To this aim, we focus on the use of Answer Set Programming as we are experienced in modeling with this paradigm and we have a strong partnership with a computer science team leader in the development of dedicated grounders and solvers (Potsdam university).

4.3. Asymptotic dynamics of a biological system

Once a model is built and its main actors are identified, the next step is to clarify **how** they combine to control the system (**Axis 2**). Roughly, the fine tuning of the system response may be of two types. Either it results from the discrete combinatorics of the actors, as the result of a genetic adaptation to extreme environmental conditions or the difference between species is rather at the enzyme-efficiency level. For instance, if Pufa's are found to be produced using a set of pathways specific to brown algae, the work in axis 2 will consist to apply constraint-based combinatorial approaches to select consistent combinations of pathways controlling the metabolite production. Otherwise, if enzymes controlling the production of Pufa's are found to be expressed in other algae, it suggests that the response of the system is rather governed by a fine quantitative tuning of pathways. In this case, we use symbolic dynamics and average-case analysis of algorithms to weight the respective importance of interactions in observed phenotypes (see Fig. 2). This specific approach is motivated by the quite restricted spectrum of available physiological observations over the asymptotic dynamics of the biological system.

4.4. Biological sequence annotation

In order to check the accuracy of in-silico predictions, a last step (**Axis 3**) is to extract genetic actors responsible of biological pathways of interest in the targeted organism and locate them in the genome. In our guiding example, active proteins implied in Pufa's controlling pathways have to be precisely identified. Actors structures are represented by syntactic models (see figure 4). We use knowledge-based induction on far instances for the recognition of new members of a given sequence family within non-model genomes (see figure 3). A main objective is to model enzyme specificity with highly expressive syntactic structures - context-free model - in order to take into account constraints imposed by local domains or long-distance interactions within a protein sequence.

FLUMINANCE Project-Team

4. Application Domains

4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing technique for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of closed surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by lot of scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

GALEN Team

4. Application Domains

4.1. Clinical Projects

- **MR & Muscular Diseases:** The use of MR and Diffusion Tensor Imaging are investigated in collaboration with the Henri Mondor University Hospital and Institut of Myology towards automatic quantification of muscular mass loss and non-invasive biopsy. The aim is to provide tools that could be used to automatically analyze MR imaging and extract useful clinical measurements (Institut of Myology), and assess the potential impact of diffusion tensor imaging towards automatic quantification either of muscular diseases progression.
- **Image-driven Radiotherapy Treatment & Surgery Guidance :** The use of CT and MR imaging for cancer guidance treatment in collaboration with the Oscar Lambert Center. The aim is to provide tools for automatic dose estimation as well as off-line and on-line positioning guidance through deformable fusion between imaging data corresponding to perioding patient treatment. The same concept will be explored in collaboration with the Saint-Antoine University Hospital towards image-driven surgery guidance through 2D to 3D registration between interventional and pre-operative annotated data.
- **MR Brain Imaging towards Low-Gliomas Tumor Brain Understanding:** The use of contrast enhanced imaging is investigated in collaboration with the Montpellier University Hospital towards better understanding of low-gliomas positioning, automatic tumor segmentation/identification and longitudinal (tumor) growth modeling.
- **CT/MR Perfusion Imaging:** The use of perfusion imaging is investigated in collaboration with the Georges Pompidou European Hospital towards compartmental analysis and measuring tissue perfusion and capillary permeability in liver tumors.

GENSCALE Team

4. Application Domains

4.1. Sequence comparison

Historically, sequence comparison has been one of the most important topics in bioinformatics. BLAST is a famous software tool particularly designed for solving problems related to sequence comparisons. Initially conceived to perform searches in databases, it has mostly been used as a general-purpose sequence comparison tool. Nowadays, together with the inflation of genomic data, other software comparison tools that are able to provide better quality solutions (w.r.t the ones provided by BLAST) have been developed. They generally target specific comparison demands, such as read mapping, bank-to-bank comparison, meta-genomic sample analysis, etc. Today, sequence comparison algorithms must clearly be revisited to scale up with the very large number of sequence objects that new NGS problems have to handle.

4.2. Genome comparison

This application domain aims at providing a global relationship between genomes. The problem lies in the different structures that genomes can have: segments of genome can be rearranged, duplicated or deleted (the alignment can no longer be done in one piece). Therefore one major aim is the study of chromosomal rearrangements, breaking points, structural variation between individuals of the same species, etc. However, even analyses focused on smaller variations such as Single Nucleotide Polymorphisms (SNP) at the whole genome scale are different from the sequence comparison problem, since one needs first to identify common (orthologous) parts between whole genome sequences and thus obtain this global relationship (or map) between genomes. New challenges in genome comparison are emerging with the evolution of sequencing techniques. Nowadays, they allow for comparing genomes at intra-species level, and to deal simultaneously with hundreds or thousands of complete genomes. New methods are needed to find the sequence and structural variants between such a large number of non-assembled genomes. Even for the comparison of more distant species, classical methods must be revisited to deal with the increasing number of genomes but more importantly their decreasing quality: genomes are no longer fully assembled nor annotated.

4.3. Protein comparison

Comparing protein is important for understanding their evolutionary relationships and for predicting their structures and their functions. While annotating functions for new proteins, such as those solved in structural genomics projects, protein structural alignment methods may be able to identify functionally related proteins when the sequence identity between a given query protein and the related proteins are low (i.e. lower than 20%). Moreover, protein comparison allows for solving the so-called protein family identification problem. Given an unclassified protein structure (query), the comparison of protein structures can be used for assigning a score measuring the "similarity" between the query and the proteins belonging to a set of families. Based on this score, the query is assigned to one of the families of the set. The knowledge acquired by performing such analyses can then be exploited in methods for protein structure prediction that are based on a homology modeling approach.

IBIS Project-Team (section vide)

MACS Project-Team

4. Application Domains

4.1. Application domains

Our researches have natural applications in all sectors of the mechanical industry: car and naval industries; aeronautics and space; civil engineering; tires; MEMs and nanotechnologies...

We also actively seek new applications in biotechnologies, although of course the economy and structuring of this sector is not as developed yet.

MAGIQUE-3D Project-Team

4. Application Domains

4.1. Seismic Imaging

The main objective of modern seismic processing is to find the best representation of the subsurface that can fit the data recorded during the seismic acquisition survey. In this context, the seismic wave equation is the most appropriate mathematical model. Numerous research programs and related publications have been devoted to this equation. An acoustic representation is suitable if the waves propagate in a fluid. But the subsurface does not contain fluids only and the acoustic representation is not sufficient in the general case. Indeed the acoustic wave equation does not take some waves into account, for instance shear waves, turning waves or the multiples that are generated after several reflections at the interfaces between the different layers of the geological model. It is then necessary to consider a mathematical model that is more complex and resolution techniques that can model such waves. The elastic or viscoelastic wave equations are then reference models, but they are much more difficult to solve, in particular in the 3D case. Hence, we need to develop new high-performance approximation methods.

Reflection seismics is an indirect measurement technique that consists in recording echoes produced by the propagation of a seismic wave in a geological model. This wave is created artificially during seismic acquisition surveys. These echoes (i.e., reflections) are generated by the heterogeneities of the model. For instance, if the seismic wave propagates from a clay layer to sand, one will observe a sharp reflected signal in the seismic data recorded in the field. One then talks about reflection seismics if the wave is reflected at the interface between the two media, or talks about seismic refraction if the wave is transmitted along the interface. The arrival time of the echo enables one to locate the position of this transition, and the amplitude of the echo gives information on some physical parameters of the two geological media that are in contact. The first petroleum exploration surveys were performed at the beginning of the 1920's and for instance, the Orchard Salt Dome in Texas (USA) was discovered in 1924 by the seismic-reflection method.

MAGNOME Project-Team

4. Application Domains

4.1. Function and history of yeast genomes

Yeasts provide an ideal subject matter for the study of eukaryotic microorganisms. From an experimental standpoint, the yeast *Saccharomyces cerevisiae* is a model organism amenable to laboratory use and very widely exploited, resulting in an astonishing array of experimental results. From a genomic standpoint, yeasts from the hemiascomycete class provide a unique tool for studying eukaryotic genome evolution on a large scale. With their relatively small and compact genomes, yeasts offer a unique opportunity to explore eukaryotic genome evolution by comparative analysis of several species.

- Yeasts are widely used as cell factories, for the production of beer, wine and bread and more recently of various metabolic products such as vitamins, ethanol, citric acid, lipids, etc.
- Yeasts can assimilate hydrocarbons (genera *Candida*, *Yarrowia* and *Debaryomyces*), depolymerise tannin extracts (*Zygosaccharomyces rouxii*) and produce hormones and vaccines in industrial quantities through heterologous gene expression.
- Several yeast species are pathogenic for humans, especially *Candida albicans*, *Candida glabrata*, *Candida tropicalis* and the Basidiomycete *Cryptococcus neoformans*.

The hemiascomycetous yeasts represent a homogeneous phylogenetic group of eukaryotes with a relatively large diversity at the physiological and ecological levels. Comparative genomic studies within this group have proved very informative [33], [48], [47], [35], [50], [30], [31], [6].

MAGNOME applies its methods for comparative genomics and knowledge engineering to the yeasts through the ten-year old *Génolevures* program (GDR 2354 CNRS), devoted to large-scale comparisons of yeast genomes with the aim of addressing basic questions of molecular evolution. We developed the software tools used by the CNRS's genolevures.org web site. For example, MAGNOME's MAGUS system for simultaneous genome annotation combines semi-supervised classification and rule-based inference in a collaborative web-based system that explicitly uses comparative genomics to simultaneously analyse groups of related genomes.

4.2. Alternative fuels and bioconversion

Oleaginous yeasts are capable of synthesizing lipids from different substrates other than glucose, and current research is attempting to understand these conversions with the goal of optimizing their throughput, production and quality. From a genomic standpoint the objective is to characterize genes involved in the biosynthesis of precursor molecules which will be transformed into fuels, which are thus not derived from petroleum. Biological experimentation by partner laboratories study lipid accumulation in the oleaginous yeasts such as *Yarrowia lipolytica* starting from:

- pentoses, produced from lignin cellulose agricultural substrates following a biorefining strategy,
- glycerol, a secondary output of chemical production of biodiesel, and
- industrial residues.

Lipases from *Y. lipolytica* are of particular interest (see [38] for review). Experimental characterization of the lipid bodies produced from these substrates will aid in the identification of target genes which may serve for genetic engineering. This in turn requires the development of molecular tools for this class of yeasts with strong industrial potential. MAGNOME's focus is in acquiring genome sequences, predicting genes using models learned from genome comparison and sequencing of cDNA transcripts, and comparative annotation. Our overall goal is to define dynamic models that can be used to predict the behavior of modified strains and thus drive selection and genetic engineering.

4.3. Winemaking and improved strain selection

Yeasts and bacteria are essential for the winemaking process, and selection of strains based both on their efficiency and on the influence on the quality of wine is a subject of significant effort in the Aquitaine region. Unlike the species studied above, yeast and bacterial starters for winemaking cannot be genetically modified. In order to propose improved and more specialized starters, industrial producers use breeding and selection strategies.

Yeast starters from the *Saccharomyces* genus are used for primary, alcohol fermentation. Recent advances have made it possible to identify the genetic causes of the different technological differences between strains [55], [54], [53]. Manipulating the genetic causes rather than the industrial consequences is far more amenable to experimental development. An essential tool in identifying these genetic causes is comparative genomics.

Bacterial starters based on *Oenococcus oeni* are used in secondary, malolactic fermentation. Genetically, *O. oeni* presents a surprising level of intra-specific diversity, and clues that it may evolve more rapidly than expected. Studying the diversity of the *O. oeni* genomes has led to genetic tools that can be used to evaluate the predisposition of different strains to respond to oenological stresses. While identifying particular genes has been the leading strategy up to now, recently a new strategy based on comparative genomics has been undertaken to understand the impact and mechanisms of genetic diversity [20], [23], [21], [32], [36], [28] [3].

Starting from historical collaborations by Pascal Durrens and Elisabeth Bon with partners from the Institute for Wine and Vine Sciences in Bordeaux (ISVV), and local industry, and in the framework of an effective partnership, we apply our tools to large-scale comparative genomics of yeast and bacterial starters in winemaking.

4.4. Knowledge bases for molecular tools

Affinity binders are molecular tools for recognizing protein targets, that play a fundamental in proteomics and clinical diagnostics. Large catalogs of binders from competing technologies (antibodies, DNA/RNA aptamers, artificial scaffolds, etc.) and Europe has set itself the ambitious goal of establishing a comprehensive, characterized and standardized collection of specific binders directed against all individual human proteins, including variant forms and modifications. Despite the central importance of binders, they presently cover only a very small fraction of the proteome, and even though there are many antibodies against some targets (for example, >900 antibodies against p53), there are none against the vast majority of proteins. Moreover, widely accepted standards for binder characterization are virtually nonexistent.

Alongside the technical challenges in producing a comprehensive binder resource are significant logistical challenges, related to the variety of producers and the lack of reliable quality control mechanisms. As part of the ProteomeBinders and Affinomics projects, MAGNOME works to develop knowledge engineering techniques for storing, exploring, and exchanging experimental data used in affinity binder characterization. This work involves databases and tools for molecular interaction data [42] [45], standards for data exchange between peers [44], [49], [41] and reporting standards [4] [58].

MASAIE Project-Team

4. Application Domains

4.1. Metapopulation models

Heterogeneity plays an important role in many infectious disease processes. For instance, spatial heterogeneity is a strong determinant of host-parasite relationships. In modeling spatial or geographic effects on the spread of a disease, a distinction is usually made between diffusion and dispersal models. In diffusion models, spread is to immediately adjacent zones, hence the phenomenon of traveling waves can appear. These models traditionally use partial differential equations. However, there are some important situations that cannot be modeled by PDE. This is the case when the space considered is discrete. For example, when we have to consider sparsely populated regions, the human population is located in patches. The organization of human-hosts into well-defined social units such as families, villages or cities, are good examples of patches. Another example arises in the study of the human African Trypanosomiasis. The vector is the tse-tse fly, and it is known that flies take fewer blood meals in villages than in coffee plantations where the villagers work during the day. For such situations where human or vectors can travel a long distance in a short period of time, dispersal models are more appropriate. These models consider migration of individuals between patches. The infection does not take place during the migration process. The situation is that of a directed graph, where the vertices represent the patches and the arcs represent the links between patches. Recently, there has been increased interest in these deterministic metapopulation disease models. We have generalized to n patches the Ross-Macdonald model which describes the dynamics of malaria. We incorporate in our model the fact that some patches can be vector free. We assume that the hosts can migrate between patches, but not the vectors. The susceptible and infectious individuals have the same dispersal rate. We compute the basic reproduction ratio \mathcal{R}_0 . We prove that if $\mathcal{R}_0 \leq 1$, then the disease-free equilibrium is globally asymptotically stable. When $\mathcal{R}_0 > 1$, we prove that there exists a unique endemic equilibrium, which is globally asymptotically stable on the biological domain minus the disease-free equilibrium.

MASAIE is developing, in the framework of the CAPES-COFECUB project (see international program), a metapopulation model for dengue. This model is for the state of Rio and is using the data of foundation FIOCRUZ.

4.2. Estimating total parasite load in falciparum malaria patients

We give a brief review of the biological features of malaria. Malaria in a human begins with an inoculum of *Plasmodium* parasites (sporozoites) from a female *Anopheles* mosquito. The sporozoites enter the liver within minutes. After a period of asexual reproduction in the liver, the parasites (merozoites) are released in the bloodstream where the asexual erythrocyte cycle begins. The merozoites enter red blood cells (RBC), grow and reproduce over a period of approximately 48 hours after which the erythrocyte ruptures releasing daughter parasites that quickly invade a fresh erythrocyte to renew the cycle. This blood cycle can be repeated many times, in the course of which some of the merozoites instead develop in the sexual form of the parasites : gametocytes. Gametocytes are benign for the host and are waiting for the mosquitoes. An important characteristic of *Plasmodium falciparum*, the most virulent malaria parasite, is sequestration. At the half-way point of parasite development, the infected erythrocyte leaves the circulating peripheral blood and binds to the endothelium in the microvasculature of various organs where the cycle is completed. A measurement of *Plasmodium falciparum* parasitaemia taken from a blood smear therefore samples young parasites only. Physician treating malaria use the number of parasites in peripheral blood smears as a measure of infection, this does not give the total parasite burden of the patient. Moreover antimalarial drugs are known to act preferentially on different stages of parasite development. Our work consists in developing tools for estimating the sequestered parasites and hence the total parasite burden of the patient.

MNEMOSYNE Team

4. Application Domains

4.1. Overview

One of the most original specificity of our team is that it is part of a laboratory in Neuroscience (with a large spectrum of activity from the molecule to the behavior), focused on neurodegenerative diseases and consequently working in tight collaboration with the medical domain. As a consequence, neuroscientists and the medical world are considered as the primary end-users of our researches. Beyond data and signal analysis where our expertise in machine learning may be possibly useful, our interactions are mainly centered on the exploitation of our models. They will be classically regarded as a way to validate biological assumptions and to generate new hypotheses to be investigated in the living. Our macroscopic models and their implementation in autonomous robots will allow an analysis at the behavioral level and will propose a systemic framework, the interpretation of which will meet aetiological analysis in the medical domain and interpretation of intelligent behavior in cognitive neuroscience.

The study of neurodegenerative diseases is targeted because they match the phenomena we model. Particularly, the Parkinson disease results from the death of dopaminergic cells in the basal ganglia, one of the main systems that we are modeling. The Alzheimer disease also results from the loss of neurons, in several cortical and subcortical regions. The variety of these regions, together with large mnemonic and cognitive deficits, require a systemic view of the cerebral architecture and associated functions, very consistent with our approach.

Of course, numerical sciences are also impacted by our researches, at several levels. At a global level, we will propose new control architectures aimed at providing a higher degree of autonomy to robots, as well as machine learning algorithms working in more realistic environment. More specifically, our focus on some cognitive functions in closed loop with a real environment will address currently open problems. This is obviously the case for planning and decision making; this is particularly the case for the domain of affective computing, since motivational characteristics arising from the design of an artificial physiology allow to consider not only cold rational cognition but also hot emotional cognition. The association of both kinds of cognition is undoubtedly an innovative way to create more realistic intelligent systems but also to elaborate more natural interfaces between these systems and human users.

At last, we think that our activities in well-founded distributed computations and high performance computing are not just intended to help us design large scale systems. We also think that we are working here at the core of informatics and, accordingly, that we could transfer some fundamental results in this domain.

MODEMIC Project-Team

4. Application Domains

4.1. Preservation of water resources

The biological decontamination of wastewater is our main application domain, in the continuation of the long collaboration with the INRA research laboratory LBE. We target applications from the decontamination industry, held by large groups as well as small companies specialized in specific pollutants (for instance in fish farming). We aim also to study connected application domains for

- the aquatic ecology where microorganisms play an important role in the quality of natural water resources,
- the re-use of water in arid climates for countries of North of Africa, within the euro-Mediterranean project TREASURE.

4.2. Microbial ecology of soil

This application domain is more recent for the team members. We target

- the research questions raised by agronomists, about the better understanding of the interactions and the biodiversity of microbial communities in soils, with the help of models and numerical simulations,
- the role of spatial structures on the functions or *ecological services* of microbial ecosystems, notably the soil fertility and the carbon sequestration.

4.3. Control of fermentation processes

Very closely to our studies about wastewater bioreactors and chemostat models, we target applications in fermentation processes:

- either for agro-food products. A typical application is the control of cascade fermenters in the study of wine fermentation with UMR SPO (Sciences Pour l'Enologie, Montpellier), within the European project CAFE.
- either for the green chemistry. A typical application is the consideration of spatialization in enzymatic models of production of agro-polymers with UMR IATE (Ingénierie des Agro-polymères et Technologies Émergentes, Montpellier).

4.4. Animal digestive ecosystem

Ruminants absorb plant cells, mainly constituted by cellulose, from which the microbial population of their digestive system extracts carbon and energy to provide proteins and energetic molecules. This bio-conversion produces also important quantities of methane, a gas responsible of part of the greenhouse effect (the billion of cows on earth reject 20% of the methane linked to human activities). INRA researchers have shown that this methane production could be reduced by 30% by changing the proportion of fat acids in the their food, that also implies that the composition of their microbial ecosystem is modified.

This application domain of the microbial ecology is at an early stage. URH team (Unité de Recherche sur les Herbivores, Clermont) has developed an artificial rumen that is close to a chemostat, for testing different kind of nutrition diets. Preliminaries contacts have been taken, and a modelling demand has been clearly formulated and will be taken up by MODEMIC. This theme falls into the research priorities for the environment preservation.

MOISE Project-Team

4. Application Domains

4.1. Introduction

The evolution of natural systems, in the short, mid, or long term, has extremely important consequences for both the global Earth system and humanity. Forecasting this evolution is thus a major challenge from the scientific, economic, and human viewpoints.

Humanity has to face the problem of **global warming**, brought on by the emission of greenhouse gases from human activities. This warming will probably cause huge changes at global and regional scales, in terms of climate, vegetation and biodiversity, with major consequences for local populations. Research has therefore been conducted over the past 15 to 20 years in an effort to model the Earth's climate and forecast its evolution in the 21st century in response to anthropic action.

With regard to short-term forecasts, the best and oldest example is of course **weather forecasting**. Meteorological services have been providing daily short-term forecasts for several decades which are of crucial importance for numerous human activities.

Numerous other problems can also be mentioned, like **seasonal weather forecasting** (to enable powerful phenomena like an El Niño event or a drought period to be anticipated a few months in advance), **operational oceanography** (short-term forecasts of the evolution of the ocean system to provide services for the fishing industry, ship routing, defense, or the fight against marine pollution), **air pollution** prediction systems, the prediction of **floods**, or the simulation of **mud flows** and **snow avalanches** for impact studies and regional planning.

As mentioned previously, mathematical and numerical tools are omnipresent and play a fundamental role in these areas of research. In this context, the vocation of MOISE is not to carry out numerical prediction, but to address mathematical issues raised by the development of prediction systems for these application fields, in close collaboration with geophysicists.

4.2. Oceanography and the Ocean-Atmosphere System

Participants: Eric Blayo, Pierre-Antoine Bouttier, Vincent Chabot, David Cherel, Laurent Debreu, Jérémie Demange, Marc Honnorat, Christine Kazantsev, Eugène Kazantsev, François-Xavier Le Dimet, Bénédicte Lemieux-Dudon, Xavier Meunier, Maëlle Nodet, Antoine Rousseau, Manel Tayachi, Arthur Vidard.

Multi-resolution, Coupling Methods, Data Assimilation, Ocean, Atmosphere

Understanding and forecasting the ocean circulation is currently the subject of an intensive research effort by the international scientific community. This effort was primarily motivated by the crucial role of the ocean in determining the Earth's climate, particularly from the perspective of global change. In addition, important recent research programs are aimed at developing operational oceanography, i.e. near real-time forecasting of ocean circulation, with applications for ship routing, fisheries, weather forecasting, etc. Another related field is coastal oceanography, dealing for example with pollution, littoral planning, or the ecosystems management. Local and regional agencies are currently very interested in numerical modelling systems for coastal areas.

Both ocean-alone models and coupled ocean-atmosphere models are being developed to address these issues. In this context, the MOISE project-team conducts efforts mainly on the following topics:

- *Multi-resolution approaches and coupling methods:* Many applications in coastal and operational oceanography require high resolution local models. These models can either be forced at their boundaries by some known data, or be dynamically coupled with a large-scale coarser resolution model. Such model interactions require specific mathematical studies on open boundary conditions, refinement methods (like mesh refinement or stochastic downscaling), and coupling algorithms. The latter have also to be studied in the context of ocean-atmosphere coupled systems.

- *Advanced numerical schemes*: Most ocean models use simple finite difference schemes on structured grids. We are seeking for better schemes allowing both accuracy and good conservation properties, and dealing with irregular boundaries and bottom topography.
- *Data assimilation methods for ocean modelling systems*: The main difficulties encountered when assimilating data in ocean or atmosphere models are the huge dimension of the model state vector (typically 10^6 - 10^8), the strongly nonlinear character of the dynamics, and our poor knowledge of model error statistics. In this context, we are developing reduced order sequential and variational data assimilation methods addressing the aforementioned difficulties. We are also working on the assimilation of lagrangian data, of sequences of images, and on the design of data assimilation methods for multi-resolution models and for coupled systems.

Most of these studies are led in strong interaction with geophysicists, in particular from the Laboratoire des Ecoulements Géophysiques et Industriels (LEGI, Grenoble).

4.3. Glaciology

Participants: Eric Blayo, Bertrand Bonan, Bénédicte Lemieux-Dudon, Maëlle Nodet, Habib Toye Mahamadou Kele.

Inverse Methods, Data Assimilation, Glaciology, Ice Core Dating

The study of past climate is a means of understanding climatic mechanisms. Drillings in polar ice sheets provide a huge amount of information on paleoclimates: correlation between greenhouse gases and climate, fast climatic variability during the last ice age, etc. However, in order to improve the quantitative use of the data from this archive, numerous questions remain to be answered because of phenomena occurring during and after the deposition of snow. An important research aim is therefore to optimally model ice sheets in the vicinity of drilling sites in order to improve their interpretation: age scale for the ice and for the gas bubbles, mechanical thinning, initial surface temperature and accumulation when snow is deposited, spatial origin of ice from the drilling.

In another respect, ice streams represent an important feature of ice flows since they account for most of the ice leaving the ice sheet (in Antarctic, one estimates that ice streams evacuate more than 70% of the ice mass in less than 10% of the coast line). Furthermore, recent observations showed that some important ice streams are presently accelerating. Thus, we seek to improve models of ice sheets, by developing data assimilation approaches in order to calibrate them using available observations.

Another objective is the evaluation of the state of the polar ice caps in the past, and their interactions with the other components of the earth climate, in order to forecast their evolution in the forthcoming centuries. The joint use of models and data, through data assimilation techniques, to improve system description is relatively new for the glaciological community. Therefore inverse methods have to be developed or adapted for this particular purpose.

By gaining and losing mass, glaciers and ice-sheets are playing a key role in the sea level evolution. This is obvious when regarding past as, for example, collapse of the large northern hemisphere ice-sheets after the Last Glacial Maximum has contributed to an increase of 120 m of sea level. This is particularly worrying when the future is considered. Indeed, recent observations clearly indicate that important changes in the velocity structure of both Antarctic and Greenland ice-sheets are occurring, suggesting that large and irreversible changes may have been initiated. This has been clearly emphasized in the last report published by the Intergovernmental Panel on Climate Change (IPCC). IPCC has further insisted on the poor current knowledge of the key processes at the root of the observed accelerations and finally concluded that reliable projections of sea-level rise are currently unavailable. In this context, our general aim is to develop data assimilation methods related to ice flow modelling purpose, in order to provide accurate and reliable estimation of the future contribution of ice-sheets to Sea Level Rise.

Development of ice flow adjoint models is by itself a scientific challenge. This new step forward is clearly motivated by the amount of data now available at both the local and the large scales.

4.4. River Hydraulics

Participants: Eric Blayo, Antoine Rousseau, Manel Tayachi.

Shallow Water (SW) models are widely used for the numerical modeling of river flows. Depending on the geometry of the domain, of the flow regime, and of the level of accuracy which is required, either 1D or 2D SW models are implemented. It is thus necessary to couple 1D models with 2D models when both models are used to represent different portions of the same river. Moreover, when a river flows into the sea/ocean (e.g. the Rhône river in the Mediterranean), one may need to couple a 2D SW with a full 3D model (such as the Navier-Stokes equations) of the estuary. These issues have been widely addressed by the river-engineering community, but often with somehow crude approaches in terms of coupling algorithms. This may be improved thanks to more advanced boundary conditions, and with the use of Schwarz iterative methods for example. We want to tackle these issues, in particular in the framework of a partnership with the French electricity company EDF.

MORPHEME Team (section vide)

NEUROMATHCOMP Project-Team (section vide)

NUMED Project-Team

4. Application Domains

4.1. Stroke

Stroke is a very complex pathology, involving many different time scales and phenomena. Numed is currently developing various models to describe some important aspects of stroke.

4.1.1. Inflammation modelling

MA Dronne has designed a first model of inflammation at cellular level, based on ordinary differential equations. To take into account spatial phenomena, a first partial differential equation based model is under study. Together with Taissia Lelekov Boissard (post doc of the ANR contrat "AVC in silico"), they tried to find biological data to parametrize these models, and to build a basis of qualitative facts that must be reproduced by the model.

MA Dronne has also developed a collaboration with the Mario Negri institute (Milano) through the team "inflammation and nervous system diseases" (MG de Simoni). This team currently runs in vivo experiments in rodent that should provide new data to investigate the temporal evolution of various variables of the model.

The study and validation of these two models of inflammation will continue with the study of in silico experiments which will simulate the action of various anti inflammatory drugs, acting at various levels of the inflammatory reaction, work in common with biologists (INSERM 842, neurooncologie et neuro inflammation, Lyon), with clinicians (Creatis, Umr 5515, Inserm U 630 Lyon).

4.1.2. Free radicals

A first model of free radical synthesis has been initiated by V. Lemesle (post doc of ANR AVC in silico). This model is under development with P. Vigneaux. A collaboration begins with Michel Plotkine (EA 2510 pharmacology of cerebral blood flow, Paris 5 university) to get experimental data on the temporal evolution of the various variables of the model.

This model will be used to manage in silico experiments in order to study the effects of various drugs.

4.1.3. Ionic motions

A mechanistic model of ionic motions has already been developed, studied and validated to study in silico the dual role of astrocytes during ischemia, and to study the effect of various ionic channels blockers in man and rodent.

This model is now used to study in silico the effect of the combination of several neuroprotectors acting on ionic channels, transporters or receptors. This work should help to understand antagonist or synergic effects of blockers.

4.1.4. Spreading depression

Spreading depressions are propagative waves which travel in brain during ischemia and which may have a major role in the extension of the ischemic core. Currently 3D computations in real geometry are run to study their speed and the role of brain anatomy in their propagation.

4.1.5. Apoptosis during stroke

A collaboration has begun with Christiane Charriaut Marlangue (INSERM U676, Hopital Robert Debré) to study the apoptotic cascade during stroke.

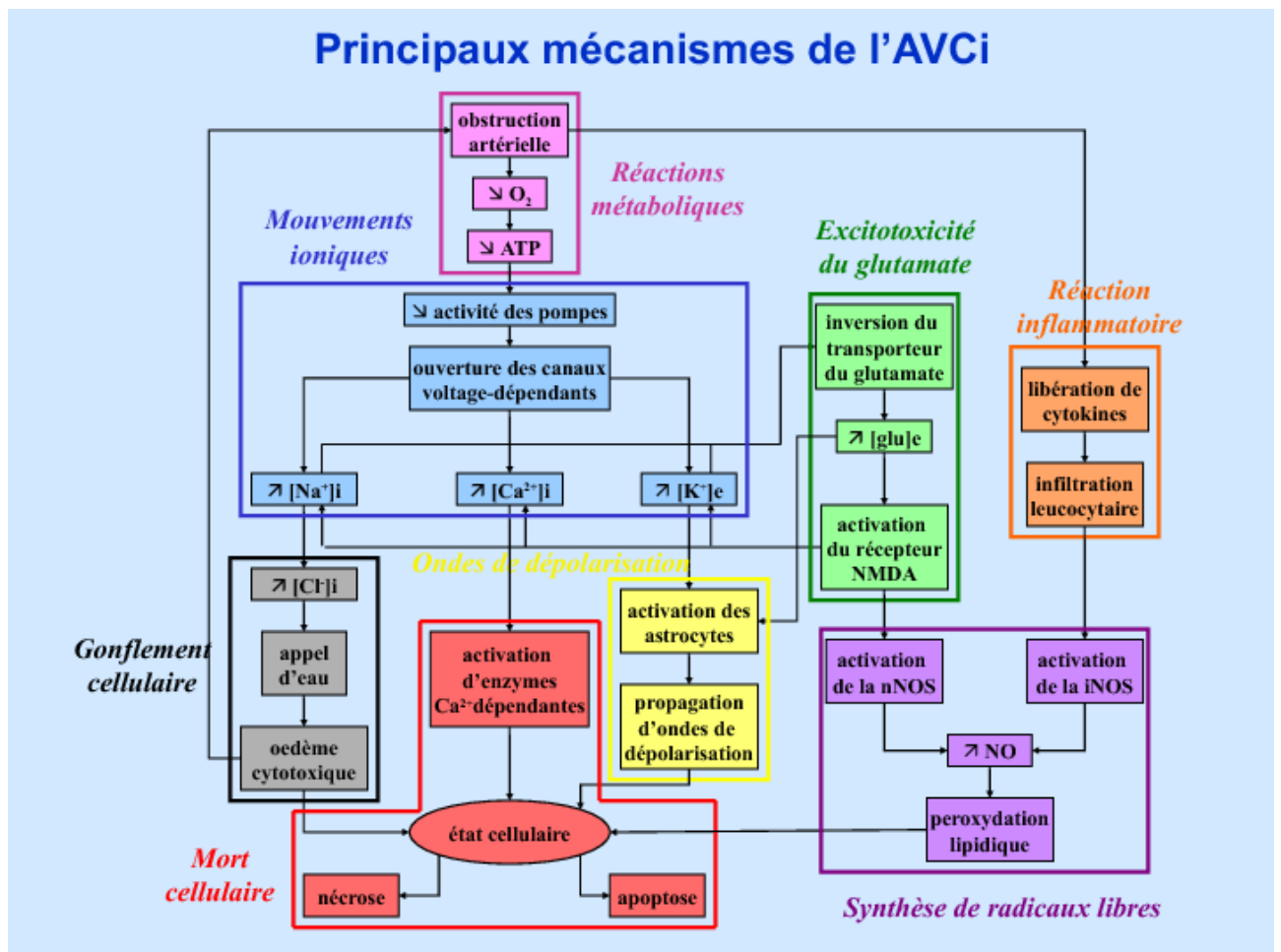


Figure 4. An example of a pdf map reconstructed by using geometrical methods in detecting landmarks

4.2. Oncology

4.2.1. Tumor growth in mice

Through a collaboration with University of Lyon and Lyon-Sud Hospital, we setup several mechanistic models to predict the evolution of tumor growth in mice including the complex biological process of angiogenesis. This work was presented at the eighteen PAGE (population approach group in Europe) meeting in Saint-Petersburg in June.

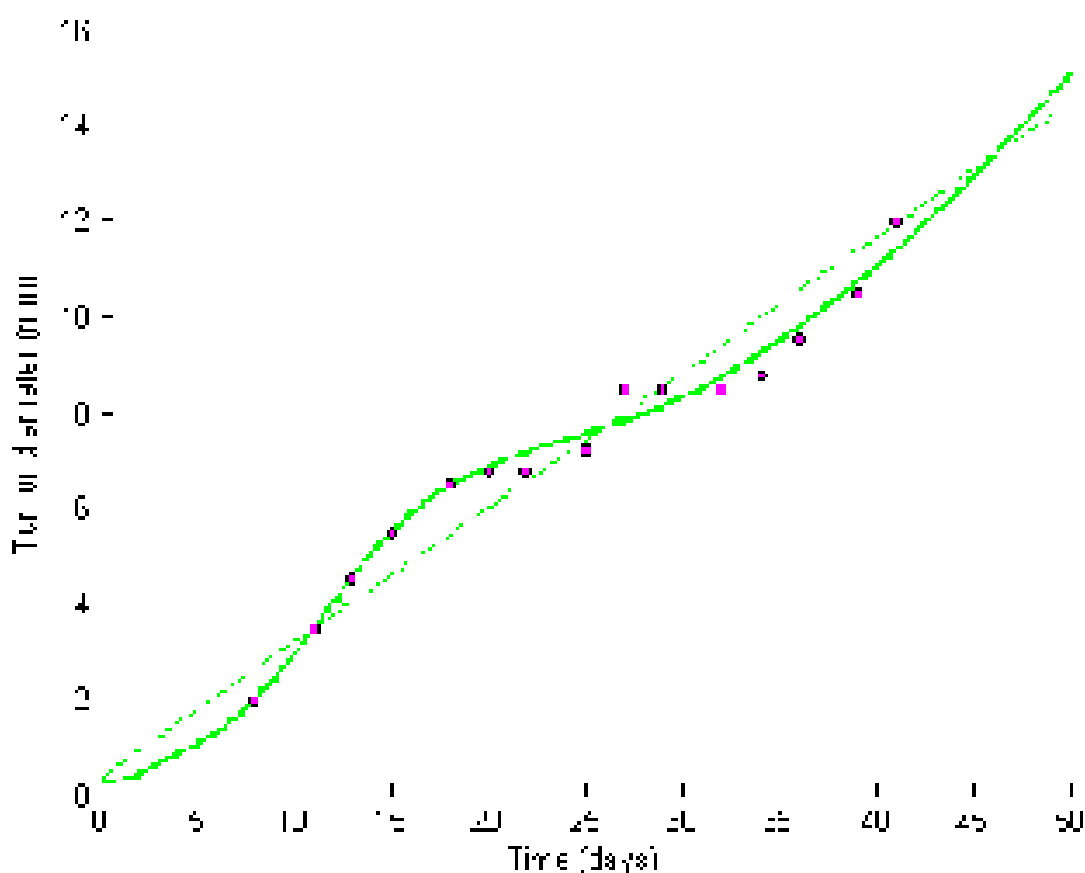


Figure 5. Modeling tumor growth in mice

4.2.2. Radioresistance

Within the framework of the project ETOILE, B. Bernard is part of the team that is building a predictive model of tumor responses to the conventional treatment and irradiation with carbon ions. During 2009, Branka has been working under supervision of Jean-Pierre Boissel and Benjamin Ribba. In collaboration with the group of Claire Rodriguez-Lafresse (Radiobiology group, Hospital Lyon Sud), her research activities included the

analysis of microarray data from different head and neck cancer cell lines, irradiated with X-ray and carbon ions. They detected differences in the irradiation response of different cancer cell lines that underlie their different radiosensitivities. Within GRAAL project, a lot of radiobiological information will be acquired on a several glioma cell lines and cell lines representing healthy brain tissue. Therefore, our interest is to model the dynamics of the glioma tumor growth and its response to radiation therapy. At the moment, we are working on the estimation of parameters describing tumor growth and diffusion from the MRI images of glioblastoma patients (collaboration with Francois Ducray).

4.3. Virology

In collaboration with Merial SA and Edouard Herriot Hospital in Lyon, B. Ribba develops mathematical models to describe the dynamic of Feline immunodeficiency virus (FIV) in infected cats. A translational approach is developed in the context of parameter estimation for complex biologically-based model. He intensively uses mixed-effect modeling approaches and its SAEM algorithm implementation in MONOLIX (Inria Saclay).

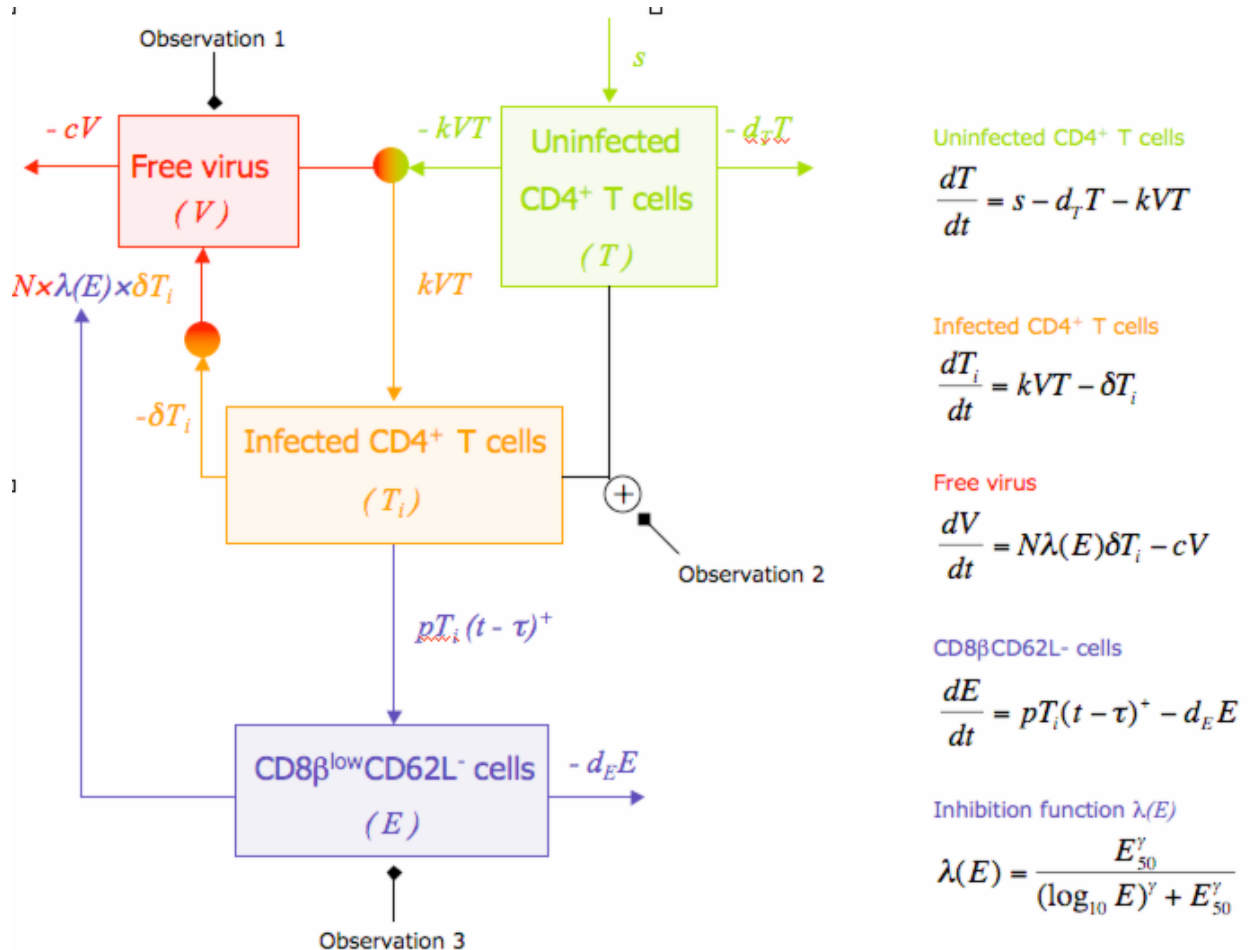


Figure 6. Model of FIV

4.4. Prion.

[In collaboration with M. Doumic (Inria Rocq.), P. Gabriel and B. Perthame (Paris 6) – ANR TOPPAZ]

We study mathematically and numerically the polymerization/fragmentation equation involved in prion aggregation. We have investigated first the case of a size-dependent polymerization rate motivated by recent experiments. We now focus on some issue in optimization of protocol. This is closely related to recent challenges in fitness optimization, and optimal control.

4.5. Atheroma

[In collaboration with N. Meunier (Paris 5)]

Following El Khatib et al. (2007) we have proposed a mathematical model for the inflammatory processes driving the growth of early atherosclerotic plaques. This model is coupled with blood flow, with particular emphasis on the influence of shear stress.

PARIETAL Project-Team

4. Application Domains

4.1. Inverse problems in Neuroimaging

Many problems in neuroimaging can be framed as forward and inverse problems. For instance, the neuroimaging *inverse problem* consists in predicting individual information (behavior, phenotype) from neuroimaging data, while an important the *forward problem* consists in fitting neuroimaging data with high-dimensional (e.g. genetic) variables. Solving these problems entails the definition of two terms: a loss that quantifies the goodness of fit of the solution (does the model explain the data reasonably well ?), and a regularization schemes that represents a prior on the expected solution of the problem. In particular some priors enforce some properties of the solutions, such as sparsity, smoothness or being piecewise constant.

Let us detail the model used in the inverse problem: Let X be a neuroimaging dataset as an (n_{subj}, n_{voxels}) matrix, where n_{subj} and n_{voxels} are the number of subjects under study, and the image size respectively, Y an array of values that represent characteristics of interest in the observed population, written as (n_{subj}, n_f) matrix, where n_f is the number of characteristics that are tested, and β an array of shape (n_{voxels}, n_f) that represents a set of pattern-specific maps. In the first place, we may consider the columns Y_1, \dots, Y_{n_f} of Y independently, yielding n_f problems to be solved in parallel:

$$Y_i = X\beta_i + \epsilon_i, \forall i \in \{1, \dots, n_f\},$$

where the vector contains β_i is the i^{th} line of β . As the problem is clearly ill-posed, it is naturally handled in a regularized regression framework:

$$\hat{\beta}_i = \operatorname{argmin}_{\beta_i} \|Y_i - X\beta_i\|^2 + \Psi(\beta_i), \quad (4)$$

where Ψ is an adequate penalization used to regularize the solution:

$$\Psi(\beta; \lambda_1, \lambda_2, \eta_1, \eta_2) = \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 + \eta_1 \|\nabla\beta\|_1 + \eta_2 \|\nabla\beta\|_2^2 \quad (5)$$

with $\lambda_1, \lambda_2, \eta_1, \eta_2 \geq 0$. In general, only one or two of these constraints will have a non-zero weighting:

- When $\lambda_1 > 0$ only (LASSO), and to some extent, when $\lambda_1, \lambda_2 > 0$ only (elastic net), the optimal solution β is (possibly very) sparse, but may not exhibit a proper image structure; it does not fit well with the intuitive concept of a brain map.
- Total Variation regularization (see Fig. 1) is obtained for $(\eta_1 > 0)$ only), and typically yields a piecewise constant solution.
- Smooth lasso is obtained with $(\eta_2 > 0)$ and $\lambda_1 > 0$ only), and yields smooth, compactly supported spatial basis functions.

The performance of the predictive model can simply be evaluated as the amount of variance in Y_i fitted by the model, for each $i \in \{1, \dots, n_f\}$. This can be computed through cross-validation, by *learning* $\hat{\beta}_i$ on some part of the dataset, and then estimating $(Y_i - X\hat{\beta}_i)$ using the remainder of the dataset.

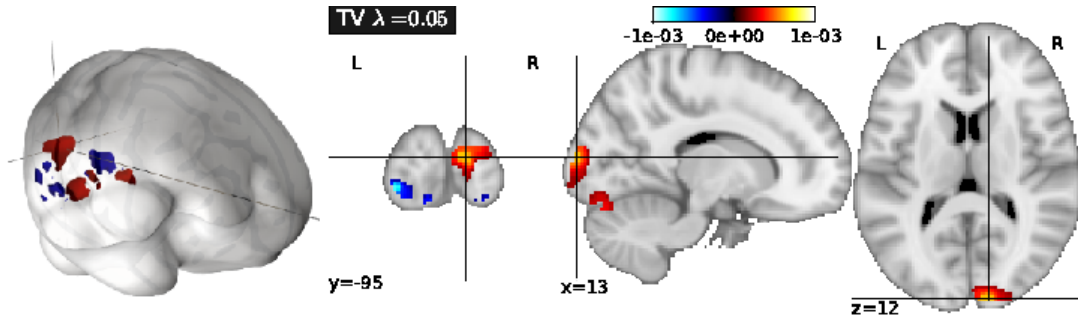


Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problem here consists in predicting the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the observation of an image. Learning and test are performed across individuals. Unlike other approaches, Total Variation regularization yields a sparse and well-localized solution that enjoys particularly high accuracy.

This framework is easily extended by considering

- *Grouped penalization*, where the penalization explicitly includes a prior clustering of the features, i.e. voxel-related signals, into given groups. This is particularly important to include external anatomical priors on the relevant solution.
- *Combined penalizations*, i.e. a mixture of simple and group-wise penalizations, that allow some variability to fit the data in different populations of subjects, while keeping some common constraints.
- *Logistic regression*, where a sigmoid non-linearity is applied to the linear model so that it yields a probability of classification in a binary classification problem.
- *Robustness to between-subject variability* is an important question, as it makes little sense that a learned model depends dramatically on the particular observations used for learning. This is an important issue, as this kind of robustness is somewhat opposite to sparsity requirements.
- *Multi-task learning*: if several target variables are thought to be related, it might be useful to constrain the estimated parameter vector β to have a shared support across all these variables. For instance, when one of the variables Y_i is not well fitted by the model, the estimation of other variables $Y_j, j \neq i$ may provide constraints on the support of β_i and thus, improve the prediction of Y_i . Yet this does not impose constraints on the non-zero parameters of the parameters β_i .

$$Y = X\beta + \epsilon, \quad (6)$$

then

$$\hat{\beta} = \operatorname{argmin}_{\beta=(\beta_i), i=1..n_f} \sum_{i=1}^{n_f} \|Y_i - X\beta_i\|^2 + \lambda \sum_{j=1}^{n_{\text{voxels}}} \sqrt{\sum_{i=1}^{n_f} \beta_{i,j}^2} \quad (7)$$

4.2. Multivariate decompositions

Multivariate decompositions are an important tool to model complex data such as brain activation images: for instance, one might be interested in extracting an atlas of brain regions from a given dataset, such as regions depicting similar activities during a protocol, across multiple protocols, or even in the absence of protocol (during resting-state). These data can often be factorized into spatial-temporal components, and thus can be estimated through *regularized Principal Components Analysis* (PCA) algorithms, which share some common steps with regularized regression.

Let X be a neuroimaging dataset written as an (n_{subj}, n_{voxels}) matrix, after proper centering; the model reads

$$X = AD + \epsilon, \quad (8)$$

where D represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}) , and A the associated subject-wise loadings. While traditional PCA and independent components analysis are limited to reconstruct components D within the space spanned by the column of X , it seems desirable to add some constraints on the rows of D , that represent spatial maps, such as sparsity, and/or smoothness, as it makes the interpretation of these maps clearer in the context of neuroimaging.

This yields the following estimation problem:

$$\min_{D,A} \|X - AD\|^2 + \Psi(D) \text{ s.t. } \|A_i\| = 1 \forall i \in \{1..n_f\}, \quad (9)$$

where (A_i) , $i \in \{1..n_f\}$ represents the columns of A . Ψ can be chosen such as in Eq. (2) in order to enforce smoothness and/or sparsity constraints.

The problem is not jointly convex in all the variables but each penalization given in Eq (2) yields a convex problem on D for A fixed, and conversely. This readily suggests an alternate optimization scheme, where D and A are estimated in turn, until convergence to a local optimum of the criterion. As in PCA, the extracted components can be ranked according to the amount of fitted variance. Importantly, also, estimated PCA models can be interpreted as a probabilistic model of the data, assuming a high-dimensional Gaussian distribution (probabilistic PCA).

4.3. Covariance estimation

Another important estimation problem stems from the general issue of learning the relationship between sets of variables, in particular their covariance. Covariance learning is essential to model the dependence of these variables when they are used in a multivariate model, for instance to assess whether an observation is aberrant or not or in classification problems. Covariance learning is necessary to model latent interactions in high-dimensional observation spaces, e.g. when considering multiple contrasts or functional connectivity data.

The difficulties are two-fold: on the one hand, there is a shortage of data to learn a good covariance model from an individual subject, and on the other hand, subject-to-subject variability poses a serious challenge to the use of multi-subject data. While the covariance structure may vary from population to population, or depending on the input data (activation versus spontaneous activity), assuming some shared structure across problems, such as their sparsity pattern, is important in order to obtain correct estimates from noisy data. Some of the most important models are:

- **Sparse Gaussian graphical models**, as they express meaningful conditional independence relationships between regions, and do improve conditioning/avoid overfit.
- **Decomposable models**, as they enjoy good computational properties and enable intuitive interpretations of the network structure. Whether they can faithfully or not represent brain networks is an important question that needs to be addressed.
- **PCA-based regularization of covariance** which is powerful when modes of variation are more important than conditional independence relationships.

Adequate model selection procedures are necessary to achieve the right level of sparsity or regularization in covariance estimation; the natural evaluation metric here is the out-of-samples likelihood of the associated Gaussian model. Another essential remaining issue is to develop an adequate statistical framework to test differences between covariance models in different populations. To do so, we will consider different means of parametrizing covariance distributions and their impact on the network. Our current work on post-stroke patients (see e.g. Fig. 2) suggests indeed that modeling may prove essential to perform sensitive inference.

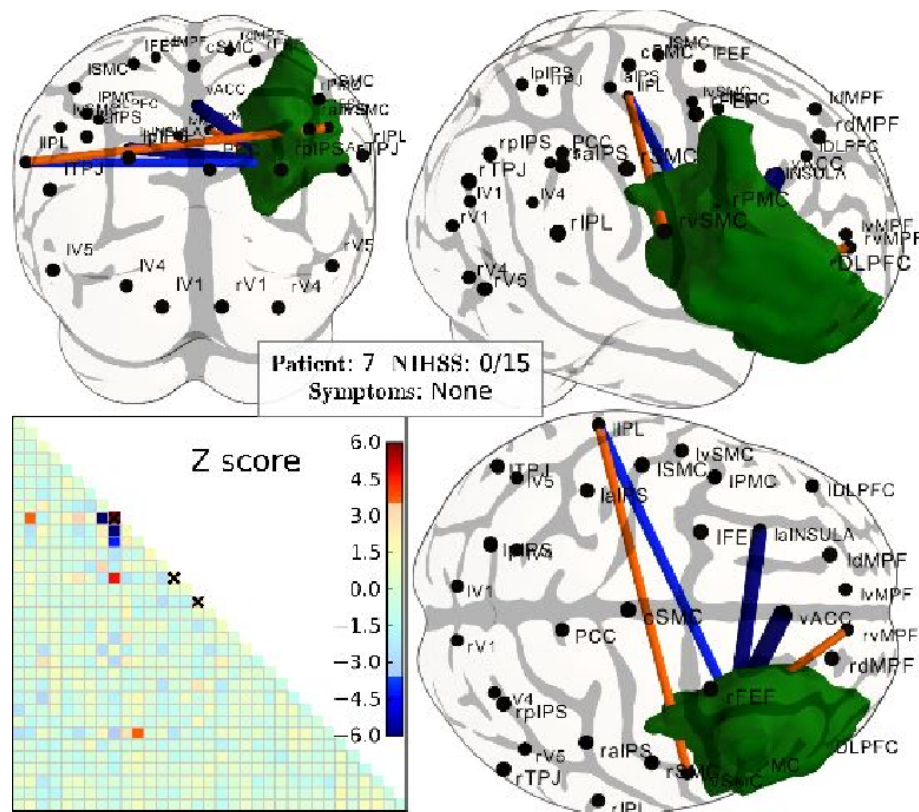


Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion outlined in green) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical detection of the difference requires a sophisticated statistical framework for the comparison of graphical models.

POMDAPI Project-Team (section vide)

REO Project-Team

4. Application Domains

4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport through the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 ml.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments. This activity has been carried out in the framework of successive research programs: RNTS “R-MOD” until 2005, ACI “le-poumon-vous-dis-je” until 2007 and ANR M3RS until 2013.

4.3. Cardiac electrophysiology

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [71], [75], [74] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). We are now working on using our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macsproject-team, we are working on the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology. In particular, we collaborated with ELA Medical company (pacemaker manufacturer, Sorin group).

SAGE Project-Team

4. Application Domains

4.1. Geophysics

The team has chosen a particular domain of application, which is geophysics. In this domain, many problems require solving large scale systems of equations, arising from the discretization of coupled models. Emphasis is put on hydrogeology, but the team also investigates geodesy, heat and mass transfer in soil, and granular materials. One of the objectives is to use high performance computing in order to tackle 3D large scale computational domains with complex physical models.

4.2. Hydrogeology

This is joint work with Geosciences Rennes, University of Poitiers and CDCSP at University of Lyon. It is also done in the context of the group Momas and Andra grant.

Many environmental studies rely on modelling geo-chemical and hydrodynamic processes. Some issues concern water resources, aquifer contamination, underground waste disposal, clean-up of former waste deposits, acid mine drainage remediation. Other issues, also related to energy, concern geothermy, unconventional gas, enhanced oil recovery, underground storage of CO₂, underground storage of nuclear waste.

Simulation of contaminant transport in groundwater is a highly complex problem, governed by coupled linear or nonlinear PDAEs. Moreover, due to the lack of experimental data, stochastic models are used for dealing with heterogeneity. The main objective of the team is to design and to implement efficient and robust numerical models, including Uncertainty Quantification methods.

Recent research showed that rock solid masses are in general fractured and that fluids can percolate through networks of inter-connected fractures. Fractured media are by nature very heterogeneous and multi-scale, so that homogenisation approaches are not relevant. The team develops a numerical model for fluid flow and contaminant transport in three-dimensional porous fractured media.

An important output is the parallel scientific platform H2OLab, running on clusters, grids and machines available in supercomputing centers.

SERPICO Team

4. Application Domains

4.1. Biological pilot models: Birbeck granule and Melanosome biogenesis

In the past recent years, research carried at UMR 144 CNRS Institut Curie contributed to a better understanding of the intracellular compartmentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight:

- multiple sorting and structural events involved in the biogenesis of these organelles;
- complexity of the endo-melanosomal network of these highly specialized cells;
- complex molecular architecture organizing and coordinating their dynamics;
- intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and within the overall biological environment as seen at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link in a single workflow, dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

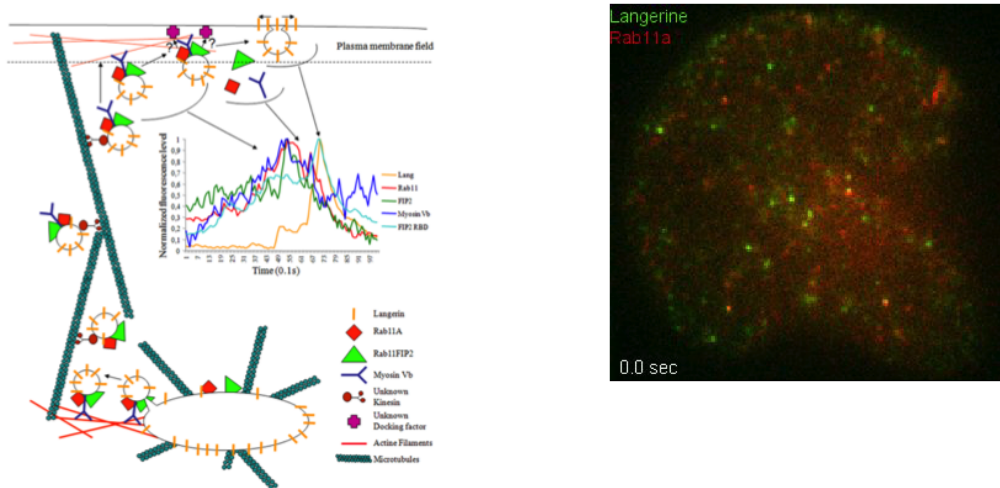


Figure 1. Rab11/Langerin (TIRF) and Birbeck granules: from interactions and trafficking to membrane biogenesis.

SHACRA Project-Team

4. Application Domains

4.1. Clinical Applications

Some of the scientific challenges described previously can be seen in a general context (such as solving constraints between different types of objects, parallel computing for interactive simulations, etc.) but often it is necessary to define a clinical context for the problem. This is required in particular for defining the appropriate assumptions in various stages of the biophysical modeling. It is also necessary to validate the results. This clinical context is a combination of two elements: the procedure we attempt to simulate and the objective of the simulation: training, planning or per-operative guidance. Below are a series of applications we plan to develop. The choice of these applications is not random: the clinical procedures we target are all technically challenging, they highlight various parts of our research, and often they represent an ideal testbed for transitioning from training to planning to guidance. It is important also to note that developing these applications raises many challenges and as such this step should be seen as an integral part of our research. It is also through the development of these applications that we can communicate with physicians, and validate our results. SOFA will be used as a backbone for the integration of our research into clinical applications.

4.1.1. *Interventional radiology*

Over the past twenty years, interventional methods such as angioplasty, stenting, and catheter-based drug delivery have substantially improved the outcomes for patients with vascular disease. Pathologies that used to require a surgical procedure can now be treated in a much less invasive way. As a consequence, interventional radiology procedures represent an increasing part of the interventions currently performed, with more than 6 million patients treated every year in Europe and about 5 millions the United States. However, these techniques require an intricate combination of tactile and visual feedback, and extensive training periods to attain competency. To reinforce the need to reach and maintain proficiency, the FDA recently required that US physicians go through simulation-based training before using newly developed carotid stents. Besides simulation for training, interventional radiology is a perfect target to illustrate the potential of planning and rehearsal of procedures. As an initial step in this direction, Alcove and Magrit were partners in an ARC project (Simple) to develop a planning tool for the treatment of aneurysms using coils. This collaboration still goes on after the end of the ARC, and led to a series of papers in key conferences [5] [28], [34], [21].

4.1.1.1. *Interventional neuro-radiology*

We will continue the development of our simulation and planning system for interventional radiology, with two principal clinical partners: Massachusetts General Hospital in Boston and University Hospital in Nancy. We have completed the integration in SOFA of improved versions of algorithms for describing the behavior of catheters, guide-wires, coils, as well as the interactive simulation of fluoroscopic images, the modeling of complex contacts. Our future efforts will focus on the development of an advanced planning system for interventional radiology, in particular for coil embolization. This will require the integration of new methods of reconstruction of vascular anatomy from medical images (in collaboration with the MAGRIT team). We will also add our recent results on blood flow simulation in aneurysms.

4.1.1.2. *Interventional cardiology using radio-frequency ablation*

Cardiac arrhythmias (or dysrhythmias) are problems that affect the electrical system of the heart muscle, producing abnormal heart rhythms, and causing the heart to pump less effectively. About 5% of people over 40 years old are affected by this pathology, with a rather high morbidity rate. Radio-frequency ablation is a non-surgical procedure that has been used for about 15 years to treat tachyarrhythmias, i.e. rapid, uncoordinated heartbeats. The procedure is performed by guiding a catheter with an electrode at its tip to the area of heart muscle where there is an accessory pathway. The catheter is guided under fluoroscopic imaging. When the catheter is positioned at the site where cells give off the electrical signals that stimulate the abnormal heart

rhythm, a low radio-frequency energy is transmitted to the pathway. This destroys heart muscle cells within a very small area near the tip of the catheter and stops the area from conducting the extra impulses that caused the arrhythmia. In this context, a simulation system would be able to provide added value in two main areas: 1) to train physicians in the early stages of their apprenticeship and 2) to provide quantitative information during the planning phase of a complex procedure, using patient-specific data. Most aspects of this simulation will rely on components developed during our research program but we will also extend our collaboration with the ASCLEPIOS team and the CardioSense3D project on the modeling of the heart the Cadiosense3D project. This involves an important integration task, and it will also validate the reusability aspects of the code developed within SOFA.

4.1.2. Minimally-invasive surgery

4.1.2.1. Laparoscopic hepatic resection

The liver is one of the major organs in the human body. It is in charge of more than 100 vital functions. Because of its many functions, its pathologies are also varied, numerous and unfortunately often lethal. This is for instance the case of hepatitides which today affect about 300,000 people in France for hepatitis B and 600,000 people for hepatitis C. The most advanced state of evolution of these pathologies is generally cirrhosis followed by cancer, which represents the third cause of cancer related death. In 2005, 14,267 liver cancer cases and 20,497 cirrhosis cases have been diagnosed in France. The surgical solution remains the option offering the best success rate for these pathologies. More than 7,000 surgical interventions have been carried out on the liver in 2005 and partial resection of the liver remains the most common approach. In this context, the ability to train surgeons, and to be able to plan complex procedures using computer-based simulations, would be a formidable help to the current apprenticeship model: “See One, Do One, Teach One”. Right now, only a few commercial systems are available to the medical community, and they are limited to basic skills training. Developing a realistic simulation system that could be used to plan and rehearse procedures would be a very important step in the introduction of new training paradigms in medicine. This is the main objective of the PASSPORT european project in which we are actively contributing at two levels. First, our research results on biomechanical modeling of solid organs and on coupling will be used to propose a realistic model of the deformation of the liver and its vascular network. Second, SOFA has been chosen in this project as the software for integrating all results from the different partners. Both aspects will help validate our models, test SOFA and obtain feedback from the clinicians.

4.1.2.2. Ophthalmology and cataract surgery

A cataract is an opacity in the natural lens of the eye. It represents an important cause of visual impairment and, if not treated, can lead to blindness. It is actually the leading cause of blindness worldwide, and its development is related to aging, sunlight exposure, smoking, poor nutrition, eye trauma, and certain medications. The best treatment for this pathology remains surgery. Cataract surgery has made important advances over the past twenty years, and in 2005, more than 5 million people in the United States and in Europe underwent cataract surgery. Most cataract surgeries are performed using microscopic size incisions, advanced ultrasonic equipment to fragment cataracts into tiny fragments, and foldable intraocular lenses to minimize the size of the incision. All these advances benefit the patient, but increase training requirements for eye surgeons. At the end of 2007, we started the development of a new training system for cataract surgery. The main objectives of this simulation are to reproduce with great accuracy the three main steps of cataract surgery: 1) capsulorhexis 2) phacoemulsification and 3) implantation of an intraocular lens. We have already started the development of this simulation. The main research effort went in the choice of appropriate deformable models for the lens and lens capsule. An important effort also went into the development of topological changes corresponding to the capsulorhexis and phacoemulsification [20]. The modeling of the intraocular implant and its deployment in the capsule has been published to the major conference in medical simulation [26].

4.1.2.3. Neurosurgery and deep brain stimulation

Deep brain stimulation (DBS) is a neurosurgical treatment which stimulates the brain with low electrical signals. The signals reorganize the brain’s electrical impulses (similarly as what was presented above for radio-frequency ablation for cardiac problems). This results in major improvements in several pathologies

such as Parkinson disease. The principle of the procedure is the following: a thin, insulated wire lead with several electrodes at the tip is surgically implanted into the affected area of the brain. A wire runs under the skin to a battery-operated pulse generator implanted near the collarbone. The generator is programmed to send continuous electrical pulses to the brain. To implant the electrodes, a neurosurgeon uses a stereotactic head frame and magnetic resonance or computed tomography imaging to map the brain and pinpoint the problem area. The main difficulty in this procedure comes from the deformation of the brain (small brain shift when the skull is opened, and local deformation of the brain due to the insertion of the electrode) and the deflection of the electrode itself during and after the procedure. This results in a difference between the planned target and the location of the end effector of the electrode. Our main objective is to use our work on soft tissue deformation, vascularized structures, as well as our recent results on constraint solving between soft tissues and flexible devices [29]. This work will be done in collaboration with the VISAGES team and we will dedicate an important effort in validating our results, analyzing post-operative medical images, and interacting with surgeons. This project has a strong potential as DBS is being increasingly used yet most research groups only consider non deformable planning systems (geometrical planning). Our proposal could make a important difference in the accuracy of the planning as it takes into account the biophysics of the brain.

SISYPHE Project-Team (section vide)

STEEP Exploratory Action

4. Application Domains

4.1. Urban economy and land use and transport modeling

Modern urban regions are highly complex entities. The understanding of the phenomena underlying urban sprawl and peri-urbanization is a key element to control the dynamics structuring urban space. Clearly, urban transport systems are intricately linked to urban structure and the distribution of activities, i.e., to land use. Urbanization generally implies an increase in travel demand. Cities have traditionally met this additional demand by expanding the transportation supply, through new highways and transit lines. In turn, an improvement of the accessibility of ever-farther land leads to an expansion of urban development, resulting in a significant feedback loop between transportation infrastructure and land use, one of the main causes of urban sprawl.

Several models have been developed in the field of urban economics to understand the complex relationship between transportation and land use and to facilitate the urban planning process. They enable the simulation of public policies and the quantification of indicators describing the evolution of urban structure. Key factors such as transport congestion, energy consumption, CO₂ emissions etc., can be evaluated or estimated, and different urban development scenarios can be tested in a quantitative manner.

Yet, very few local authorities in charge of planning issues make use of these strategic models, mostly because they are difficult to calibrate and validate, two critical steps where systematic improvement would increase the level of confidence in the obtained results. These limitations prevent dissemination in local agencies. One goal of STEEP is therefore to meet the need of better calibration and validation strategies and algorithms. This research is the core of our projects CITIES (ANR Modèles Numériques) and TRACER (Ecos Nord Venezuela).

4.2. Ecological accounting and material flow analysis

One of the major issues in the assessment of the long-term sustainability of urban areas is related to the concept of “imported sustainability”. Indeed, any city brings from the outside most of its material and energy resources, and rejects to the outside the waste produced by its activity. The modern era has seen a dramatic increase in both volume and variety of these material flows and consumption as well as in distance of origin and destination of these flows, usually accompanied by a spectacular increase in the associated environmental impacts. A realistic assessment of the sustainability of urban areas requires to quantify both local and distant environmental impacts; greenhouse gas emissions are only one aspect of this question.

In order to produce such an assessment for a given territory or urban area, one must first establish different types of ecological accounting: one must identify and quantify the different types of material and energy uses on the one hand, and the different types of impact associated to these uses. The first task is the object of Material Flow Analysis (MFA), while the second is more directly related to the logic of Life Cycle Analysis (LCA). One of the major challenges here is to obtain reliable MFA data at the region and *département* scales, either directly, or through appropriate disaggregation techniques.

The STEEP group has started a research program on this theme with three major aims in mind: 1) Creating a comprehensive database enabling such analyses; 2) Developing methodology and models resolving scaling issues, and developing algorithms allowing to rigorously and automatically obtain the adequate assessments; 3) Providing a synthetic analysis of environmental impacts associated to the major material flows, at various geographic levels (employment catchment area, *département* and *région*, for France). The database elaboration is already well underway, and apparently does not yet have any equivalent elsewhere in France. The impact evaluation and decision help strategy (i.e., evaluating alternative policy options in term of environmental impact) will be developed in collaboration with a private company ARTELIA, through a CIFRE PhD thesis that is scheduled to start early 2013, and will be based on existing LCA databases. The PhD student is co-directed with Denis Dupré from CERAG (Centre d’Etudes et de Recherches Appliquées à la Gestion).

4.3. Eco-system services

Long-term sustainability is closely related to the underlying ecosystems, on various fronts: production of renewable resources (either energy or biomass), waste and pollutant resorption, local and global climate regulations etc. These various functions constitute the “ecosystem services” provided to society by our natural environment.

The reduction of the adverse impacts of urban areas on the environment is linked not only to limiting urban sprawl and making more efficient use of the available resources, but also to developing a better grasp of the interrelations between urban/peri-urban areas and their agricultural and semi-natural surroundings. In particular, reducing distant impacts while making a better use of local resources is a major challenge for the coming decades.

In this context, the STEEP team is involved in a project bearing on the characterization of local Ecosystem Services NETworks (ESNET), piloted by LECA (*Laboratoire d'Ecologie Alpine*), and in collaboration with a number of other research laboratories (most notably, IRSTEA Grenoble) and a panel of local stakeholders; the scale of interest is typically a landscape (in the ecologic/geographic sense, i.e., a zone a few kilometers to a few tens of kilometers wide). The project aims at developing a generic modelling framework of ecosystem services, and studying their behavior under various scenarios of coupled urban/environment evolution. The contribution of the STEEP team is centered on the Land Use/Land Cover Change (LUCC) model that will be one of the major building blocks of the whole model, with the help of specifically hired personal.

The project has been supported by FRB (*Fondation pour la Recherche en Biodiversité*) and will be funded by ONEMA (*Office National de l'Eau et des Milieux Aquatiques*) for the three years of its duration.

VIRTUAL PLANTS Project-Team (section vide)

VISAGES Project-Team

4. Application Domains

4.1. Neuroimaging

neuroimaging, clinical neuroscience, multiple sclerosis, multispectral MRI, brain atlas

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions.

Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modelling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define *in vivo* those four different MS lesion patterns. An *in vivo* distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. Since MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types, but a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and grey matter, our major objective within this application domain is to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g. nD+t MRI). In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labelling neuroimaging (e.g. MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g. USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved *in vivo* characterization of MS pathology.

4.3. Modelling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field [52], [54], [53], [55], and by continuing our local and wider collaborations.

An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g. variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

ACES Project-Team (section vide)

ADAM Project-Team

4. Application Domains

4.1. Introduction

The ADAM project-team targets the software engineering of adaptive service-oriented applications and middleware. The application domain covered by ADAM is broad and spans from distribution applications to middleware. In all these cases, adaptability is the property which is sought: applications and middleware must be adaptable to new execution contexts, they must react to changes in the environment and they must be able to discover and integrate new services.

The ADAM project-team produces software and middleware building blocks. This explains why the application domain is broad, yet targeting applications where adaptability is the key requirement. This includes electronic commerce, embedded systems, health care information systems, and terrestrial transport information systems. These domains are in direct relation with our currently funded activities. They act as testbeds for the solutions that we propose in terms of middleware services, middleware platforms, runtime kernels, component libraries, languages design or domain modeling.

4.2. Electronic Commerce

Applications in the domain of electronic commerce are by essence distributed. They involve many different participants with heterogeneous information systems which cannot be changed. The challenge is then to provide an adaptation layer to be able to compose and let these systems interoperate. In the context of the ANR TLog SCOrWare, the ICT SOA4All and the FUI CAPPUCINO projects, our activities in this domain will aim at supporting service-oriented architectures. We want to have adaptive architectures that can be composed and orchestrated seamlessly. In this domain, the business relationship with customers is vital and many different usage scenarios must be supported. Customers are roaming, and the services must be kept operational across different devices. This puts some constraints on the server tier where technical services must be adapted to manage, for instance, long lasting transactions. The application server infrastructure must then provide a support for adapting technical services.

4.3. Embedded Systems

Embedded systems form a domain where adaptation is a key requirement. The design and the implementation of modern embedded software uses advanced software engineering techniques such model-driven development or software component frameworks. In this domain, we are involved in several projects, such as the ANR TLog Flex-eWare, and the FUI MIND projects. Several challenges must be addressed here. For example, when a model-driven developed application is adapted, designers have to ensure that the models and the operational level are kept synchronized. The co-evolution of these two levels is one of the challenges that we are addressing. A second challenge is related to software components that need to be customized in order to fit the requirements imposed by constrained environments. It is, for example, a matter of providing component frameworks that can accommodate various granularities of services.

4.4. Health Care Information Systems

Health care information systems form a third application domain in which the ADAM project-team is involved, for instance through demonstrators which will be implemented in the context of the ANR TLog FAROS project. The challenge here is to provide a distributed infrastructure where information will be available to medical staff wherever they are. This imposes to be able to provide this information on many different devices (from high resolution screens to embedded devices on the scene of an accident), while ensuring the privacy of the medical data of a patient (several level of data access must be granted depending on the categories of medical staff). Given the vital role of such an information system, we want to provide guarantees that the services will be highly available and trustworthy. We envision to provide a service-oriented architecture which will be extended to support software contracts and multi-scale environments.

4.5. Information Systems for Terrestrial Transport

Information systems for terrestrial transport are also a domain that we are relying on, to apply our research activities in accordance with the ANR ARA REVE project and the INRETS collaboration. Applications are here characterized by frequent disconnections, poor quality network links, and high mobility. We want to provide an infrastructure where the technical services, and among others the communication services, can be adapted to support new requirements. One of the paths that we propose to investigate is to include such a scenario in the general context of the adaptiveness of component frameworks.

ALGORILLE Project-Team

4. Application Domains

4.1. Scientific computation

In the context of studying GPU cluster programming as well as asynchronous algorithms, we have developed a parallel code to solve a PDE problem that is representative of scientific computations. This is a 3D version of the *advection-diffusion-reaction* problem modeling the interactions of two chemical species in shallow waters. Different versions of the application have been implemented. A first version has been the subject of extensive studies about computing and energy performance. Other versions, especially the last one using MPI+OpenMP+CUDA, are still the subject of studies.

Also in a close collaboration with Fatmir Asllanaj from the LEMTA laboratory, Nancy, France, we launched the development of a code for Radiative Transport Equations, ETR. As a first step to higher efficiency the code (originally in FORTRAN) was completely rewritten in C. Already this sequential rewrite has largely improved the time and memory efficiency of the code. The next step will be a parallelization for multi-core machines and clusters that will make this a unique tool to tackle ETRs in 2 and 3 dimensions.

4.2. Financial computation

AmerMal application is a parallel *American Option Pricer* developed in collaboration with Lokman Abbas-Turki during his PhD thesis at University Paris-Est. This pricing mathematic algorithm has been designed in order to be parallelized on clusters of GPUs. It is based on a Maillavin calculus and on Monte Carlo stochastic computations. However, the Monte Carlo trajectories are coupled, and our final parallel algorithm includes many communications between the computing nodes. Nevertheless, we achieved good scalability on our 16-nodes GPU clusters.

The initial version of AmerMal has been implemented using MPI+OpenMP+CUDA in 2010 and 2011. In 2012, we have ported this application on ORWL+CUDA, both in order to test and improve ORWL semantic and ORWL performances.

4.3. Signal processing

ParSONS is a parallel *sound sources classifier* developed at SUPELEC by Stephane Rossignol. It allows to identify different kinds of sound sources in an audio signal file (like a radio or TV archive file), and to tag different part of the file. Then it is possible to know which part of the archive file include human voice, or music, or more accurately jazz or rock music. From an algorithmic point of view, this application performs a lot of Fourier Transform computations and a lot of IO operations.

We have developed an MPI+OpenMP version, optimizing input data file reading and overlapping these IO operations with computations (signal processing). We run some experiments on a 256-nodes dual-core PC cluster and on a 16-nodes 4-hyperthreaded-cores experimentation PC cluster, both located at SUPELEC. We identified the most efficient configurations, considering the number of input file reading processes, the number of threads per process, the number of processes per node... Finally, we have designed and implemented a deployment tool, in order to hide these configuration issues to signal processing researchers and users.

In a near future, we aim to port this application to ORWL, to experiment, improve and validate ORWL on distributed applications achieving a high ratio of IO compared to computations and classic communications.

ARLES Project-Team

4. Application Domains

4.1. Application Domains

The ARLES project-team is interested in the application of pervasive computing, and as such considers various application domains. Indeed, our application domain is voluntarily broad since we aim at offering generic solutions. However, we examine exploitation of our results for specific applications, as part of the experiments that we undertake to validate our research results through prototype implementation. Applications that we consider in particular include demonstrators developed in the context of the European and National projects to which we contribute (§ [7.1](#) & [7.2](#)).

ASAP Project-Team

4. Application Domains

4.1. Overview

The results of the research targeted in ASAP span a wide range of applications. Below are a few examples.

- Personalized Web Search.
- Recommendation.
- Social Networks.
- Notification Systems.
- Distributed Storage.
- Video Streaming.

ASCOLA Project-Team

4. Application Domains

4.1. Enterprise Information Systems and Services

Large IT infrastructures typically evolve by adding new third-party or internally-developed components, but also frequently by integrating already existing information systems. Integration frequently requires the addition of glue code that mediates between different software components and infrastructures but may also consist in more invasive modifications to implementations, in particular to implement crosscutting functionalities. In more abstract terms, enterprise information systems are subject to structuring problems involving horizontal composition (composition of top-level functionalities) as well as vertical composition (reuse and sharing of implementations among several top-level functionalities). Moreover, information systems have to be more and more dynamic.

Service-Oriented Computing (SOC) that is frequently used for solving some of the integration problems discussed above. Indeed, service-oriented computing has two main advantages:

- Loose-coupling: services are autonomous, in that they do not require other services to be executed;
- Ease of integration: Services communicate over standard protocols.

Our current work is based on the following observation: similar to other compositional structuring mechanisms, SOAs are subject to the problem of crosscutting functionalities, that is, functionalities that are scattered and tangled over large parts of the architecture and the underlying implementation. Security functionalities, such as access control and monitoring for intrusion detection, are a prime example of such a functionality in that it is not possible to modularize security issues in a well-separated module. Aspect-Oriented Software Development is precisely an application-structuring method that addresses in a systemic way the problem of the lack of modularization facilities for crosscutting functionalities.

We are considering solutions to secure SOAs by providing an aspect-oriented structuring and programming model that allows security functionalities to be modularized. Two levels of research have been identified:

- Service level: as services can be composed to build processes, aspect weaving will deal with the orchestration and the choreography of services.
- Implementation level: as services are abstractly specified, aspect weaving will require to extend service interfaces in order to describe the effects of the executed services on the sensitive resources they control.

In 2012, we have developed techniques for the Service-Level Agreement (SLA) management for Cloud elasticity, see Sec. 6.3 , as well as models and type systems for service-oriented systems, see Sec. 6.1 . Furthermore, we take part in a starting new European project A4Cloud on accountability challenges, that is, the responsible stewardship of third-party data and computations, see Sec. 8.2 .

4.2. Cluster, Grid and Cloud Computing

Cluster, Grid and more recently Cloud computing platforms aim at delivering a larger capacity of computing power compared to a single computer configuration. This capacity can be used to improve performance (for scientific applications) or availability (e.g., for Internet services hosted by a data center). These distributed infrastructures consist of a group of coupled computers that work together. This group can be spread across a LAN (cluster), across a WAN (Grid), and across the Internet (Clouds). Due to their large scale, these architectures require permanent adaptation, from the application to the system level and calls for automation of the adaptation process. We focus on self-configuration and self-optimization functionalities across the whole software stack: from the lower levels (systems mechanisms such as distributed file systems for instance) to the higher ones (i.e. the applications themselves such as J2EE clustered servers or scientific grid applications).

In 2012, we have developed the DVMS system, which contains one of the most highly scalable scheduling algorithm for virtual machines; we have also generated several results on the energy efficient management of Cloud applications and infrastructures, see Sec. 6.3 .

4.3. Pervasive Systems

Pervasive systems are another class of systems raising interesting challenges in terms of software structuring. Such systems are highly concurrent and distributed. Moreover, they assume a high-level of mobility and context-aware interactions between numerous and heterogeneous devices (laptops, PDAs, smartphones, cameras, electronic appliances...). Programming such systems requires proper support for handling various interfering concerns like software customization and evolution, security, privacy, context-awareness... Additionally, service composition occurs spontaneously at runtime.

In 2012, we have developed the language EScala, which integrates reactive programming through events with aspect-oriented and object-oriented mechanisms, see Sec. 6.1 .

ATLANMOD Team

4. Application Domains

4.1. Introduction

By definition, MDE can be applied to any software domain. Core MDE techniques developed by the team have been successfully applied to a large variety of industrial domains from information systems to embedded systems. MDE is not even restricted to software engineering, but also applies to data engineering [50] and to system engineering [41]. There are a lot of problems in these application domains that may be addressed by means of modeling and model transformation techniques. For the core techniques, we are now focusing on solving the scalability problem to make sure they can be successfully adopted by our industrial partners in the context of large and complex software systems.

Nevertheless, the team has indeed selected a set of vertical and horizontal domains that AtlanMod finds specially interesting. In what follows we describe three of them.

4.2. Reverse Engineering

One important and original domain that is being investigated by the AtlanMod team is the reverse engineering of existing IT systems. We do believe that efficiently dealing with such legacy systems is one of the main challenges in Software Engineering and related industry today. Having a better understanding of these systems in order to document, maintain, improve or migrate them is thus a key requirement for both academic and industrial actors in this area. However, it is not an easy task and it still raises interesting challenging issues to be explored [49].

We have shown how reverse engineering practices may be advantageously revisited with the help of the MDE approach and techniques, applying (as base principle) the systematic representation as models of the required information discovered from the legacy software artifacts (e.g.; source code, configuration files, documentation, metadata, etc). The rise in abstraction allowed by MDE can bring new hopes that reverse engineering is now able to move beyond more traditional ad-hoc practices. For instance, an ongoing industrial PhD in partnership with IBM France aims to investigate the possibilities of conceptualizing a generic framework enabling the extraction of business rules from a legacy application, as much as possible, independently of the language used to code it. Moreover, different pragmatic solutions for improving the overall scalability when dealing with large-scale legacy systems (handling huge data volumes) are intensively studied by the team.

In this context, AtlanMod has set up and is developing accordingly the open source Eclipse-MDT MoDisco project (see 5.2). MoDisco is notably being referenced by the OMG ADM (Architecture Driven Modernization) normalization task force as the reference implementation for several of its standard metamodels.

We have also opened a novel research line focused on integration of APIs into MDE. In the application of reverse engineering processes while modernizing software system it is very common to face the need of integrating Application Programming Interfaces (APIs). Indeed, building any application usually involves managing a plethora of APIs to access different software assets such as: basic infrastructures (e.g., operating system, databases, or middleware), general-purpose or domain specific libraries, frameworks, software components, web services, and even other applications. Thus, to promote the interoperability between the API and model technical spaces, we have developed API2MoL, which is an approach aimed at automating the implementation of API-MDE bridges. This new language allows defining mappings between the artefacts of a given API (e.g., API classes in object-oriented APIs) and the elements of a metamodel that represents this API in the MDE technical space. A mapping definition provides the information which is necessary to build a bridge for a concrete API specification and metamodel. Thanks to the API-MDE bridges automatically created, developers are liberated from having to manually implement the tasks of obtaining models from API objects and generating such objects from models. Therefore, API2MoL may improve the productivity and quality of the part of the MDE application that deals with the APIs.

4.3. Security Engineering

Several components are required to build up a system security architecture, such as firewalls, database user access control, intrusion detection systems, and VPN (Virtual Private Network) routers. These components must be properly configured to provide an appropriate degree of security to the system. The configuration process is highly complex and error-prone. In most organizations, security components are either manually configured based on security administrators expertise and flair; or simply recycled from existing configurations already deployed in other systems (even if they may not be appropriated for the current one). These practices put at risk the security of the whole organization.

We have started a Phd thesis in this domain intended to investigate the construction of a model-driven automatic reverse engineering mechanism (implemented as an extension of the MoDisco project) capable of analyzing deployed security aspects of components (e.g., concrete firewall configurations) to derive the abstract model (e.g., network security global policy) that is actually enforced over the system. Once the model is obtained, it can be reconciled with the expected security directives, to check its compliance, can be queried to test consistency or used in a process of forward engineering to generate validated security configurations.

As a first step we intend to apply model-driven techniques for the extraction of high level model representations of security policies enforced by firewalls. Firewalls, core components in network security systems, are generally configured by using very low level vendor specific rule-based languages, difficult to understand and to maintain. As a consequence, as the configuration files grow, understanding which security policy is being actually enforced or checking if inconsistencies has been introduced becomes a very complex and time consuming task. We propose to raise the level of abstraction so that the user can deal directly with the high level policies. Once a model representation of the enforced policy is available, model-driven techniques will ease some of the tasks we need to perform, like consistency checking, validation, querying and visualization. Easy migration between different firewall vendors will be also enabled.

4.4. Software Quality

As with any type of production, an essential part of software production is determining the quality of the software. The level of quality associated to a software product is inevitably tied to properties such as how well it was developed and how useful it is to its users. The AtlanMod team is considering software quality aspects in two areas that are described in the following paragraphs: formal verification of models and corpus-based domain-specific language (DSL) analysis.

In the context of MDE, models are expressions of a conceptualization (metamodel) of their respective domain. Therefore precise definitions of our metamodels are important to get the subsequent MDE activities right, such as testing, concrete syntax definition, model interchange, code generation, and any more. In this sense, metamodels establish a single point of truth. The Object Constraint Language (OCL) of the OMG is widely accepted as a standard language to provide such precise definitions.

However, precise metamodels are complex artifacts. To get them right, systematic approaches to quality assurance are required. As a solution to this problem, we propose formal checking of model satisfiability (can we express the desired things), and formal checking of unsatisfiability (does the metamodel not have models with undesirable properties). As both activities constitute NP-hard problems, they have to be put carefully into algorithms. The team maintains the tool EMFtoCSP which translates the problem into the domain of constrain logic programming (CLP) for which sophisticated decision procedures exist. The tool integrates the described functionality in the Eclipse Modeling Framework (EMF) and the Eclipse Modeling Tools (MDT), making the functionality available for MDE in practice.

In terms of DSLs, an equally important aspect of software language engineering other than the initial development process is the identification of quality characteristics of the language. We are currently investigating the use of corpus-based analysis to identify language characteristics. A corpus in this case consists of artifacts or models of the DSL that have been generated by its users. We aim to extract information based on the corpus of a DSL to identify various characteristics that can potentially prove useful for the language engineer in his or her efforts to improve the language. Such information can complement other quality measurements including

the formal verification method described in the previous paragraph and user feedback-based evaluations. In addition to an initial investigation of cloning in DSLs, our corpus-based analysis is also investigating metamodel element instance and relationship analyses

AVALON Team

4. Application Domains

4.1. Overview

The Avalon team targets applications with large computing and/or data storage needs, which are yet difficult to program and maintain while achieving performance. Those applications can be parallel and/or distributed applications. It is typically the case of large simulations and/or code coupling applications from “classical” computational problems such as climate forecasting or cosmology. The team also deals with bioinformatics as they raise some interesting research issues linked to data management. For example, some of such applications rely on MapReduce algorithmic skeletons.

4.2. Bioinformatics

Large-scale data management is certainly one of the most important applications of distributed systems of the future. Bioinformatics is a field producing such kinds of applications. For example, DNA sequencing applications make use of MapReduce skeletons.

The MapReduce programming model publicized by Google is a widely used model for deploying application services on platforms for data processing on a large scale (Grids and Clouds). MapReduce implementations extremely are scalable and efficient when located on the datacenters of big companies. However, technological (use of a parallel file system) and algorithmic (hard-coded replication) choices of an implementation such as Hadoop compromise performance on platforms like grids of heterogeneous clusters. Applying it to large scale and volatile environments such as desktop grids still remain a challenge.

4.3. Climate Forecasting Simulations

World’s climate is currently changing due to the increase of the greenhouse gases in the atmosphere. Climate fluctuations are forecasted for the years to come. For a proper study of the incoming changes, numerical simulations are needed, using general circulation models of a climate system.

As for most applications the team is targeting, our goal is to thoroughly analyze climate forecasting applications to model its needs in terms of programming model, execution model, data access pattern, and computing needs. Once a proper model of the application has been derived, appropriate scheduling heuristics could be designed, tested, and compared.

4.4. Cosmological Simulations

*Ramses*¹ is a typical computational intensive application used by astrophysicists to study the formation of galaxies. *Ramses* is used, among other things, to simulate the evolution of a collisionless, self-gravitating fluid called “dark matter” through cosmic time. Individual trajectories of macro-particles are integrated using a state-of-the-art “N body solver”, coupled to a finite volume Euler solver, based on the Adaptive Mesh Refinement technique. The computational space is decomposed among the available processors using a *mesh partitioning* strategy based on the Peano-Hilbert cell ordering.

Cosmological simulations are usually divided into two main categories. Large scale periodic boxes requiring massively parallel computers are performed on a very long elapsed time (usually several months). The second category stands for much faster small scale “zoom simulations”.

¹<http://www.itp.uzh.ch/~teyssier/Site/RAMSES.html>

4.5. Code Coupling Applications

Different kinds of code coupling applications are considered. The simplest form is within a parallel code. For example, OpenAtom, a molecular dynamics simulation application exhibits several coupling of different pieces of codes, whose arrangement and configuration depend on many parameters, and some depend on input simulation parameters. The challenge is to let application designers express the functionality of the application while the actual execution code can be automatically configured.

Another class of code coupling applications is the coupling at temporal level. Our collaboration with EDF R&D provides us with several use cases including scenarios with *Code_Aster* (thermo-mechanics simulations), *Syrthes* (transient thermal simulations in complex solid geometries), etc. These couplings face other issues such as combining several large scale resources, managing data transfers, etc.

CEPAGE Project-Team

4. Application Domains

4.1. Resource Allocation and Scheduling

4.1.1. Project-team positioning

CEPAGE has undertaken tasks related to the *high level modeling* of heterogeneous networks, both at logical level (overlay networks design) and performance level (latency, bandwidth prediction, connectivity artifacts) in order to optimize tasks such as *resource allocation* and *scheduling* of computations and communications. Objectives include replica placement, broadcasting (streaming) of large messages, independent tasks scheduling and optimization of OLAP databases. Such problems have received a lot of attention in research centers in the USA (Armherst, Colorado, ...), in Spain (Madrid), Poland (Wroclaw), Germany (Dortmund), and others. Papers on algorithmic aspects of platform modeling, scheduling and resource allocation appear at parallel processing conferences and journals in Parallel and Distributed Computing (IPDPS, EuroPar, HIPC, SPAA, IEEE TPDS, JPDC) and members of CEPAGE are strongly involved in many of these events (IPDPS, EuroPar, TPDS) as well as helping to animate well-established specialized workshops, such as HCW and HeteroPar.

Within Inria, studies on overlay networks are performed in the ASAP and GANG projects, and studies related to scheduling and resource allocation are done within the ROMA and the MOAIS projects (and to some extent within ALGORILLE).

4.1.2. Scientific achievements

The approach followed in the CEPAGE project, and our main originality, is to consider the whole chain, from gathering actual data on the networks to platform modeling and complexity analysis. Indeed, many complexity analysis studies are performed on models whose parameters cannot actually be evaluated (this applies, for instance, to all algorithms that assume that the topology of a platform running over the Internet is known in advance) and many platform models are intractable from an algorithmic perspective (this applies, for instance, to all models that represent latencies or bandwidths between all pairs of nodes as a general matrix). Our general goal is to provide models whose parameters can be evaluated at runtime using actual direct measurements, to propose algorithms whose worst-case (or average-case) behavior can be proved for this model, and finally to evaluate the whole chain (model + algorithm + implementation).

From an applicative perspective, in the framework of the PhD Thesis of Hejer Rejeb, we have considered several storage and resource allocation problems in collaboration with Cyril Banino-Rokkones at Yahoo! Trondheim (dealing with actual datasets enabled us to improve known approximation results in this specific context). We have in particular studied the modeling of TCP mechanism for handling contentions and its influence on the performance of several scheduling algorithms and advocated the use of QoS mechanisms for prescribed bandwidth sharing (IPDPS 2010 [67], ICPADS 2008 [52], AlgoTel 2009 [64], ICPADS 2009 [63], PDP 2010 [65]). In the PhD thesis of Hubert Larchevêque, we have considered the problem of aggregating resources (or placing replicas) in a distributed network (Sirocco 2008 [54], Opodis 2008 [55], ICPP 2011 [60], AlgoTel 2011 [56]) so that each group satisfies some properties (in terms of aggregated memory, CPU and maximal distance in terms of latency within a group). We proved several multi-criteria approximation results for this problem, and we compared several embedding tools (Vivaldi, Sequoia) in the context of resource aggregation. For these applications, we have also provided when possible distributed algorithms based on sophisticated overlay networks, in particular in order to deal with heterogeneity (IPDPS 2008 [61]). In the PhD Thesis of Przemyslaw Uznanski, we focus on the design of efficient streaming and broadcasting strategies, in particular in presence of connectivity artifacts like firewalls (IPDPS 2010 [62], ICPADS 2011 [59]). We have also worked on establishing under the bounded multiport model several new complexity results for classical distributed computing models such as divisible load theory (HCW 2008 [57], IPDPS 2008 [107], IPDPS 2012 [58]) that have been later extended to Continuous Integration (HCW 2012 [53]).

In the context of database query optimization, materializing some queries results for optimization is a standard solution when execution time performance is crucial. In the datacube context, the problem has been studied for a long time under the storage space limit constraint. Here also, we were able to reformulate this problem by considering instead the execution time as the hard constraint while the objective is to reduce the storage space. Even if the problem turns to be NP-hard, this reformulation allowed us to provide effective approximate solutions with both space and performance bounded guarantees (EDBT 2009 [96]). Moreover, reducing the storage space tends to reduce the maintenance time since the latter is linearly proportional to the former. Finally, we characterized the minimal number of updates to be performed before performance becomes no more guaranteed and a new solution must be recomputed (ADBIS 2008 [97]). One of the key concepts we used for solving this problem was that of a *border*. It turns out that this notion is equivalent to e.g., maximal frequent itemsets or minimal functional dependencies extensively studied by data mining community. In contrast to all previous proposals, we proposed the only parallel algorithm computing these borders with a speed-up guarantee regarding the number of processing units (CIKM 2011 [95]). Besides the analytical study, its implementation in maximal frequent itemset mining outperforms state of the art implementations (see Section 5.1).

To achieve these results, our efforts have also focused on analyzing and building realistic datasets (AlgoTel 2012 [86]) and proposing data analysis results for specific distributions (ISAAC 2011 [48]). On the modeling side, in general, for bandwidth and contention modeling, we have proved that the bounded multiport model (where each node is associated to an incoming bandwidth, an outgoing bandwidth and a maximal number of simultaneous TCP connexions) is both implementable, realistic and tractable (EuroPar 2011 [66]). In particular, we have proved in strongly different contexts (allocation of virtual machines to physical machines, overlay design for broadcasting, server allocation for volunteer computing) that the use of resource augmentation enables to obtain quasi-optimal results. All our modeling efforts and algorithms have been included into the SimGRID Software (<http://simgrid.gforge.inria.fr>), which enables us both to compare several algorithms under the same exact conditions and to compare the results obtained with several communication models (see Section 5.1).

Perspectives: We believe that our approach based on sound models, approximation algorithms for these models, followed by experimental validation is a strong one and we intend to continue in this direction in the following years. Our goal of designing realistic solutions pushes towards considering average case analysis of our algorithms, as well as robust optimization techniques. Furthermore, the recent strong interest in Cloud systems from the community entices us to use our expertise in resource allocation for the optimization of Cloud systems, both from the provider and from the user points of view. We already have some interesting contacts with local companies to share start collaborating on these topics. In this context, reliability issues are very important, and we believe that robust optimization is a very relevant approach for these problems.

4.2. Compact Routing

4.2.1. Project-team positioning

In this axis, CEPAGE mainly works on the design on distributed and light data structures. One of the techniques consists in summarizing the topology and metric of the networks allowing to route or to approximate the original distances within the network. Such structures, often called *spanners*, does not require the storage of all the original network links. Then we get economic distributed data structures that can be updated without a high communication cost. Our main collaborations are done with the best specialists world-wide, in particular: Israel (Weizmann), USA (MIT, Microsoft, Chicago), Belgium (Alcatel Lucent-Bell), France (Paris, Nice).

Algorithms and Routing are also intensively studied in research labs in the USA (CAIDA). Our contributions appear regularly at all of the major conferences in Distributed Computing (PODC, DISC, SPAA), as well as at top venues with a more general algorithmic audience (STOC, SODA, ICALP, ESA). Members of CEPAGE actively participate in these events (ICALP 2010 and DISC 2009 were organized by members of CEPAGE).

Within Inria, studies of mobile agents are also performed in the GANG project and to some extent also within MASCOTTE within the european project EULER.

4.2.2. Scientific achievements

There are several techniques to manage sub-linear size routing tables (in the number of nodes of the platform) while guaranteeing almost shortest paths. Some techniques provide routes of length at most $1 + \epsilon$ times the length of the shortest one while maintaining a poly-logarithmic number of entries per routing table. However, these techniques are not universal in the sense that they apply only on some class of underlying topologies. Universal schemes exist. Typically they achieve $O(\sqrt{n})$ -entry local routing tables for a stretch factor of 3 in the worst case. Some experiments have shown that such methods, although universal, work very well in practice, in average, on realistic scale-free or existing topologies.

The space lower bound of $O(\sqrt{n})$ -entry for routing with *multiplicative* stretch 3 is due to the existence of dense graphs with large girth. Dense graphs can be sparsified to subgraphs (spanners), with various stretch guarantees. There are spanners with *additive* stretch guarantees (some even have constant additive stretch) but only very few additive routing schemes are known.

In (SPAA 2012 [90]), we give reasons why routing in unweighted graphs with *additive* stretch is difficult in the form of space lower bounds for general graphs and for planar graphs. On the positive side, we give an almost tight upper bound: we present the first non-trivial compact routing scheme with $o(\lg^2 n)$ -bit addresses, *additive* stretch $O(n^{1/3})$, and table size $O(n^{1/3})$ bits for planar graphs.

We have recently considered the *forbidden-set* extension of distance oracles and routing schemes. Given an arbitrary set of edge/node failure F , a source s and a target t such that $s, t \notin F$, the goal is to route (or evaluate the distance) between s and t in the graph $G \setminus F$, so avoiding F . The classical problem is for $F = \emptyset$. This extension is considered as a first step toward fully dynamic data-structures, a challenging goal. For graphs of low doubling dimension we have shown in (PODC 2012 [47]) that it is possible to route from s to t in $G \setminus F$ with stretch $1 + \epsilon$, for all s, t, F , given poly-logarithmic size labels of all the nodes invoked in the query (s, t, F) . This has been generalized to all planar graphs achieving similar stretch and label size performances. As a byproduct we have designed a fully dynamic algorithm for maintaining $1 + \epsilon$ approximate distances in planar graphs supporting edge/node addition/removal within update and query time \sqrt{n} in the worst-case (STOC 2012 [46]).

Θ_k -graphs are geometric graphs that appear in the context of graph navigation. The shortest-path metric of these graphs is known to approximate the Euclidean complete graph up to a factor depending on the cone number k and the dimension of the space. We have introduced in (WG 2010 [68]) a specific subgraph of the Θ_6 -graph defined in the 2D Euclidean space, namely the half- Θ_6 -graph, composed of the even-cone edges of the Θ_6 -graph. Our main contribution is to show that these graphs are exactly the TD-Delaunay graphs, and are strongly connected to the geodesic embeddings of orthogonal surfaces of coplanar points in the 3D Euclidean space. We also studied the asymptotic behavior of these spanners (*Adv. in Appl. Proba.* [105]) and in collaboration with Ljubomir Perković, we worked on the question of bounded degree planar spanner. We proposed an algorithm that computes a plane 6-spanner of degree at most 6 in (ICALP 2010 [69]). The previous best bound on the maximum degree for constant stretch plane spanners was only 14.

In order to cope with network dynamism and failures, and motivated by multipath routing, we introduce a multi-connected variant of spanners. For that purpose we introduce in (OPODIS 2011 [91]) the p -multipath cost between two nodes u and v as the minimum weight of a collection of p internally vertex-disjoint paths between u and v . Given a weighted graph G , a subgraph H is a p -multipath s -spanner if for all u, v , the p -multipath cost between u and v in H is at most s times the p -multipath cost in G . The s factor is called the stretch. Building upon recent results on fault-tolerant spanners, we show how to build p -multipath spanners of constant stretch and of $O(n^{1+1/k})$ edges, for fixed parameters p and k , n being the number of nodes of the graph. Such spanners can be constructed by a distributed algorithm running in $O(k)$ rounds. Additionally, we give an improved construction for the case $p = k = 2$. Our spanner H has $O(n^{3/2})$ edges and the p -multipath cost in H between any two node is at most twice the corresponding one in G plus $O(W)$, W being the maximum edge weight.

We also worked on compact coding in data warehouses: in order to get quick answer in large data, we have to estimate, select and materialize (store) partial data structures. We got several solutions with a prescribed

guarantee in different models for the following problems: view size estimation with small samples, view selection, parallel computation of frequent itemsets. In (*Theor. Comp. Sci.* [94]) a new algorithm that allow the administrator or user of a DBMS to choose which part of the data cube to optimize (known as the *the views selection problem*), that takes as input a fact table and computes a set of views to store in order to speed up queries.

Perspectives: The compact coding activity in data-warehouse is promising since the amount of data collected keeps on increasing and being able to answer in real-time complex requests (data mining) is still challenging.

Some robust data structures already exist which, given a small number of k changes of topology or k faults, tolerate these faults, i.e., alternative routes with bounded stretch can be provided without any updates. This is a first step toward dynamic networks but the updates of these data structures are currently still quite complicated with a high communication cost.

4.3. Mobile Agents

4.3.1. Project-team positioning

CEPAGE has undertaken tasks related to the design of algorithms which control the behavior of so called *mobile agents*, moving around a network or a geometric environment, with the goal of achieving a specified objective. Objectives of central importance to the study include: exploration of unknown environments, terrain patrolling, network maintenance, and coordination of activities with other agents. Such problems have in recent years been the object of interest of numerous research teams working on Distributed Computing worldwide, in particular, at research centers in Canada (Quebec), Israel (Tel Aviv, Haifa), France (Paris, Marseille), the UK (London, Liverpool), and Switzerland (Zurich). Algorithms for mobile agents in social networking applications are also intensively studied in research labs in the USA (Stanford, Facebook). Papers on mobile agents appear regularly at all of the major conferences in Distributed Computing (PODC, DISC, SPAA), as well as at top venues with a more general algorithmic audience (SODA, ICALP, ESA). Members of CEPAGE actively participate in these events, and are also a recognizable part of the European community focused around mobile agents, helping to animate well-established specialized conferences, such as SIROCCO and OPODIS.

Within Inria, studies of mobile agents are also performed in the GANG project, and to some extent also within MASCOTTE. CEPAGE has active research links with both of these teams.

4.3.2. Scientific achievements

The work of CEPAGE has focused on contributing new decentralized algorithms for controlling mobile entities known as *agents*, deployed in unknown environments. We mainly considered the network setting, in which agents moving around the nodes of the network graph may be used to analyze the structure of the network and to perform maintenance tasks, such as detecting dynamic faults, improving/monitoring dissemination of information, etc. Our theoretical studies focused on designing new strategies for controlling the behavior of agents and answering crucial questions concerning the feasibility of solving fundamental problems, subject to different model assumptions and constraints on the knowledge and computational power of agents.

One major line of our research focused on the so called *anonymous graph model* in which an agent is unable to determine the identifier of the node of its current location, but can only see a local ordering of the links around it. Such a study is motivated e.g. by scenarios in which the identifiers of nodes may be too large for the agent to process using its bounded resources, or may change in time. In this model, we studied two of the most fundamental problems: that of traversing all of the nodes of the network (exploration) and of meeting another agent in the network (rendezvous), so as to coordinate with it. Our contributions include a precise characterization of the space requirements for agents solving both of these problems deterministically: exploration in (*Trans. Alg.* 2008 [73]) and rendezvous in (*Dist. Comp.* 2012 [81]), in a paper presented at the Best Paper Session of PODC 2010. We have also studied fast solutions for specific scenarios of the rendezvous problem (DISC 2010 [49], DISC 2011 [74], SPAA 2012 [82]) and the problem of approximate map construction within an anonymous graph (OPODIS 2010 [71]). A separate problem, intensively studied in recent years by several research teams, concerns the exploration of a network with pre-configured ports

so as to assist the agent. In our work on the topic, our team has proposed several new techniques for graph decomposition, leading in particular to the shortest currently known strategies of periodic exploration for both the case of memoryless (*Algorithmica* 2012 [101]) and small-memory agents (SIROCCO 2009 [77]).

A closely related line of research was devoted to the design of network exploration strategies which guarantee a fast and fair traversal of all the nodes, making use of agents with extremely restricted capabilities. Such strategies were inspired by the random walk, but had the additional advantage of deterministic and desirable behavior in worst-case scenarios. We presented a series of results in the area at notable conferences, involving both the design of new exploration strategies (ICALP 2009 [75]) and completely new insights into previously known approaches such as the so called “rotor-router model” (DISC 2009 [50], OPODIS 2009 [51]). All of the proposed algorithms were shown to be viable alternatives to the random walk, competing in terms of such parameters as cover time, steady-state exploration frequency, and stabilization in the event of faults.

Our efforts have also focused on the theory of coordinating activities of large groups of agents. We have conducted pioneering work in the so called look-compute-move model in networks, in which extremely restricted (asynchronous and oblivious) agents, relying on snapshot views of the system, are nevertheless able to perform useful computational tasks. Our solutions to the problems of collective exploration in trees (*Theor. Comp. Sci.* 2010 [88]) and gathering agents on a ring (*Theor. Comp. Sci.* 2008 [99] and 2010 [98]) have sparked a long line of follow-up research, accumulating more than 120 citations in total (according to Google Scholar). In a slightly different scenario, we have considered computations with teams of agents whose task is to collaboratively detect and mark potentially dangerous (faulty) links of the network, called “black holes”, which are capable of destroying agents which enter them. We have provided important contributions to the theory of black hole search in both undirected (SIROCCO 2008 [76], DISC 2008 [89]) and directed (*Theor. Comp. Sci.* [102]) graphs.

It is expected that the mobile agent theme of CEPAGE will give rise to 2 PhD theses. In 2013, Ahmed Wade will defend his thesis on mobile agent protocols for dynamic networks, whereas in 2014 Dominik Pajak will defend his thesis on multi-agent protocols for efficient graph exploration. Our scientific interests also include mobile agent protocols for geometric applications, more remote from the central themes of CEPAGE, but having extensive applications in robotics (providing protocols, e.g., for efficient patrolling and guarding of terrains, traversing terrains using groups of robots, etc.). We have already published several papers in this area (SIROCCO 2010 [79], SWAT 2010 [80], ESA 2011 [78]), building up the theoretical fundamentals of a new field, and already attracting the attention of a wider community of researchers working in robotics and AI.

Perspectives: Our goal is to explore applications of mobile agent techniques in domains of growing importance, namely, social networks and robotics. We are currently discussing applications of our techniques in problems of brand recognition on the web with a local industrial partner (Systonic KeepAlert), and other companies (through our research collaborators in Liverpool). We intend to undertake collaboration with European/American research labs and industrial partners.

CIDRE Project-Team

4. Application Domains

4.1. Application Domains

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, where security (and safety) is a major concern, may benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by general public (basically, the internet and services such as web services, social networks, etc.) can also benefit from results obtained by CIDRE, especially with regards to privacy. Systems are getting more and more complex, decentralized, distributed, or spontaneous. The emergence of cloud computing brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

DANTE Team

4. Application Domains

4.1. Life Science & Health

In parallel to the advances in modern medicine, health sciences and public health policy, epidemic models aided by computer simulations and information technologies offer an increasingly important tool for the understanding of transmission dynamics and of epidemic patterns. The increased computational power and use of Information and Communication Technologies makes feasible sophisticated modeling approaches augmented by detailed in vivo data sets, and allow to study a variety of possible scenarios and control strategies, helping and supporting the decision process at the scientific, medical and public health level. The research conducted in the DANTE project finds direct applications in the domain of LSH since modeling approaches crucially depend on our ability to describe the interactions of individuals in the population. In the MOSAR project we are collaborating with the team of Pr. Didier Guillemot (Inserm/Institut. Pasteur/Université de Versailles). Within the TUBEXPO and ARIBO projects, we are collaborating with Pr. Jean-Christophe Lucet (Professeur des université Paris VII ? Praticien hospitalier APHP).

4.2. Network Science / Complex networks

In the last ten years, the study of complex networks has received an important boost with large interdisciplinary efforts aimed at their analysis and characterization. Two main points explain this large activity: on the one hand, many systems coming from very different disciplines (from biology to computer science) have a convenient representation in terms of graphs; on the other hand, the ever-increasing availability of large data sets and computer power have allowed their storage and manipulation. Many maps have emerged, describing many networks of practical interest in social science, critical infrastructures, networking, and biology. The DANTE project targets the study of dynamically evolving networks, from the point both of their structure and of the dynamics of processes taking place on them.

DIONYSOS Project-Team (section vide)

DISTRIBCOM Project-Team

4. Application Domains

4.1. Telecommunication network management

The management of telecommunication networks is traditionally a human performed activity that covers the five FCAPS functions: Fault management, network Configuration, Accounting, Performances and Security. This simple classification has exploded in the last decade, under the pressure of several phenomena. The first one concerns the growth in size and complexity of networks, with the emergence of new (possibly virtual) operators, the multiplication of vendors, new core and (wireless) access technologies, the variety of terminal devices, the convergence of phone/computer/radio/TV networks, the multiplication of services over the top, the necessity to provide QoS for a wide variety of traffic demands, etc. As a consequence, the management task is reaching the limits of human operators and demands automation. It is estimated that telecommunication companies spend over 50% of their manpower on management tasks. They naturally want to reduce it and dedicate their effort to the design and offer of innovative services, where the added value is more important (as witnessed by the success of some over-the-top companies). The result of these trends is that network management now covers a much wider variety of problems, for which automatic solutions are requested. This takes the name of self-management, or autonomic management: one wishes to manage networks by high-level objectives, and networks should be able to adapt themselves automatically to fulfill these objectives. DistribCom is contributing to this field with its background on the modeling of distributed/concurrent systems, and its expertise in distributed algorithms. Networks are perfect examples of large distributed and concurrent systems, with specific features like the dynamicity (their structure evolves) and a hierarchical structure (multiple layers, multiple description granularities). We have proposed model-based distributed algorithms to solve problems like failure diagnosis, negotiation of QoS (quality of service) parameters, parameter optimization, graceful shutdown of OSPF routers for maintenance operations... The present activities in this domain are related to the joint diagnosis for access network + core network + services, within the European IP UniverSelf. The challenges cover self-modelling methods (how to obtain the network model that is used by the management algorithms), active diagnosis methods that both adapt the scope of their network model and perform tests to explain a fault situation, and self-healing methods.

4.2. Web services and active structured documents

Keywords: Active documents, Web services, choreographies, orchestrations, QoS.

Web services architectures are usually composed of distant services, assembled in a composite framework. This raises several practical issues: one of them is how to choose services, assemble them, and coordinate their executions in a composite framework. Another issue is to guarantee good properties of a composite framework (safety but also QoS properties). All this has to be done in a context where a distant service provided by a subcontractor is only perceived as an interface, specifying legal inputs and outputs, and possibly a quality contract. The standard in industry for Web-services is now BPEL [43] but most of the problems listed above are untractable for this language. Composition of services can also be performed using choreography languages such as ORC [55]. The implementation of orchestration and choreography description languages raises a number of difficulties related to efficiency, clean semantics, and reproducibility of executions, issues of composite QoS associated with orchestrations. We develop studies in these areas, with the aim of proposing service composition frameworks equipped with tools to specify, but also to monitor and analyze the specified architectures. Another issue is the convergence between data and workflows. Web Services architectures are frequently considered exclusively as workflows, or as information systems. Many approaches to Web Service orchestration and choreography abstract data away. Symmetrically, modern approaches to Web data management typically based on XML and Xqueries rely on too simplistic forms of control. We develop a line of research on Active documents. Active documents are structured data embedding references to services,

which allow for the definitions of complex workflows involving data aspects. The original model was proposed by S. Abiteboul (see for instance [42]), but the concept of active document goes beyond AXML, and offers a document oriented alternative to Web services orchestrations and choreographies. This approach is in particular well adapted to the modeling of E-business processes, or information processing in organizations, etc. Our aim is to extend and promote the concept of active document. This means developing verification and composition tools for document-based architectures, considered not only as theoretical models but also as effectively running systems. To this extend, we develop an active document platform.

FOCUS Project-Team

4. Application Domains

4.1. Ubiquitous Systems

The main application domain for Focus are ubiquitous systems, broadly systems whose distinctive features are: mobility, high dynamicity, heterogeneity, variable availability (the availability of services offered by the constituent parts of a system may fluctuate, and similarly the guarantees offered by single components may not be the same all the time), open-endedness, complexity (the systems are made by a large number of components, with sophisticated architectural structures). In Focus we are particularly interested in the following aspects.

- *Linguistic primitives* for programming dialogues among components.
- *Contracts* expressing the functionalities offered by components.
- *Adaptability and evolvability* of the behaviour of components.
- *Verification* of properties of component systems.
- Bounds on component *resource consumption* (e.g., time and space consumed).

4.2. Service Oriented Computing and Cloud Computing

Today the component-based methodology often refers to Service Oriented Computing. This is a specialized form of component-based approach. According to W3C, a service-oriented architecture is “a set of components which can be invoked, and whose interface descriptions can be published and discovered”. In the early days of Service Oriented Computing, the term services was strictly related to that of Web Services. Nowadays, it has a much broader meaning as exemplified by the XaaS (everything as a service) paradigm: based on modern virtualization technologies, Cloud computing offers the possibility to build sophisticated service systems on virtualized infrastructures accessible from everywhere and from any kind of computing device. Such infrastructures are usually examples of sophisticated service oriented architectures that, differently from traditional service systems, should also be capable to elastically adapt on demand to the user requests.

4.3. Software Product Lines

A Software Product Line is a set of software systems that together address a particular market segment or fulfill a particular mission. Today, Software Product Lines are successfully applied in a range of industries, including telephony, medical imaging, financial services, car electronics, and utility control [54]. Customization and integration are keywords in Software Product Lines: a specific system in the family is constructed by selecting its properties (often technically called “features”), and, following such selection, by customizing and integrating the needed components and deploying them on the required platform.

FUN Team (section vide)

GANG Project-Team

3. Application Domains

3.1. Application Domains

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale ad hoc networks and mapping the web.

- The application of measuring and modeling Internet metrics such as latencies and bandwidth is to provide tools for optimizing Internet applications. This concerns especially large scale applications such as web site mirroring and peer-to-peer applications.
- Peer-to-peer protocols are based on a all equal paradigm that allows to design highly reliable and scalable applications. Besides the file sharing application, peer-to-peer solutions could take over in web content dissemination resistant to high demand bursts or in mobility management. Envisioned peer-to-peer applications include video on demand, streaming, exchange of classified ads,...
- Wifi networks have entered our every day life. However, enabling them at large scale is still a challenge. Algorithmic breakthrough in large ad hoc networks would allow to use them in fast and economic deployment of new radio communication systems.
- The main application of the web graph structure consists in ranking pages. Enabling site level indexing and ranking is a possible application o f such studies.

GRAND-LARGE Project-Team (section vide)

HIEPACS Project-Team

4. Application Domains

4.1. Introduction

Currently, we have one major application which is material physics, and for which we contribute to all steps that go from modelling aspects to the design and the implementation of very efficient algorithms and codes for very large multi-scale simulations. Moreover, we apply our algorithmic research about linear algebra (see Section 3) in the context of several collaborations with industrial and academic partners. Our high performance libraries are or will be integrated in several complex codes and will be used and validated for very large simulations.

4.2. Material physics

Participants: Béranger Bramas, Olivier Coulaud, Aurélien Esnard, Pierre Fortin, Luc Giraud, Jean Roman.

Due to the increase of available computer power, new applications in nano science and physics appear such as study of properties of new materials (photovoltaic materials, bio- and environmental sensors, ...), failure in materials, nano-indentation. Chemists, physicists now commonly perform simulations in these fields. These computations simulate systems up to billion of atoms in materials, for large time scales up to several nanoseconds. The larger the simulation, the smaller the computational cost of the potential driving the phenomena, resulting in low precision results. So, if we need to increase the precision, there is two ways to decrease the computational cost. In the first approach, we improve classical methods and algorithms and in the second way, we will consider a multiscale approach.

Many applications in material physics need to couple several models like quantum mechanic and molecular mechanic models, or molecular and mesoscopic or continuum models. These couplings allow scientists to treat larger solids or molecules in their environment. Many of macroscopic phenomena in science depend on phenomena at smaller scales. Full simulations at the finest level are not computationally feasible in the whole material. Most of the time, the finest level is only necessary where the phenomenon of interest occurs; for example in a crack propagation simulation, far from the tip, we have a macroscopic behavior of the material and then we can use a coarser model. The idea is to limit the more expensive level simulation to a subset of the domain and to combine it with a macroscopic level. This implies that atomistic simulations must be speeded up by several orders of magnitude.

We will focus on two applications; the first one concerns the computation of optical spectra of molecules or solids in their environment. In the second application, we will develop faster algorithms to obtain a better understanding of the metal plasticity, phenomenon governing by dislocation behavior. Moreover, we will focus on the improvement of the algorithms and the methods to build faster and more accurate simulations on modern massively parallel architectures.

4.2.1. Hybrid materials

There is current interest in hybrid pigments for cosmetics, phototherapy and paints. Hybrid materials, combining the properties of an inorganic host and the tailorable properties of organic guests, particularly dyes, are also of wide interest for environmental detection (oxygen sensors) and remediation (trapping and elimination of dyes in effluents, photosensitised production of reactive oxygen species for reduction of air and water borne contaminants). A thorough understanding of the factors determining the photo and chemical stability of hybrid pigments is thus mandated by health, environmental concerns and economic viability.

Many applications of hybrid materials in the field of optics exploit combinations of properties such as transparency, adhesion, barrier effect, corrosion, protection, easy tuning of the colour and refractive index, adjustable mechanical properties and decorative properties. It is remarkable that ancient pigments, such as Maya Blue and lacquers, fulfill a number of these properties. This is a key to the attractiveness of such materials. These materials are not simply physical mixtures, but should be thought of as either miscible organic and inorganic components, or as a heterogeneous system where at least one of the component exhibits a hierarchical order at the nanometer scale. The properties of such materials no longer derive from the sum of the individual contributions of both phases, since the organic/inorganic interface plays a major role. Either organic and inorganic components are embedded and only weak bonds (hydrogen, van der Waals, ionic bonds) give the structure its cohesion (class I) or covalent and ionic-covalent bonds govern the stability of the whole (class II).

These simulations are complex and costly and may involve several length scales, quantum effects, components of different kinds (mineral-organic, hydro-philic and -phobic parts). Computer simulation already contributes widely to the design of these materials, but current simulation packages do not provide several crucial functions, which would greatly enhance the scope and power of computer simulation in this field.

The computation of optical spectra of molecules and solids is the greatest use of the Time Dependent Density Functional Theory (TDDFT). We compute the ground state of the given system as the solution of the Kohn-Sham equations (DFT). Then, we compute the excited states of the quantum system under an external perturbation - electrical field of the environment - or thanks to the linear theory, we compute only the response function of the system. In fact, physicists are not only interested by the spectra for one conformation of the molecule, but by an average on its available configurations. To do that, they sample the trajectory of the system and then compute several hundred of optical spectra in one simulation. But, due to the size of interesting systems (several thousands of atoms) and even if we consider linear methods to solve the Kohn-Sham equations arising from the Density Functional Theory, we cannot compute all the system at this scale. In fact, such simulations are performed by coupling Quantum mechanics (QM) and Molecular mechanic (MM). A lot of works are done on the way to couple these two scales, but a lot of work remains in order to build efficient methods and efficient parallel couplings.

The most consuming time in such coupling is to compute optical spectra is the TDDFT. Unfortunately, examining optical excitations based on contemporary quantum mechanical methods can be especially challenging because accurate methods for structural energies, such as DFT, are often not well suited for excited state properties. This requires new methods designed for predicting excited states and new algorithms for implementing them. Several tracks will be investigated in the project:

- Typically physicists or chemists consider spectral functions to build a basis (orbital functions) and all the computations are performed in a spectral way. Due to our background, we want to develop new methods to solve the system in the real space by finite differences or by wavelets methods. The main expectation is to construct error estimates based on for instance the grid-size h parameter.
- For a given frequency in the optical spectra, we have to solve a symmetric non Hermitian system. With our knowledge on linear solvers, we think that we can improve the methods commonly used (Lanczos like) to solve the system (see Section 3.3).
- Improving the parallel coupling is crucial for large systems because the computational cost of the atomic and quantum models are really different. In parallel we have the following order of magnitude: one second or less per time step for the molecular dynamics, several minutes or more for the DFT and the TDDFT. The challenge to find the best distribution in order to have the same CPU time per time step is really important to reach high performance. Another aspect in the coupling is the coupling with the visualization to obtain online visualization or steerable simulations. Such steerable simulations help the physicists to construct the system during the simulation process by moving one or a set of molecules. This kind of interaction is very challenging in terms of algorithmic and this is a good field for our software platform EPSN.

4.2.2. Material failures

Another domain of interest is the material aging for the nuclear industry. The materials are exposed to complex conditions due to the combination of thermo-mechanical loading, the effects of irradiation and the harsh operating environment. This operating regime makes experimentation extremely difficult and we must rely on multi-physics and multi-scale modelling for our understanding of how these materials behave in service. This fundamental understanding helps not only to ensure the longevity of existing nuclear reactors, but also to guide the development of new materials for 4th generation reactor programs and dedicated fusion reactors. For the study of crystalline materials, an important tool is dislocation dynamics (DD) modelling. This multiscale simulation method predicts the plastic response of a material from the underlying physics of dislocation motion. DD serves as a crucial link between the scale of molecular dynamics and macroscopic methods based on finite elements; it can be used to accurately describe the interactions of a small handful of dislocations, or equally well to investigate the global behavior of a massive collection of interacting defects.

To explore, i.e., to simulate these new areas, we need to develop and/or to improve significantly models, schemes and solvers used in the classical codes. In the project, we want to accelerate algorithms arising in those fields. We will focus on the following topics (in particular in the starting OPTIDIS ANR-COSINUS project in collaboration with CEA Saclay, CEA Ile-de-france and SIMaP Laboratory in Grenoble) in connection with research described at Sections 3.4 and 3.5 .

- The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. In dislocation codes, the computation of interaction forces between dislocations is still the most CPU time consuming and has to be improved to obtain faster and more accurate simulations.
- In such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically construct a good load balancing are crucial to achieve high performance.
- From a physical and a simulation point of view, it will be interesting to couple a molecular dynamics model (atomistic model) with a dislocation one (mesoscale model). In such three-dimensional coupling, the main difficulties are firstly to find and characterize a dislocation in the atomistic region, secondly to understand how we can transmit with consistency the information between the two micro and meso scales.

4.3. Application framework customers of high performance linear algebra solvers

Participants: Emmanuel Agullo, Luc Giraud, Abdou Guermouche, Stojce Nakov, Jean Roman, Xavier Vasseur.

We are currently collaborating with various research groups involved in geophysics, electromagnetics and structural mechanics. For all these application areas, the current bottleneck is the solution of huge sparse linear systems often involving multiple right-hand sides either available simultaneously or given in sequence. The robustness, efficiency and scalability of the numerical tools designed in Section 3.3 will be preliminary investigated in the parallel simulation codes of these partners.

For the solution of large systems arising from PDE discretization, the geometric full multigrid technique based on a few levels in the grid hierarchy and an efficient parallel sparse direct solver on the coarsest level can be considered. Originally developed for 3D Maxwell solution in collaboration with CEA-CESTA, the approach can be extended to other application fields.

Many simulation codes need the solution with simultaneous right-hand sides but also with right-hand sides given in sequence. The first situation arises in RCS calculations, but is generic in many parametric studies, while the second one comes from the nature of the solver such as implicit time stepping schemes or inverse iterations. Many of the numerical approaches and possible outcoming software are well suited to tackle these challenging problems.

On more academic sides, some ongoing collaborations with other Inria EPIs will be continued and others will be started. In collaboration with the NACHOS Inria project team, we will continue to investigate the use of efficient linear solvers for the solution of the Maxwell equations in the time and frequency domains where discontinuous Galerkin discretizations are considered. Additional funding will be sought out in order to foster this research activity in connection with actions described in Section 3.3 .

The efficient solution of linear systems strongly relies on the activities described in Section 3.2 (e.g. complex load balancing problem) and in Section 3.3 (for the various parallel linear algebra kernels).

4.4. Scalable numerical schemes for scientific applications

Participants: Olivier Coulaud, Yohann Dudouit, Luc Giraud, Guillaume Latu, Alexis Praga, Jean Roman, Pablo Salas Medina, Xavier Vasseur.

We are also collaborating with application research group to design or improve numerical schemes in the view of large scale parallel simulations.

Seismic wave propagation in heterogeneous media requires to properly capture the local heterogeneity and consequently requires locally refined meshes. In close collaboration with TOTAL we study new parallelizable schemes for the solution of the elastodynamic system with local spatial refinements based on discontinuous Galerkin techniques. The objective is to design novel parallel scalable implementations for large 3D simulations. A second work is currently carried on with TOTAL for Seismic modeling and Reverse Time Migration (RTM) based on the full wave equation discretization. These tools are of major importance since they give an accurate representation of complex wave propagation areas. Unfortunately, they are highly compute intensive. To address this challenge we have designed a fast parallel simulator that solves the acoustic wave equation on a GPU cluster.

Thermoacoustic instabilities are an important concern in the design of gas turbine combustion chambers. Most modern combustion chambers have annular shapes and this leads to the appearance of azimuthal acoustic modes. These modes are often powerful and can lead to structural vibrations being sometimes damaging. Therefore, they must be identified at the design stage in order to be able to eliminate them. However, due to the complexity of industrial combustion chambers with a large number of burners, numerical studies of real configurations are a challenging task. Such a challenging calculations performed in close collaboration with the Computational Fluid Dynamic project at CERFACS.

The chemistry and transport models (CTM) play a central role in global geophysical models. The solution of the CTM represents up-to 50 % on the computing resources involved in global geophysical simulations. Therefore, the availability of efficient scalable parallel numerical schemes on emerging and future supercomputers is crucial. The purpose of this research activity is to study, design and implement novel numerical schemes following the work initiated by D. Cariolle in the framework of the ANR Solstice project. Alexi Praga, PhD hired by CERFACS, is conducting this research action under the joint supervision of HiePACS and the Aviation and Environment project at CERFACS in close collaboration with CNRM/Meteo-France.

HIPERCOM Project-Team

4. Application Domains

4.1. Introduction

The HIPERCOM project-team is mainly concerned by six domains:

- wireless mobile ad hoc networks,
- services over mobile networks,
- community networks,
- vehicular networks,
- large ad hoc networks with sensor nodes,
- energy-efficient wireless sensor networks.

4.2. Wireless mobile ad hoc networks

Wireless mobile ad hoc networks, Services over mobile networks, Community Networks.

Abstract. Mobile wireless networks have numerous applications in rescue and emergency operation, military tactical networking and in wireless high speed access to the internet.

A mobile ad hoc network is a network made of a collection of mobile nodes that gather spontaneously and communicate without requiring a pre-existing infrastructure. Of course a mobile ad hoc network use a wireless communication medium. They can be applied in various contexts:

- military;
- rescue and emergency;
- high speed access to internet.

The military context is the most obvious application of mobile ad hoc networks.

Soldiers invading a country won't subscribe in advance to the local operator. On the reverse side, home units won't use their local operators firstly because they will likely be disrupted in the first hours of the conflict, and secondly because a wireless communication via an operator is not stealth enough to protect the data and the units. In Chechnya, a general has been killed by a missile tracking the uplink signal of his portable phone.

The rescue context is halfway between military and civilian applications. In the september 11 disaster, most of the phone base station of the area have knocked out in less than twenty minutes. The remaining base stations were unable to operate because they could not work in ad hoc mode. The Wireless Emergency Rescue Team recommended afterward that telecom operators should provide ad hoc mode for their infrastructure in order to operate in emergency situation in plain cooperation with police, firemen and hospital networks.

Mobile ad hoc network provide an enhanced coverage for high speed wireless access to the internet. The now very popular WLAN standard, WiFi, provides much larger capacity than mobile operator networks. Using a mobile ad hoc network around hot spots will offer high speed access to much larger community, including cars, busses, trains and pedestrians.

4.3. Services over mobile networks

Abstract. New wireless network calls for new services that fulfill the requirement in terms of mobility and capacity.

The generalization of a new generation of mobile networks calls for a new set of services and applications. For example:

- Indoor and outdoor positioning
- Service discovery and localisation
- Multicast and quality of services

Quality of service has become the central requirement that users expect from a network. High throughput, service continuity are critical issue for multimedia application over the wireless internet where the bandwidth is more scarce than in the wired world. A significant issue in the ad-hoc domain is that of the integrity of the network itself. Routing protocols allow, according to their specifications, any node to participate in the network - the assumption being that all nodes are behaving well and welcome. If that assumption fails - then the network may be subject to malicious nodes, and the integrity of the network fails. An important security service over mobile networks is to ensure that the integrity of the network is preserved even when attacks are launched against the integrity of the network.

4.4. Community Networks

Abstract. There is an increasing demand to deploy network within a community, rural or urban, with cabled or wireless access.

Community networks or citizen network are now frequent in big cities. In America most of the main cities have a community network. A community network is using the communication resource of each member (ADSL, Cable and wireless) to provide a general coverage of a city. Pedestrian in the street or in city mails can communicate via a high speed mobile mesh network. This new trend now appears in Europe with many experiments of the OLSR routing protocol in Paris, Lille, Toulouse, Berlin, Bruxelles, Seattle. The management of such networks is completely distributed and makes them very robust to faults. There is room for smart operators in this business.

4.5. Vehicular Networks

Abstract. Intelligent transport systems require efficient wireless telecommunications.

Vehicular ad hoc networks (VANET) are based on short- to medium-range transmission systems that support both vehicle-to-vehicle and vehicle-to-roadside communications. Vehicular networks will enable vehicular safety applications (safety warnings) as well as non-safety applications (real-time traffic information, routing support, mobile entertainment, and many others). We are interested in developing an efficient routing protocol that takes advantage of the fixed network infrastructure deployed along the roads. We are also studying MAC layer issues in order to provide more priority for security messages which have stringent delivery constraints.

4.6. Large ad hoc networks with sensor nodes

Abstract. Large autonomous wireless sensors in the internet of the things need very well tuned algorithms.

Self-organization is considered as a key element in tomorrow's Internet architecture. A major challenge concerning the integration of self-organized networks in the Internet is the accomplishment of light weight network protocols in large ad hoc environments.

In this domain, Hipercom's activity with wireless sensor nodes in collaboration with the Freie Universitaet in Berlin explores various solutions, including extensions of OLSR (for example DHT-OLSR) using programmable sensor nodes co-designed by the Freie Universitaet, and provides one of the largest testbeds of this kind, to date.

4.7. Energy efficient wireless sensor networks

Abstract. Energy efficiency is a key property in wireless sensor networks.

Various techniques are used to contribute to energy efficiency. In the OCARI network, an industrial wireless sensor network, we have designed and implemented an energy efficient routing protocol and a node activity scheduling algorithm allowing router nodes to sleep. We have applied a cross-layering approach allowing the optimization of MAC and network protocols taking into account the application requirements and the environment in which the network operates. This activity has been done in collaboration with our partners EDF, LIMOS and TELIT.

INDES Project-Team

4. Application Domains

4.1. Web programming

Along with games, multimedia applications, electronic commerce, and email, the web has popularized computers in everybody's life. The revolution is engaged and we may be at the dawn of a new era of computing where the web is a central element. The web constitutes an infrastructure more versatile, polymorphic, and open, in other words, more powerful, than any dedicated network previously invented. For this very reason, it is likely that most of the computer programs we will write in the future, for professional purposes as well as for our own needs, will extensively rely on the web.

In addition to allowing reactive and graphically pleasing interfaces, web applications are de facto distributed. Implementing an application with a web interface makes it instantly open to the world and accessible from much more than one computer. The web also partially solves the problem of platform compatibility because it physically separates the rendering engine from the computation engine. Therefore, the client does not have to make assumptions on the server hardware configuration, and vice versa. Lastly, HTML is highly durable. While traditional graphical toolkits evolve continuously, making existing interfaces obsolete and breaking backward compatibility, modern web browsers that render on the edge web pages are still able to correctly display the web pages of the early 1990's.

For these reasons, the web is arguably ready to escape the beaten track of n-tier applications, CGI scripting and interaction based on HTML forms. However, we think that it still lacks programming abstractions that minimize the overwhelming amount of technologies that need to be mastered when web programming is involved. Our experience on reactive and functional programming is used for bridging this gap.

4.2. Multimedia

Electronic equipments are less and less expensive and more and more widely spread out. Nowadays, in industrial countries, computers are almost as popular as TV sets. Today, almost everybody owns a mobile phone. Many are equipped with a GPS or a PDA. Modem, routers, NASes and other network appliances are also commonly used, although they are sometimes sealed under proprietary packaging such as the Livebox or the Freebox. Most of us evolve in an electronic environment which is rich but which is also populated with mostly isolated devices.

The first multimedia applications on the web have appeared with the Web 2.0. The most famous ones are Flickr, YouTube, or Deezer. All these applications rely on the same principle: they allow roaming users to access the various multimedia resources available all over the Internet via their web browser. The convergence between our new electronic environment and the multimedia facilities offered by the web will allow engineers to create new applications. However, since these applications are complex to implement this will not happen until appropriate languages and tools are available. In the Indes team, we develop compilers, systems, and libraries that address this problem.

4.3. House Automation

The web is the de facto standard of communication for heterogeneous devices. The number of devices able to access the web is permanently increasing. Nowadays, even our mobile phones can access the web. Tomorrow it could even be the turn of our wristwatches! The web hence constitutes a compelling architecture for developing applications relying on the "ambient" computing facilities. However, since current programming languages do not allow us to develop easily these applications, ambient computing is currently based on ad-hoc solutions. Programming ambient computing via the web is still to be explored. The tools developed in the Indes team allow us to build prototypes of a web-based home automation platform. For instance, we experiment with controlling heaters, air-conditioners, and electronic shutters with our mobile phones using web GUIs.

KERDATA Project-Team

4. Application Domains

4.1. Application Domains

Below are three examples which illustrate the needs of large-scale data-intensive applications with respect to storage, I/O and data analysis. They illustrate the classes of applications that can benefit from our research activities.

4.1.1. *Joint genetic and neuroimaging data analysis on Azure clouds*

Joint acquisition of neuroimaging and genetic data on large cohorts of subjects is a new approach used to assess and understand the variability that exists between individuals, and that has remained poorly understood so far. As both neuroimaging- and genetic-domain observations represent a huge amount of variables (of the order of millions), performing statistically rigorous analyses on such amounts of data is a major computational challenge that cannot be addressed with conventional computational techniques only. On the one hand, sophisticated regression techniques need to be used in order to perform significant analysis on these large datasets; on the other hand, the cost entailed by parameter optimization and statistical validation procedures (e.g. permutation tests) is very high.

The A-Brain (AzureBrain) Project started in October 2010 within the Microsoft Research-Inria Joint Research Center. It is co-led by the KerData (Rennes) and Parietal (Saclay) Inria teams. They jointly address this computational problem using cloud related techniques on Microsoft Azure cloud infrastructure. The two teams bring together their complementary expertise: KerData in the area of scalable cloud data management, and Parietal in the field of neuroimaging and genetics data analysis.

In particular, KerData brings its expertise in designing solutions for optimized data storage and management for the Map-Reduce programming model. This model has recently arisen as a very effective approach to develop high-performance applications over very large distributed systems such as grids and now clouds. The computations involved in the statistical analysis designed by the Parietal team fit particularly well with this model.

4.1.2. *Structural protein analysis on Nimbus clouds*

Proteins are major components of the life. They are involved in lots of biochemical reactions and vital mechanisms for the living organisms. The three-dimensional (3D) structure of a protein is essential for its function and for its participation to the whole metabolism of a living organism. However, due to experimental limitations, only few protein structures (roughly, 60,000) have been experimentally determined, compared to the millions of proteins sequences which are known. In the case of structural genomics, the knowledge of the 3D structure may be not sufficient to infer the function. Thus, an usual way to make a structural analysis of a protein or to infer its function is to compare its known, or potential, structure to the whole set of structures referenced in the *Protein Data Bank* (PDB).

In the framework of the MapReduce ANR project led by KerData, we focus on the SuMo application (*Surf the Molecules*) proposed by Institute for Biology and Chemistry of the Proteins from Lyon (IBCP, a partner in the MapReduce project). This application performs structural protein analysis by comparing a set of protein structures against a very large set of structures stored in a huge database. This is a typical data-intensive application that can leverage the Map-Reduce model for a scalable execution on large-scale distributed platforms. Our goal is to explore storage-level concurrency-oriented optimizations to make the SuMo application scalable for large-scale experiments of protein structures comparison on cloud infrastructures managed using the Nimbus IaaS toolkit developed at Argonne National Lab (USA).

If the results are convincing, then they can immediately be applied to the derived version of this application for drug design in an industrial context, called MED-SuMo, a software managed by the MEDIT SME (also a partner in this project). For pharmaceutical and biotech industries, such an implementation run over a cloud computing facility opens several new applications for drug design. Rather than searching for 3D similarity into biostructural data, it will become possible to classify the entire biostructural space and to periodically update all derivative predictive models with new experimental data. The applications in that complete chemo-proteomic vision concern the identification of new druggable protein targets and thereby the generation of new drug candidates.

4.1.3. I/O intensive climate simulations for the Blue Waters post-Petascale machine

A major research topic in the context of HPC simulations running on post-Petascale supercomputers is to explore how to efficiently record and visualize data during the simulation without impacting the performance of the computation generating that data. Conventional practice consists in storing data on disk, moving it off-site, reading it into a workflow, and analyzing it. It becomes increasingly harder to use because of the large data volumes generated at fast rates, in contrast to limited back-end speeds. Scalable approaches to deal with these I/O limitations are thus of utmost importance. This is one of the main challenges explicitly stated in the roadmap of the Blue Waters Project (<http://www.ncsa.illinois.edu/BlueWaters/>), which aims to build one of the most powerful supercomputers in the world when it comes online in 2012.

In this context, the KerData project-team started to explore ways to remove the limitations mentioned above through a collaborative work in the framework of the Joint Inria-UIUC Lab for Petascale Computing (JLPC, Urbana-Champaign, Illinois, USA), whose research activity focuses on the Blue Waters project. As a starting point, we are focusing on a particular tornado simulation code called CM1 (Cloud Model 1), which is intended to be run on the Blue Waters machine. Preliminary investigation demonstrated the inefficiency of the current I/O approaches, which typically consists in periodically writing a very large number of small files. This causes burst of I/O in the parallel file system, leading to poor performance and extreme variability (jitter) compared to what could be expected from the underlying hardware. The challenge here is to investigate how to make an efficient use of the underlying file system by avoiding synchronization and contention as much as possible. In collaboration with the JLPC, we started to address those challenges through an approach based on dedicated I/O cores.

LOGNET Team

3. Application Domains

3.1. Introduction

We study overlay networks and peer-to-peer systems. Our skills are applied to studying protocols to interconnect heterogeneous networks, while guaranteeing backward compatibility. We experiment with those networks and systems in various fields, such as social networks and video streaming.

We design and implement a generic social platform, which is able to “program” and “run” (in a cloud based platform hosting a NoSQL data base) generic social networks. This is the first step towards a full decentralized P2P-based social network platform.

We also study Trust and Reputation Systems for P2P networks, and for Network Web Economy.

The final objective of those research veins is to move the computer and the computability at the edge of the network.

As another topic, we also study logics and type theory for improving proof assistants based on the Curry-Howard Isomorphism.

3.2. Applications

Because of its generality, our overlay network can target many applications. We would like to list a small number of useful programmable overlay-network-related case studies that can be considered as “LogNet Grand Challenges”, to help potential readers understand the interest of our research program.

- Interconnecting overlay networks transparently;
- building a programmable social network platform relying on a cloud + P2P architecture;
- experimenting with our interconnecting algorithm in the domain of video streaming;
- studying and integrating mobile devices and mobile networks 3G/4G as a real peer in actual P2P systems;
- studying trust and reputation systems applied to P2P and web economy;
- studying new distributed models of computation (long term objective);
- studying new type theories and lambda-calculi to be the basis of new proof assistants based on Curry-Howard isomorphism.

MADYNES Project-Team

4. Application Domains

4.1. Mobile, ad-hoc and constrained networks

The results coming out from MADYNES can be applied to any dynamic infrastructure that contributes to the delivery of value added services. While this is a potentially huge application domain, we focus on the following environments at the network level:

1. multicast services,
2. ad-hoc networks,
3. mobile devices and IPv6 networks,
4. voice over IP infrastructure.

All these selected application areas exhibit different dynamicity features. In the context of multicast services, we focus on distribution, monitoring and accounting of key distribution protocols. On *ad-hoc* and dynamic networks we are investigating the provisioning, monitoring, configuration and performance management issues.

Concerning mobile devices, we are interested in their configuration, provisioning and monitoring. IPv6 work goes on in Information Models and on self-configuration of the agents.

4.2. Dynamic service infrastructures

At the service level, dynamics is also increasing very fast. We apply the results of our work on autonomous management on infrastructures which support dynamic composition and for which self-instrumentation and management automation is required.

The target service environments are:

- sensor networks,
- peer-to-peer infrastructures,
- information centric networks,
- ambient environments.

MAESTRO Project-Team

4. Application Domains

4.1. Application Domains

MAESTRO's main application area is networking and in particular, modeling, performance evaluation, optimization and control of protocols and network architectures. It includes:

- Wireless (cellular, ad hoc, sensor) networks: WLAN, WiMAX, UMTS, LTE, HSPA, delay tolerant networks (DTN), power control, medium access control, transmission rate control, redundancy in source coding, mobility models, coverage, routing, green base stations,
- Internet applications: social networks, content distribution systems, peer-to-peer systems, overlay networks, multimedia traffic, video-on-demand, multicast;
- Information-Centric Networking (ICN) architectures: Content-Centric Network (CCN, also called Content-Oriented Networks);
- Internet infrastructure: TCP, high speed congestion control, voice over IP, service differentiation, quality of service, web caches, proxy caches.

MASCOTTE Project-Team

4. Application Domains

4.1. Application Domains

In the last year the main application domain of the project remained Telecommunications. Within this domain, we consider applications that follow the needs and interests of our industrial partners, in particular Orange Labs or Alcatel-Lucent Bell-Labs, but also SMEs like 3-Roam and Avisto .

MASCOTTE is mainly interested in the design and management of heterogeneous networks. The project has kept working on the design of backbone networks (optical networks, radio networks, IP networks).

The project has also been working on routing algorithms such as dynamic and compact routing schemes in the context of the FP7 EULER led by Alcatel-Lucent Bell-Labs (Belgium). It also studied the evolution of the routing in case of any kind of topological modifications (maintenance operations, failures, capacity variations, etc.). Finally, an emphasis is done on green networks with low power consumption. This work is in collaboration with Orange Labs and the SME 3-Roam and partly supported by the ANR DIMAGREEN.

MESCAL Project-Team

4. Application Domains

4.1. On-demand Geographical Maps

Participant: Jean-Marc Vincent.

This joint work involves the UMR 8504 Géographie-Cité, LIG, UMS RIATE and the Maisons de l'Homme et de la Société.

Improvements in the Web developments have opened new perspectives in interactive cartography. Nevertheless existing architectures have some problems to perform spatial analysis methods that require complex calculus over large data sets. Such a situation involves some limitations in the query capabilities and analysis methods proposed to users. The HyperCarte consortium with LIG, Géographie-cité and UMR RIATE proposes innovative solutions to these problems. Our approach deals with various areas such as spatio-temporal modeling, parallel computing and cartographic visualization that are related to spatial organizations of social phenomena.

Nowadays, analysis are done on huge heterogeneous data set. For example, demographic data sets at nuts 5 level, represent more than 100.000 territorial units with 40 social attributes. Many algorithms of spatial analysis, in particular potential analysis are quadratic in the size of the data set. Then adapted methods are needed to provide “user real time” analysis tools.

4.2. Wireless Networks

Participants: Bruno Gaujal, Corinne Touati, Panayotis Mertikopoulos.

MESCAL is involved in the common laboratory between Inria and Alcatel-Lucent. Bruno Gaujal is leading the Selfnets research action. This action was started in 2008 and was renewed for four more years (from 2012 to 2016). In our collaboration with Alcatel we use game theory techniques as well as evolutionary algorithms to compute optimal configurations in wireless networks (typically 3G or LTE networks) in a distributed manner.

4.3. Cloud and Desktop Computing

Participants: Derrick Kondo, Arnaud Legrand, Olivier Richard.

The research of MESCAL on desktop grids has been very active and fruitful during the evaluation period. The main achievements concern the collection and statistical exploitation of traces in volunteer computing systems. Such models have enabled to optimize the behavior of volunteer computing systems or to extend the scope of their applicability. Such traces have also been used in SIMGRID to simulate volunteer computing systems at unprecedented scale.

MOAIS Project-Team

4. Application Domains

4.1. Outline

The scientific methodology of MOAIS consists in:

- designing algorithms with provable performance on generic theoretical models. In particular we develop randomized algorithms for distributed scheduling and approximate multi-objective optimization theory..
- implementing and evaluating those algorithms with our main softwares:
 - Kaapi for fine grain scheduling of compute-intensive applications;
 - FlowVR for coarse-grain scheduling of interactive applications;
 - TakTuk, a tool for large scale remote executions deployment.
 - Triva, for the visualization of heterogeneous parallel executions.
 - KRASH, to generate reproducible CPU load on many-cores machines.
- customizing our softwares for their use in real applications studied and developed by other partners. Applications are essential to the validation and further development of MOAIS results. Application fields are: virtual reality and scientific computing (simulation, visualization, combinatorial optimization, biology, computer algebra). Depending on the application the target architecture ranges from MPSoCs (multi-processor system on chips), multicore and GPU units to clusters and heterogeneous grids. In all cases, the performance is related to the efficient use of the available, often heterogeneous, parallel resources.

MOAIS research is not only oriented towards theory but also focuses on applicative software and hardware platforms developed with external partners. Significant efforts are made to build, manage and maintain these platforms. We are involved with other teams in four main platforms:

- SOFA, a real-time physics simulation engine (<http://www.sofa-framework.org/>);
- Grimage, a 3D modeling and high performance 3D rendering platform (<http://www.inrialpes.fr/grimage>) and its evolution with the new Kinovis platform.
- Digitalis, a 780 core cluster based on Intel Nehalem processors and Infiniband network. Digitalis is used both for batch computations and interactive applications;
- Grid'5000, the experimental national grid (<http://www.grid5000.fr/>).

4.2. Virtual Reality

Participants: Thierry Gautier, Bruno Raffin, Jean-Louis Roch.

We are pursuing and extending existing collaborations to develop virtual reality applications on PC clusters and grid environments:

- Real time 3D modeling. An on-going collaboration with the MORPHEO project focuses on developing solutions to enable real time 3D modeling from multiple cameras using a PC cluster. This work is tightly coupled to the FlowVR software. Our recent developments take two main directions:
 - to provide the user a high level of interaction and immersion in the mixed reality environment. This work is focused on the Grimage platform and its successor, the new equipex Kinovis managed by Morpheo team. The camera position and orientation need to be precisely known at anytime, requiring to develop on-line calibration approaches. The background subtraction cannot anymore be based on a static background learning for the mobile camera, required here too new algorithms.

- Distributed collaboration across distant sites. In the context of the ANR DALIAN we have developed a collaborative application where multiple users, distributed in several sites each using a real time 3D modeling platform, can meet in a virtual world with a user in Grenoble also using a similar platform. The main issues are related to data transfers that need to be carefully managed to ensure a good latency while keeping a good quality, and the development of new interaction paradigms. Focusing on distributed scientific simulation, we extend those technologies in the context of the FVNANO and PetaFlow contracts.
- Real time physical simulation. We are collaborating with the Imagine project on the SOFA simulation framework. Marie Durand a Ph.D. co-advised by François Faure (IMAGINE) and Bruno Raffin, works on parallelizing SOFA using the KAAPI programming environment. The challenge is to provide SOFA with a parallelization that is efficient (real-time) while not being invasive for SOFA programmers (usually not parallel programmer). We developed a first version using the Kaapi environment for SMP machines that relies on a mix of work-stealing and dependency graph analysis and partitioning. A second version targets machines with multiples CPUs and multiple GPUs. We extended the initial framework to support a work stealing based load balancing between CPUs and GPUs. It required to extend Kaapi to support heterogeneous tasks (GPU and CPU ones) and to adapt the work stealing strategy to limit data transfers between CPUs and GPUs (the main bottleneck for GPU computing).
- Distant collaborative work. We conduct experiments using FlowVR for running applications on Grid environments. Two kinds of experiments will be considered: collaborative work by coupling two or more distant VR environments ; large scale interactive simulation using computing resources from the grid. For these experiments, we are collaborating with the LIFO and the LABRI.
- Parallel cache-oblivious algorithms for scientific visualization. In collaboration with the CEA DAM, we have developed a cache-oblivious algorithm with provable performance for irregular meshes. Based on this work, we are studying parallel algorithms that take advantage of the shared cache usually encountered on multi-core architectures (L3 shared cache) and of hardware accelerators. In collaboration with EDF, we develop new parallel algorithms for scientific visualization (eg VTK) on multicore (phD thesis of Mathias Ettinger). We are also considering adaptive algorithms to take advantage of the new trend of computers to integrate several computing units that may have different computing abilities (multicore arithmetic and graphical processing units, eventually integrated on one chip). We study balancing workload on multi GPU and CPU architectures for scientific visualization problems.

4.3. Code Coupling and Grid Programming

Participants: François Broquedis, Thierry Gautier, Jean-Louis Roch, Vincent Danjean, Frédéric Wagner.

Code coupling aim is to assemble component to build distributed applications by reusing legacy code. The objective here is to build high performance applications for cluster and grid infrastructures.

- **Grid programming model and runtime support.** Programming the grid is a challenging problem. The MOAIS Team has a strong knowledge in parallel algorithms and develop a runtime support for scheduling grid program written in a very high level interface. The parallelism from recursive divide and conquer applications and those from iterative simulation are studied. Scheduling heuristics are based on online work stealing for the former class of applications, and on hierarchical partitioning for the latter. The runtime support provides capabilities to hide latency by computation thanks to a non-blocking one-side communication protocol and by re-ordering computational tasks.
- **Grid application deployment.** To test grid applications, we need to deploy and start programs on all used computers. This can become difficult if the real topology involves several clusters with firewall, different runtime environments, etc. The MOAIS Team designed and implemented a new tool called karun that allows a user to easily deploy a parallel application wrote with the KAAPI software. This KAAPI tool relies on the TakTuk software to quickly launch programs on all nodes. The user only

needs to describe the hierarchical networks/clusters involved in the experiment with their firewall if any.

- **Visualization of grid applications execution.** The analysis of applications execution on the grid is challenging both because of the large scale of the platform and because of the heterogeneous topology of the interconnections. To help users to understand their application behavior and to detect potential bottleneck or load unbalance, the MOAIS team designed and implemented a tool named **Triva**. This tool proposes a new three dimensional visualization model that combines topological information to space time data collected during the execution. It also proposes an aggregation mechanism that eases the detection of application load unbalance.

4.4. Safe Distributed Computations

Participants: Vincent Danjean, Thierry Gautier, Clément Pernet, Jean-Louis Roch.

Large scale distributed platforms, such as the GRID and Peer-to-Peer computing systems, gather thousands of nodes for computing parallel applications. At this scale, component failures, disconnections (fail-stop faults) or results modifications (malicious faults) are part of operation, and applications have to deal directly with repeated failures during program runs. Indeed, since failure rate in such platform is proportional to the number of involved resources, the mean time between failure is dramatically decreased on very large size architectures. Moreover, even if a middleware is used to secure the communications and to manage the resources, the computational nodes operate in an unbounded environment and are subject to a wide range of attacks able to break confidentiality or to alter the resources or the computed results. Beyond fault-tolerancy, yet the possibility of massive attacks resulting in an error rate larger than tolerable by the application has to be considered. Such massive attacks are especially of concern due to Distributed Denial of Service, virus or Trojan attacks, and more generally orchestrated attacks against widespread vulnerabilities of a specific operating system that may result in the corruption of a large number of resources. The challenge is then to provide confidence to the parties about the use of such an unbound infrastructure. The MOAIS team addresses two issues:

- fault tolerance (node failures and disconnections): based on a global distributed consistent state, for the sake of scalability;
- security aspects: confidentiality, authentication and integrity of the computations.

Our approach to solve those problems is based on the efficient checkpointing of the dataflow that described the computation at coarse-grain. This distributed checkpoint, based on the local stack of each work-stealer process, provides a causally linked representation of the state. It is used for a scalable checkpoint/restart protocol and for probabilistic detection of massive attacks.

Moreover, we study the scalability of security protocols on large scale infrastructures. Within the SHIVA contract (global competitiveness cluster Minalogic in Grenoble) and in collaboration with C-S company, the Ph.D. of Ludovic Jacquin (coadvised with the PLANETE EPI) we developed a high-rate systematic ciphering platform based on the coupling of a multicore architecture with security components (FPGA and smart card) developed by industrial partners.

4.5. Embedded Systems

Participants: Jean-Louis Roch, Guillaume Huard, Denis Trystram, Vincent Danjean.

To improve the performance of current embedded systems, Multiprocessor System-on-Chip (MPSoC) offers many advantages, especially in terms of flexibility and low cost. Multimedia applications, such as video encoding, require more and more intensive computations. The system should be able to exploit the resources as much as possible to save power and time. This challenge may be addressed by parallel computing coupled with performant scheduling. On-going work focuses on developing the scheduling and monitoring technologies developed in MOAIS for embedded systems.

In the framework of our cooperation with STM (Miguel Santana) and within the SocTrace project, we are developping tools to manage distributed large scale traces. We especially focus on visualization, developping visual aggregation techniques (Phd Damien Dosimont, started in 2/2012 advised by Guillaume Huard in collaboration with Jean-Marc Vincent).

MYRIADS Project-Team

4. Application Domains

4.1. Application Domains

The MYRIADS research activities address a broad range of applications domains. We validate our research results with selected use cases from the following application domains:

- Web services, Service oriented Applications,
- Business applications,
- Bio-informatics applications,
- Computational science applications,
- Numerical simulation.

OASIS Project-Team

4. Application Domains

4.1. Grid and Cloud Computing

Grid, peer-to-peer, group communication, mobile object systems, Cloud, fault tolerance, distribution, security, synchronisation

As distributed systems are becoming ubiquitous, Grid computing, and the more recent concept of Cloud computing are facing a major challenge for computer science: seamless access and use of large-scale computing resources, world-wide. It is believed that by providing pervasive, dependable, consistent and inexpensive access to advanced computational capabilities, large-scale computing infrastructures allow new classes of applications to emerge.

There is a need for models and infrastructures for grid and peer-to-peer computing, and we promote programming models based on communicating mobile and distributed objects and components that can allow to harness these infrastructures. Another challenge is to use, for a given computation, unused CPU cycles of desktop computers in a Local Area Network, or even from wide area interconnected nodes. This is local or wide area Computational Peer-To-Peer, a concept that can contribute to a global energy footprint reduction. This is a challenge that also appears to be pregnant in more stable and homogeneous environments compared to P2P systems, such as datacenters under the problematics known as Virtual Machines consolidation.

4.2. Service Oriented Architectures (SOA)

Distributed Service Composition, Distributed Service Infrastructure, Peer-to-Peer data storage and lookup, Autonomic Management, Large-Scale deployment and monitoring

Service Oriented Architectures aim at the integration of distributed services and more generally at the integration of distributed and heterogeneous data, at the level of the Enterprise or of the whole Internet.

The team seeks solutions to the problems encountered here, with the underlying motivation to demonstrate the usefulness of a large-scale distributed programming approach and runtime support as featured by ProActive and GCM:

- Interaction between services: the uniform usage of web services based *client-server* invocations, through the possible support of an Enterprise Service Bus, can provide a simple interoperability between them. GCM components can be exposed as web services [61], and we have conducted research and development to permit a GCM component to invoke an external web service through a client interface. We also have provided a Service Component Architecture (see SCA specifications at <http://www.oasis-opencsa.org/sca>) personality for GCM components (GCM/SCA) so they can be integrated in SCA-based applications relying on SCA bindings configured as web services. For more loosely coupled interactions between services (e.g. compliant to the Web Services Notification standard), we pursue efforts to support *publish-subscribe* interaction models. Scalability in terms of number of notified events per time unit, and full interoperability through the use of semantic web notations applied to these events are some of the key challenges the community is addressing and we too. Events also correspond to data that may be worth to store, for future analytics, besides being propagated to interested parties (in the form of the event content). Our research can thus also contribute to the Big Data domain, a currently hot topic in ICT.
- Services compositions on a possibly large set of machines: if service compositions can even be turned as autonomic activities, these capabilities will really make SOA ready for the Open Internet scale (because at such a scale, a global management of all services is not possible). For service compositions represented as GCM-based component assemblies, we are indeed exploring the use of control components put in the components membranes, acting as sensors or actuators, that can

drive the self-deployment and self-management of composite services, according to negotiated Service Level Agreements. For service orchestrations usually expressed as BPEL like processes, and expressing the *composition in time* aspect of the composition of services, supports for deployment, management, and execution capable to support dynamic adaptations are also needed. Here again we believe a middleware based upon distributed and autonomous components as GCM can really help.

4.3. Simulation tools and methodology

simulation, component-based modeling, parallel and distributed simulation, reproducibility, architecture description language

Components are being used in simulation since many years. However, given its many application fields and its high computation needs, simulation is still a challenging application for component-based programming techniques and tools.

We have been exploring the application of Oasis programming methods to simulation problems in various areas of engineering problems, but also of financial applications.

More recently, with the arrival of O. Dalle in the team, and following a work previously started in the Mascotte project-team in 2006 [59], we are pursuing research on applying distributed component-based programming techniques to simulation. More precisely, new results are sought in three directions:

- improvement of simulation tools and methodology with techniques and tools borrowed from latest research in component-based software engineering;
- improvement of simulation scalability using high performance and distributed computing techniques;
- application of the results in the previous directions, in particular to the simulation of very large-scale distributed systems, such as peer-to-peer networks.

PHOENIX Project-Team

4. Application Domains

4.1. Introduction

Building on our previous work, we are studying software development in the context of communication services, in their most general forms. That is, going beyond human-to-human interactions, and covering human-to-machine and machine-to-machine interactions. Software systems revolving around such forms of communications can be found in a number of areas, including telephony, pervasive computing, and assisted living; we view these software systems as coordinating the communication between networked entities, regardless of their nature, human, hardware or software. In this context, our three main application domains are pervasive computing, avionics and assisted living.

4.2. Pervasive Computing

Pervasive computing systems are being deployed in a rapidly increasing number of areas, including building automation and supply chain management. Regardless of their target area, pervasive computing systems have a typical architectural pattern. They aggregate data from a variety of distributed sources, whether sensing devices or software components, analyze a context to make decisions, and carry out decisions by invoking a range of actuators. Because pervasive computing systems are standing at the crossroads of several domains (*e.g.*, distributed systems, multimedia, and embedded systems), they raise a number of challenges in software development:

- *Heterogeneity.* Pervasive computing systems are made of off-the-shelf entities, that is, hardware and software building blocks. These entities run on specific platforms, feature various interaction models, and provide non-standard interfaces. This heterogeneity tends to percolate in the application code, preventing its portability and reusability, and cluttering it with low-level details.
- *Lack of structuring.* Pervasive computing systems coordinate numerous, interrelated components. A lack of global structuring makes the development and evolution of such systems error-prone: component interactions may be invalid or missing.
- *Combination of technologies.* Pervasive computing systems involve a variety of technological issues, including device intricacies, complex APIs of distributed systems technologies and middleware-specific features. Coping with this range of issues results in code bloated with special cases to glue technologies together.
- *Dynamicity.* In a pervasive computing system, devices may either become available as they get deployed, or unavailable due to malfunction or network failure. Dealing with these issues explicitly in the implementation can quickly make the code cumbersome.
- *Testing.* Pervasive computing systems are complicated to test. Doing so requires equipments to be acquired, tested, configured and deployed. Furthermore, some scenarios cannot be tested because of the nature of the situations involved (*e.g.*, fire and smoke). As a result, the programmer must resort to writing specific code to achieve ad hoc testing.

4.3. Assisted Living

Cognitive impairments (memory, attention, time and space orientation, *etc*) affect a large part of the population, including elderly, patients with brain injuries (traumatic brain injury, stroke, *etc*), and people suffering from cognitive disabilities, such as Down syndrome.

The emerging industry of digital assistive technologies provide hardware devices dedicated to specific tasks, such as a telephone set with a keyboard picturing relatives (<http://www.doro.fr>), or a device for audio and video communication over the web (<http://www.technosens.fr>). These assistive technologies apply a traditional approach to personal assistance by providing an equipment dedicated to a single task (or a limited set of tasks), without leveraging surrounding devices. This traditional approach has fundamental limitations that must be overcome to significantly improve assistive technologies:

- they are *not adaptable to one's needs*. They are generally dedicated to a task and have very limited functionalities: no networking, limited computing capabilities, a limited screen and rudimentary interaction modalities. This lack of functionality may cause a proliferation of devices, complicating the end-user life. Moreover, they are rarely designed to adapt to the cognitive changes of the user. When the requirements evolve, the person must acquire a new device.
- they are often *proprietary*, limiting innovation. As a result, they cannot cope with the evolution of users' needs.
- they have limited or *no interoperability*. As a result, they cannot rely on other devices and software services to offer richer applications.

To break this model, we propose to offer an assistive solution that is open-ended in terms of applications and entities. (1) An on-line catalog of available applications enables every user and caregiver to define personalized assistance in the form of an evolving and adapted set of applications; this catalog provides a community of developers with a mechanism to publish applications for specific daily-activity needs. (2) New types of entities can be added to a platform description to enhance its functionalities and extend the scope of future applications.

4.4. Avionics

In avionics, an aircraft can be seen as an environment full of sensors (*e.g.*, accelerometers, gyroscopes, and GPS sensors) and actuators (*e.g.*, ailerons and elevator trim). For example, a flight guidance system controls the aircraft using data produced by sensors. In a critical platform such as an aircraft, software systems have to be certified. Moreover the safety-critical nature of the avionics domain takes the form of stringent non-functional requirements, resulting in a number of challenges in software development:

- *Traceability*. Traceability is the ability to trace all the requirements throughout the development process. In the avionics certification processes, traceability is mandatory for both functional and non-functional requirements.
- *Coherence*. Functional and non-functional aspects of an application are inherently coupled. For example, dependability mechanisms can potentially deteriorate the overall performance of the application. The coherence of the requirements is particularly critical when the software evolves: even minor modifications to one aspect may tremendously impact the others, leading to unpredicted failures.
- *Separation of concerns*. Avionics platforms involve the collaboration of several experts (from low-level system to software, safety, QoS), making requirements traceability significantly more challenging. Providing development methodologies that allow a clear separation of concerns can tremendously improve traceability.

Our approach consists of enriching a design language with non-functional declarations. Such declarations allow the safety expert to specify at design time how errors are handled, guiding and facilitating the implementation of error handling code. The design is also enriched with Quality of Service (QoS) declarations such as time constraints. For each of these non-functional declarations, specific development support can be generated. We have validated this approach by developing flight guidance applications for avionics and drone systems.

PLANETE Project-Team

4. Application Domains

4.1. Next Generation Networks

The next-generation network must overcome the limitations of existing networks and allow adding new capabilities and services. Future networks should be available anytime and anywhere, be accessible from any communication device, require little or no management overhead, be resilient to failures, malicious attacks and natural disasters, and be trustworthy for all types of communication traffic. Studies should therefore address a balance of theoretical and experimental researches that expand the understanding of large, complex, heterogeneous networks, design of access and core networks based on emerging wireless and optical technologies, and continue the evolution of Internet. On the other hand, it is also highly important to design a next-generation Internet which we will call the "Future Internet" from core functionalities in order to ensure security and robustness, manageability, utility and social need, new computing paradigms, integration of new network technologies and higher-level service architectures.

To meet emerging requirements for the Internet's technical architecture, the protocols and structures that guide its operation require coordinated, coherent redesign. A new approach will require rethinking of the network functions and addressing a range of challenges. These challenges include, but are not limited to, the following examples:

- New models for efficient data dissemination;
- Coping with intermittent connectivity;
- The design of secured, privacy protecting, and robust networked systems;
- Understanding the Internet behavior;
- Building network evaluation platforms.

The following research directions are essential building blocks we are contributing to the future Internet architecture.

Towards Data-Centric Networking

From the Internet design, back to 1970, the resources to be addressed and localized are computers. Indeed, at that time there were few machines interconnected, and nobody believed this number would ever be larger than a few tens of thousand of machines. Moreover, those machines were static machines with well identified resources (e.g., a given hierarchy of files) that were explicitly requested by the users. Today, the legacy of this architecture is the notion of URLs that explicitly address specific resources on a specific machine. Even if modern architectures use caches to replicate contents with DNS redirection to make those caches transparent to the end-users, this solution is only a hack that does not solve today's real problem: Users are only interested in data and do not want anymore to explicitly address where those data are. Finding data should be a service offered by the network. In this context of data-centric network, which means that the network architecture is explicitly built to transparently support the notion of content, a data can be much more than a simple content. In such a network you can, of course, request a specific file without explicitly specifying its location, the network will transparently return the closest instance of the content. You can also request a specific service from a person without knowing its explicit network location. This is in particular the case of a VoIP or an instant messaging conversation. A data-centric architecture is much more than a simple modification of the naming scheme currently used in the Internet. It requires a major rethinking of many fundamental building blocks of the current Internet. Such networking architecture will however allow seamless handling of the tricky problematic of *episodic connectivity*. It also shifts the focus from transmitting data by geographic location, to *disseminating* it via named content. In the Planète project-team, we start to work on such data-centric architectures as a

follow-up and federating axe for three of our current activities (adaptive multimedia transmission protocols for heterogeneous networks, data dissemination paradigms and peer-to-peer systems). It is important to study such data-centric architectures considering in particular the corresponding naming problem, routing and resource allocation, reliable transport, data security and authentication, content storage.

Today's Internet is characterized by high node and link heterogeneity. Nodes may vary substantially in terms of their processing, storage, communication, and energy capabilities. They may also exhibit very different mobility characteristics, from static nodes to nodes that are considerably mobile (e.g., vehicles). Links may be wired or wireless and thus operate at widely varying rates and exhibit quite different reliability characteristics. One of the challenges of data-centric architecture is to provide access to data anytime anywhere in the presence of high degree of heterogeneity. This means that the network will not be connected all the time, due to a number of factors such as node mobility, link instability, power-aware protocols that, for example, turn nodes off periodically, etc. Additionally, disconnections may last longer than what "traditional" routing protocols (e.g., MANET routing) can handle. These types of network, a.k.a, intermittently connected networks, or even episodically connected networks, have recently received considerable attention from the networking research community. Several new routing paradigms have been proposed to handle possibly frequent, long-lived disconnections. However, a number of challenges remain, including: (1) The support of scalable and transparent integration with "traditional" routing mechanisms including wired infrastructure, infrastructure-based wireless and MANET routing. (2) The study of heuristics for selecting forwarding nodes (e.g., based on node's characteristics such as node's speed, node's resources, sociability level, node's historic, etc. (3) The design of unicast and multicast transmission algorithms with congestion and error control algorithms tailored for episodically connected networks and taking into account the intrinsic characteristics of flows. (4) The design of incentive-based mechanisms to ensure that nodes forward packets while preventing or limiting the impact of possible misbehaving nodes. The solutions proposed, which are likely to extensively use cross-layer mechanisms, will be evaluated using the methodology and the tools elaborated in our new *Experimental Platform* research direction.

On the other hand, multicast/broadcast content delivery systems are playing an increasingly important role in data-centric networking. Indeed, this is an optimal dissemination technology, that enables the creation of new commercial services, like IPTV over the Internet, satellite-based digital radio and multimedia transmission to vehicles, electronic service guide (ESG) and multimedia content distribution on DVB-H/SH networks. This is also an efficient way to share information in WiFi, WiMax, sensor networks, or mobile ad hoc infrastructures. Our goal here is to take advantage of our strong background in the domain to design an *efficient, robust (in particular in case of tough environments) and secure (since we believe that security considerations will play an increasing importance) broadcasting system*. We address this problem by focusing on the following activities: (1) The protocols and applications that enable the high level control of broadcasting sessions (like the FLUTE/ALC sessions) are currently missing. The goal is to enable the content provider to securely control the underlying broadcasting sessions, to be able to launch new sessions if need be, or prematurely stop an existing session and to have feedback and statistics on the past/current deliveries. (2) The AL-FEC building block remains the cornerstone on which the whole broadcasting system relies. The goal is to design and evaluate new codes, capable of producing a large amount of redundancy (thereby approaching rateless codes), over very large objects, while requiring a small amount of memory/processing in order to be used on lightweight embedded systems and terminals. (3) The security building blocks and protocols that aim at providing content level security, protocol level security, and network level security must be natively and seamlessly integrated. This is also true of the associated protocols that enable the initialization of the elementary building blocks (e.g. in order to exchange security parameters and keys). Many components already exist. The goal here is to identify them, know how to optimally use them, and to design/adapt the missing components, if any. (4) It is important seamlessly integrated these broadcasting systems to the Internet, so that users can benefit from the service, no matter where and how he is attached to the network. More precisely we will study the potential impacts of a merge of the broadcasting networks and the Internet, and how to address them. For instance there is a major discrepancy when considering flow control aspects, since broadcasting networks are using a constant bit rate approach while the Internet is congestion controlled.

When a native broadcasting service is not enabled by the network, data should still be able to be disseminated to a large population in a scalable way. A peer-to-peer architecture supports such an efficient data dissemination. We have gained a fundamental understanding of the key algorithms of BitTorrent on the Internet. We plan to continue this work in two directions. First, we want to study how a peer-to-peer architecture can be natively supported by the network. Indeed, the client-server architecture is not robust to increase in load. The consequence is that when a site becomes suddenly popular, it usually becomes unreachable. The peer-to-peer architecture is robust to increase in load. However, a native support in the network of this architecture is a hard problem as it has implications on many components of the network (naming, addressing, transport, localization, etc.). Second, we want to evaluate the impact of wireless and mobile infrastructures on peer-to-peer protocols. This work has started with the European project Expeshare. The wireless medium and the mobility of nodes completely change the properties of peer-to-peer protocols. The dynamics becomes even more complex as it is a function of the environment and of the relative position of peers.

Network security and Privacy

The Internet was not designed to operate in a completely open and hostile environment. It was designed by researchers that trust each other and security at that time was not an issue. The situation is quite different today and the Internet community has drastically expanded. The Internet is now composed of more than 300 millions computers worldwide and the trust relationship has disappeared. One of the reason of the Internet success is that it provides ubiquitous inter-connectivity. This is also one of the its main weakness since it allows to launch attacks and to exploit vulnerabilities in a large-scale basis. The Internet is vulnerable to many different attacks, for example, Distributed Denial-of Service (DDoS) attacks, epidemic attacks (Virus/Worm), spam/phishing and intrusion attacks. The Internet is not only insecure but it also infringes users' privacy. Those breaches are due to the Internet protocols but also to new applications that are being deployed (VoIP, RFID,...). A lot of research is required to improve the Internet security and privacy. For example, more research work is required to understand, model, quantify and hopefully eliminate (or at least mitigate) existing attacks. Furthermore, more and more small devices (RFIDs or sensors) are being connected to the Internet. Current security/cryptographic solutions are too expensive and current trust models are not appropriate. New protocols and solutions are required : security and privacy must be considered in the Internet architecture as an essential component. The whole Internet architecture must be reconsidered with security and privacy in mind. Our current activities in this domain are on security in wireless, ad hoc and sensor networks, mainly the design of new key exchange protocols and of secured routing protocols. We also work on location privacy techniques, authentication cryptographic protocols and opportunistic encryption. We plan to continue our research on wireless security, and more specifically on WSN and RFID security focusing on the design of real and deployable systems. We started a new research topic on the security of the Next-Generation Internet. The important goal of this new task is to rethink the architecture of the Internet with security as a major design requirement, instead of an after-thought.

Wireless Sensor Networks: A lot of work has been done in the area of WSN security in the last years, but we believe that this is still the beginning and a lot of research challenges need to be solved. On the one hand it is widely believed that the sensor networks carry a great promise: Ubiquitous sensor networks will allow us to interface the physical environment with communication networks and the information infrastructure, and the potential benefits of such interfaces to society are enormous, possibly comparable in scale to the benefits created by the Internet. On the other hand, as with the advent of the Internet, there is an important associated risk and concern: How to make sensor network applications resilient and survivable under hostile attacks? We believe that the unique technical constraints and application scenarios of sensor networks call for new security techniques and protocols that operate above the link level and provide security for the sensor network application as a whole. Although this represents a huge challenge, addressing it successfully will result in a very high pay-off, since targeted security mechanisms can make sensor network operation far more reliable and thus more useful. This is the crux of our work. Our goal here is to design new security protocols and algorithms for constrained devices and to theoretically prove their soundness and security. Furthermore, to

complement the fundamental exploration of cryptographic and security mechanisms, we will simulate and evaluate these mechanisms experimentally.

RFID: As already mentioned, the ubiquitous use of RFID tags and the development of what has become termed "the Internet of things" will lead to a variety of security threats, many of which are quite unique to RFID deployment. Already industry, government, and citizens are aware of some of the successes and some of the limitations or threats of RFID tags, and there is a great need for researchers and technology developers to take up some of daunting challenges that threaten to undermine the commercial viability of RFID tags on the one hand, or to the rights and expectations of users on the other. We will focus here on two important issues in the use of RFID tags: (1) *Device Authentication*: allows us to answer several questions such as: Is the tag legitimate? Is the reader a tag interacts with legitimate? (2) *Privacy*: is the feature through which information pertaining to a tag's identity and behavior is protected from disclosure by unauthorized parties or by unauthorized means by legitimate parties such as readers. In a public library, for example, the information openly communicated by a tagged book could include its title or author. This may be unacceptable to some readers. Alternatively, RFID-protected pharmaceutical products might reveal a person's pathology. Turning to authenticity, if the RFID tag on a batch of medicines is not legitimate, then the drugs could be counterfeit and dangerous. Authentication and privacy are concepts that are relevant to both suppliers and consumers. Indeed, it is arguable that an RFID deployment can only be successful if all parties are satisfied that the integrity between seller and buyer respects the twin demands of authentication and privacy. Our main goal here, therefore, is to propose and to prototype the design of cryptographic algorithms and secure protocols for RFID deployment. These algorithms and protocols may be used individually or in combination, and we anticipate that they will aid in providing authentication or privacy. One particular feature of the research in the RFID-AP project is that the work must be practical. Many academic proposals can be deeply flawed in practice since too little attention has been paid to the realities of implementation and deployment. This activity will therefore be notable for the way theoretical work will be closely intertwined with the task of development and deployment. The challenges to be addressed in the project are considerable. In particular there are demanding physical limits that apply to the algorithms and protocols that can be implemented on the cheapest RFID tags. While there often exist contemporary security solutions to issues such as authentication and privacy, in an RFID-based deployment they are not technically viable. And while one could consider increasing the technical capability of an RFID-tag to achieve a better range of solutions, the solution is not economically viable.

Next Generation Internet Security: The current Internet has reached its limits; a number of research groups around the world are already working on future Internet architectures. The new Internet should have built-in security measures and support for wireless communication devices, among other things. A new network design is needed to overcome unwanted traffic, malware, viruses, identity theft and other threats plaguing today's Internet infrastructure and end hosts. This new design should also enforce a good balance between privacy and accountability. Several proposals in the area have been made so far, and we expect many more to appear in the near future. Some mechanisms to mitigate the effects of security attacks exist today. However, they are far from perfect and it is a very open question how they will behave on the future Internet. Cyber criminals are very creative and new attacks (e.g. VoIP spam, SPIT) appear regularly. Furthermore, the expectation is that cyber criminals will move into new technologies as they appear, since they offer new attack opportunities, where existing countermeasures may be rendered useless. The ultimate goal of this research activity is to contribute to the work on new Internet architecture that is more resistant to today's and future security attacks. This goal is very challenging, since some of future attacks are unpredictable. We are analyzing some of the established and some of the new architectural proposals, attempting to identify architectural elements and patterns that repeat from one architectural approach to another, leading to understanding how they impact the unwanted traffic issue and other security issues. Some of the more prominent elements are rather easy to identify and understand, such as routing, forwarding, end-to-end security, etc. Others may well be much harder to identify, such as those related to data-oriented networking, e.g., caching. The motivation for this work is that the clean slate architectures provide a unique opportunity to provide built in security capabilities that would enable the prevention of phenomenon like unwanted traffic. New architectures will most likely introduce additional name-spaces for the different fundamental objects in the network and in particular for routing objects. These

names will be the fundamental elements that will be used by the new routing architectures and security must be a key consideration when evaluating the features offered by these new name-spaces.

Network Monitoring

The Planète project-team contributes to the area of network monitoring. Our focus is on the monitoring of the Internet for the purpose of access quality assessment, problem detection and troubleshooting. Indeed, in the absence of an advanced management and control plan in the Internet, and given the simplicity of the service provided by the core of the network and the increase in its heterogeneity, it is nowadays common that users experience a service degradation and are unable to understand the reasons for the access quality they perceive. Problems at the access can be in the form of a pure disconnection, a decrease in the bandwidth or an increase in the delay or loss rate of packets. Service degradation can be caused by protocol anomalies, an attack, an increase in the load, or simply a problem at the source or destination machines. Actually, it is not easy to diagnose the reasons for service degradation. Basic tools exist as ping and trace-route, but they are unable to provide detailed answers on the source of the problem nor on its location. From operator point of view, the situation is not better since an operator has only access to its own network and can hardly translate local information into end-to-end measurements. The increase in the complexity of networks as is the case of wireless mesh networks will not ease the life of users and operators. The purpose of our work in this direction is to study to which extent one can troubleshoot the current Internet and estimate the quality at the access either with end-to-end solutions or core network solutions. Our aim is to propose an architecture that allows end-users by collaborating together to infer the reasons for service degradation and to estimate the quality of access they perceive. This architecture can be purely end-to-end or can rely on some information from the core of the network as BGP routing information. We will build on this study to understand the limitations in the current Internet architecture and propose modifications that will ease the troubleshooting and make it more efficient in future network architectures. The proposed architecture will be the subject of validation over large scale experimental platforms as PlanetLab and OneLab.

Experimental Environment for future Internet architecture

The Internet is relatively resistant to fundamental change (differentiated services, IP multicast, and secure routing protocols have not seen wide-scale deployment). A major impediment to deploy these services is the need for coordination: an Internet service provider (ISP) that deploys the service garners little benefit until other domains follow suit. Researchers are also under pressure to justify their work in the context of a federated network by explaining how new protocols could be deployed one network at a time, but emphasizing incremental deployability does not necessarily lead to the best architecture. In fact, focusing on incremental deployment may lead to solutions where each step along the path makes sense, but the end state is wrong. The substantive improvements to the Internet architecture may require fundamental change that is not incrementally deployable.

Network virtualisation has been proposed to support realistic large scale shared experimental facilities such as PlanetLab and GENI. We are working on this topic in the context of the European OneLab project.

Testing on PlanetLab has become a nearly obligatory step for an empirical research paper on a new network application or protocol to be accepted into a major networking conference or by the most prestigious networking journals. If one wishes to test a new video streaming application, or a new peer-to-peer routing overlay, or a new active measurement system for geo-location of internet hosts, hundreds of PlanetLab nodes are available for this purpose. PlanetLab gives the researcher login access to systems scattered throughout the world, with a Linux environment that is consistent across all of them.

However, network environments are becoming ever more heterogeneous. Third generation telephony is bringing large numbers of handheld wireless devices into the Internet. Wireless mesh and ad-hoc networks may soon make it common for data to cross multiple wireless hops while being routed in unconventional ways. For these new environments, new networking applications will arise. For their development and evaluation, researchers and developers will need the ability to launch applications on endhosts located in these different environments.

It is sometimes unrealistic to implement new network technology, for reasons that can be either technological - the technology is not yet available -, economical - the technology is too expensive -, or simply pragmatical - e.g. when actual mobility is key. For these kinds of situations, we believe it can be very convenient and powerful to resort to emulation techniques, in which real packets can be managed as if they had crossed, e.g., an ad hoc network.

In our project-team, we work to provide a realistic environment for the next generation of network experiments. Such a large scale, open, heterogeneous testbed should be beneficial to the whole networking academic and industrial community. It is important to have an experimental environment that increases the quality and quantity of experimental research outcomes in networking, and to accelerate the transition of these outcomes into products and services. These experimental platforms should be designed to support both research and deployment, effectively filling the gap between small-scale experiments in the lab, and mature technology that is ready for commercial deployment. As said above, in terms of experimental platforms, the well-known PlanetLab testbed is gaining ground as a secure, highly manageable, cost-effective world-wide platform, especially well fitted for experiments around New Generation Internet paradigms like overlay networks. The current trends in this field, as illustrated by the germinal successor known as GENI, are to address the following new challenges. Firstly, a more modular design will allow to achieve federation, i.e. a model where reasonably independent Management Authorities can handle their respective subpart of the platform, while preserving the integrity of the whole. Secondly, there is a consensus on the necessity to support various access and physical technologies, such as the whole range of wireless or optical links. It is also important to develop realistic simulators taking into account the tremendous growth in wireless networking, so to include the many variants of IEEE 802.11 networking, emerging IEEE standards such as WiMax (802.16), and cellular data services (GPRS, CDMA). While simulation is not the only tool used for data networking research, it is extremely useful because it often allows research questions and prototypes to be explored at many orders-of-magnitude less cost and time than that required to experiment with real implementations and networks.

Simulations allow a fast evaluation process, fully controlled scenarios, and reproducibility. However, they lack realism and the accuracy of the models implemented in the simulators is hard to assess. Emulation allows controlled environment and reproducibility, but it also suffers from a lack of realism. Experiments allow more realistic environment and implementations, but they lack reproducibility and ease of use. Therefore, each evaluation technique has strengths and weaknesses. However, there is currently no way to combine them in a scientific experimental workflow. Typical evaluation workflows are split into four steps: topology description and construction, traffic pattern description and injection, trace instrumentation description and configuration, and, analysis based on the result of the trace events and the status of the environment during the experimentation. To achieve the integration of experimental workflows among the various evaluation platforms, the two following requirements must be verified:

- **Reproducibility:** A common interface for each platform must be defined so that a same script can be run transparently on different platforms. This also implies a standard way to describe scenarios, which includes the research objective of the scenario, topology description and construction, the description of the traffic pattern and how it is injected into the scenario, the description and configuration of the instrumentation, and the evolution of the environment during the experimentation
- **Comparability:** As each platform has different limitations, a way to compare the conclusions extracted from experiments run on different platforms, or on the same platform but with different conditions (this is in particular the case for the wild experimental platforms) must be provided.

Benchmarking is the function that provides a method of comparing the performance of various subsystems across different environments. Both reproducibility and comparability are essential to benchmarking. In

order to facilitate the design of a general benchmarking methodology, we plan to integrate and automate a networking experiments workflow within the OneLab platform. This requires that we:

- Automate the definition of proper scenario definition taking in consideration available infra-structure to the experiment.
- Automate the task of mapping the experimentation topology on top of the available OneLab topology. We propose to first focus on a simple one-to-one node and link mapping the beginning.
- Define and provide extensive instrumentation sources within the OneLab system to allow users to gather all interesting trace events for offline analysis
- Measure and provide access to "environment variables" which measure the state of the OneLab system during an experimentation
- Define an offline analysis library which can infer experimentation results and comparisons based on traces and "environment variables".

To make the use of these components transparent, we plan to implement them within a simulation-like system which should allow experiments to be conducted within a simulator and within the OneLab testbed through the same programming interface. The initial version will be based on the ns-3 programming interface.

RAP Project-Team (section vide)

REGAL Project-Team

4. Application Domains

4.1. Research domain

To address the evolution of distributed platforms in recent years, we focus on the following areas:

- *Distributed algorithms for dynamic and large networks.* Network topology is no more static; distributed systems are increasingly dynamic, i.e., nodes can join, fail, recover, disconnect and reconnect, and change location. Examples include IaaS cloud computing infrastructures, where virtual machines can be moved according to load peaks, opportunistic networks such as DTNs (Delay-Tolerant Networks), and networks of robots.
- *Management of distributed data.* In emerging architectures such as distributed hash tables (DHTs) and cloud computing, our research topics include replica placement, responsiveness, load balancing, consistency maintenance, consensus algorithms, and synchronisation. This research direction is funded by several new collaborative projects (ConcoRDanT, MyCloud, Nu@age, Odisea, Prose, Shaman, Spades, Streams, R-Discover) and by industrial funding (Google).
- *Performance and robustness of Systems Software in multicore architectures.* Our research focuses on the efficient management of system resources at the user level. Issues considered include efficient synchronization and memory management in large-scale multicore architectures. At the same time, we focus on the robustness of systems software, based on the Coccinelle technology. This work is funded by ANR ABL and InfraJVM.

RMOD Project-Team (section vide)

ROMA Team

3. Application Domains

3.1. Application of sparse direct solvers

Sparse direct (multifrontal) solvers in distributed-memory environments have a wide range of applications as they are used at the heart of many numerical methods in simulation: whether a model uses finite elements or finite differences, or requires the optimization of a complex linear or nonlinear function, one often ends up solving a linear system of equations involving sparse matrices. There are therefore a number of application fields, among which some of the ones cited by the users of our sparse direct solver MUMPS (see Section 4.1) are: structural mechanics, biomechanics, medical image processing, tomography, geophysics, ad-hoc networking modeling (e.g., Markovian processes), electromagnetics, fluid dynamics, econometric models, oil reservoir simulation, magneto-hydro-dynamics, chemistry, acoustics, glaciology, astrophysics, circuit simulation.

RUNTIME Project-Team

4. Application Domains

4.1. Application Domains

HPC, simulation

The RUNTIME group is working on the design of efficient runtime systems for parallel architectures. We are currently focusing our efforts on High Performance Computing applications that merely implement numerical simulations in the field of Seismology, Weather Forecasting, Energy, Mechanics or Molecular Dynamics. These time-consuming applications need so much computing power that they need to run over parallel machines composed of several thousands of processors.

Because the lifetime of HPC applications often spreads over several years and because they are developed by many people, they have strong portability constraints. Thus, these applications are mostly developed on top of standard APIs (e.g. MPI for communications over distributed machines, OpenMP for shared-memory programming). That explains why we have long standing collaborations with research groups developing parallel language compilers, parallel programming environments, numerical libraries or communication software. Actually, all these “clients” are our primary target.

Although we are currently mainly working on HPC applications, many other fields may benefit from the techniques developed by our group. Since a large part of our efforts is devoted to exploiting multicore machines and GPU accelerators, many desktop applications could be parallelized using our runtime systems (e.g. 3D rendering, etc.).

SARDES Project-Team (section vide)

SCORE Team

4. Application Domains

4.1. Introduction

Our research is inspired by potential problems and use cases from the real world. The rapid pace of technology evolution oblige us to adapt our perspective on some of the problems we consider. Recently, the raise of the "As a Service" movement has both validated our past work but also obliged us to rethink some of our directions. Then, we have chosen to apply our work to domains because we think that the problems that they exhibit resonate with our work. This is the case with Crisis Management and with Software Engineering that are in some ways extreme case for collaboration and service consumption. We are also exploring more general business oriented applications.

4.2. E-government

E-government is a well established domain that provides its own requirements in the field of service and information management. From our perspective, e-government applications have very strong requirements regarding security, privacy and interoperability between different organizations, belonging potentially to different countries. One of the prominent contributions we have made in this domain is related to our collaboration with SAP on the relationship between processes, security policies and the problem of delegation that we considered as a important for organizational flexibility. This work resonate also with its current continuation in crisis management.

Crisis management is a special case of e-government application as it involves mostly governmental agencies in coordination with other organizations like the Red Cross or other NGO. Moreover, it brings with it a lot of requirements that are very interesting for us in the domain of coordination: a crisis process shall be very flexible, adaptable and distributed. It is mostly human driven and can be critical. In this domain, we are collaborating with SAP to define a new model of coordination that should support people involved in crisis resolution.

4.3. Groupware Systems and Software Forge

Software engineering environment can be seen as distributed collaborative systems. Software Forges are social software (e.g. Github). They transform strangers into collaborators. Forges are online services that allow instantiating, composing and managing collaborative services. Traditionally, collaborative services are version control systems, issue trackers, forums, mailing lists or wikis. We are applying our research results on coordination and data sharing into this context foreseeing a forge, not as a monolithic framework but as a composition of services that can be deployed in a Cloud infrastructure.

SOCRATE Team

4. Application Domains

4.1. Example of SDR applications

SDR concept is not new and many research teams have been working on its implementation and use in various contexts, however two elements are in favor of Socrate's orientation towards this technology:

1. The mobile SDR technology is becoming mature. Up to now, Software-Defined Radio terminals were too expensive and power consuming for mobile terminal, this should change soon. For instance, CEA's Magali platform has demonstrated part of LTE-Advanced standard recently. It is important for applied researchers to be ready when a new technology rises up, opening to many new software issues.
2. Rhône-Alpes is a strategic place for this emerging technology with important actors such as ST-Microelectronics, CEA, Minalogic and many smaller actors in informatics for telecommunication and embedded systems.

SDR technologies enables the following scenarios:

- *Transparent radio adaptation*: Depending on the available wireless protocols in the air (e.g. Wifi versus UMTS), a terminal may choose to communicate on the cheapest, or the fastest channel.
- *Radio resource allocation*: In order to minimize expensive manual cell planning and achieve "tighter" frequency reuse patterns, resulting in improved system spectral efficiency, dynamic radio resource management is a promising application of SDR.
- *White space*: By sensing the air, a terminal is able to communicate using a particular frequency which is not used even if it is reserved for another kind of application.
- *Cooperation*: Using the neighboring terminals, a user can reduce power consumption by using relay communication with the base station.
- *Saturated bands*: A fixed wireless object, e.g. a gas meter sending regular data through the air, might check if the frequency it uses is saturated and choose, alone or in a distributed manner with other gas meters, to use another frequency (or even protocol) to communicate.
- *Radars*: With numerical communications, passive radar technology is changing, these radars will have to be updated regularly to be able to listen to new communication standards.
- *Internet of things*: With the predicted huge venue of wireless object, some reconfigurability will be needed even on the simplest smart object as mentioned above for facing the band saturation problem or simply communicating in a new environment.

4.2. Public wireless access networks

The commercial markets for wireless technologies are the largest markets for SDR and cognitive radio. these markets includes *i*) the cellular market (4G, LTE), *ii*) the Wireless Local Area Network market (WLAN, e.g. Wifi), and *iii*) the Broadband Wireless Access market (e.g. WiMax). The key objective here is to improve spectrum efficiency and availability, and to enable cognitive radio and SDR to support multimedia and multi-radio initiatives.

The future mobile radio access network referred to as 4G (4th generation) is expected to provide a wireless access of 100 Mbps in extended mobility and up to 1Gbps in reduced mobility as defined by the group IMT-Advanced of the ITU-R(adiocommunication) section. On the road towards the 4G, IMT-2000 standards evolutions are driven by the work of the WiMAX forum (IEEE 802.16e) on the one hand and by those of the LTE (Long Term Evolution) group of the 3GPP on the other hand. Both groups announced some targeted evolutions that could comply with the 4G requirements, namely the Gigabit Wimax (802.16m) and the LTE-Advanced proposal from the 3GPP.

In both technologies, the scarcity of the radio spectrum is taken care of by the use of MIMO and OFDMA technologies, combining the dynamic spatial and frequency multiple access. However, a better spectral efficiency will be achieved if the radio spectrum can be shared dynamically between primary and secondary networks, and if the terminals are reconfigurable in real-time. Socrate is active in this domain because of its past activity in Swing and its links to the telecommunication teaching department of Insa. The development of the FIT platform [36] is a strong effort in this area.

4.3. Military SDR and Public Safety

Military applications have developed specific solutions for SDR. In France, Thales is a major actor (e.g. project Essor defining inter-operability between European military radio) and abroad the Joint Tactical Radio System, and Darpa focus on Mobile Ad-hoc Networks (MANETS) have brought important deliverables, like the Software Communications Architecture (SCA) for instance [37].

Recent natural disasters have brought considerable attention to the need of enhanced public safety communication abroad [35]. Socrate is not currently implied in any military or public safety research programs but is aware of the potential importance this domain may take in Europe in a near future.

4.4. Ambient Intelligence: WSN and IoT

Sensor networks have been investigated and deployed for decades already; their wireless extension, however, has witnessed a tremendous growth in recent years. This is mainly attributed to the development of wireless sensor networks (WSNs): a large number of sensor nodes, reliably operating under energy constraints. It is anticipated that within a few years, sensors will be deployed in a variety of scenarios, ranging from environmental monitoring to health care, from the public to the private sector. Prior to large-scale deployment, however, many problems have to be solved, such as the extraction of application scenarios, design of suitable software and hardware architectures, development of communication and organization protocols, validation and first steps of prototyping, etc. The Citi laboratory has a long experience in WSN which led recently to the creation of a start-up company, led by two former Citi members: HIKOB(<http://openlab.hikob.com>).

The Internet of Things (IoT) paradigm is defined as a very large set of systems interconnected to provide a virtual twin world interacting with the real world. In our work we will mostly focus on wireless systems since the wireless link is the single media able to provide a full mobility and ubiquitous access. Wireless IoT is not a reality yet but will probably result from the convergence between mobile radio access networks and wireless sensor networks. If radio access networks are able to connect almost all humans, they would fail to connect a potential of several billions of objects. Nevertheless, the mutation of cellular systems toward more adaptive and autonomous systems is on going. This is why Socrate develops a strong activity in this applicative area, with its major industrial partners: Orange Labs and Alcatel-Lucent Bell labs.

For instance, the definition of a *smart node* intermediate between a WSN and a complex SDR terminal is one of the research directions followed in Socrate, explicitly stated in the ADT Snow project. Other important contributions are made in the collaboration with SigFox and Euromedia and in the EconHome project.

4.5. Body Area Networks

Body Area Network is a relatively new paradigm which aims at promoting the development of wireless systems in, on and around the human body. Wireless Body Area Networks (BAN) is now a well known acronym which encompasses scenarios in which several sensors and actuators are located on or inside the human body to sense different data, e.g. physiological information, and transfer them wirelessly towards a remote coordination unit which processes, forwards, takes decisions, alerts, records, etc. The use of BAN spans a wide area, from medical and health care to sport through leisure applications, which definitely makes the definition of a standard air interface and protocol highly challenging. Since it is expected that such devices and networks would have a growing place in the society and become more stringent in terms of quality of service, coexistence issues will be critical. Indeed, the radio resource is known to be scarce. The recent regulation difficulties of UWB systems as well as the growing interest for opportunistic radios show that any new system

have to make an efficient use of the spectrum. This also applies to short range personal and body area network systems which are subject to huge market penetrations.

Socrate was involved in the Banet ANR project (2008-2010), in which we contributed to the development of a complete PHY/MAC standard in cooperation with Orange Labs and CEA Leti, who participated to the standardization group 802.15.6. Recently, Inria has been added as a partner the FET flagship untitled *Guardian Angels* (<http://www.fet-f.eu/>), an important european initiative to develop the BANS of the futur.

We consider that BANS will probably play an important role in the future of Internet as the multiple objects connected on body could also be connected to Internet by the mobile phone hosted by each human. Therefore the BAN success really depends on the convergence of WSN and radio access networks, which makes it a very interesting applicative framework for Socrate team.

TREC Project-Team

4. Application Domains

4.1. Application Domains

We have investigated various applications of our research results with the following industrial partners and user associations:

- **Wireless Networks**
 - Alcatel-Lucent Bell Laboratories (L. Thomas and L. Roulet) on self optimization in cellular networks.
 - Qualcomm (T. Richardson and his group) on improvements of CSMA CA.
 - Orange (M. Karray) on cellular networks.
- **Network Dynamics**
 - Thalès and Real-Time-at-Work on embedded networks.
 - Grenouille on probing in access networks.
- **Networks Economics**
 - Technicolor (J. Bolot) on economic incentives.

TRISKELL Project-Team

4. Application Domains

4.1. Application Domains

SOA, telecommunication, distributed systems, Embedded Systems, software engineering, test, UML

From small embedded systems such as home automation products or automotive systems to medium sized systems such as medical equipment, office equipment, household appliances, smart phones; up to large Service Oriented Architectures (SOA), building a new application from scratch is no longer possible. Such applications reside in (group of) machines that are expected to run continuously for years without unrecoverable errors. Special care has then to be taken to design and validate embedded software, making the appropriate trade-off between various extra-functional properties such as reliability, timeliness, safety and security but also development and production cost, including resource usage of processor, memory, bandwidth, power, etc.

Leveraging ongoing advances in hardware, embedded software is playing an evermore crucial role in our society, bound to increase even more when embedded systems get interconnected to deliver ubiquitous SOA. For this reason, embedded software has been growing in size and complexity at an exponential rate for the past 20 years, pleading for a component based approach to embedded software development. There is a real need for flexible solutions allowing to deal at the same time with a wide range of needs (product lines modeling and methodologies for managing them), while preserving quality and reducing the time to market (such as derivation and validation tools).

We believe that building flexible, reliable and efficient embedded software will be achieved by reducing the gap between executable programs, their models, and the platform on which they execute, and by developing new composition mechanisms as well as transformation techniques with a sound formal basis for mapping between the different levels.

Reliability is an essential requirement in a context where a huge number of softwares (and sometimes several versions of the same program) may coexist in a large system. On one hand, software should be able to evolve very fast, as new features or services are frequently added to existing ones, but on the other hand, the occurrence of a fault in a system can be very costly, and time consuming. While we think that formal methods may help solving this kind of problems, we develop approaches where they are kept “behind the scene” in a global process taking into account constraints and objectives coming from user requirements.

Software testing is another aspect of reliable development. Testing activities mostly consist in trying to exhibit cases where a system implementation does not conform to its specifications. Whatever the efforts spent for development, this phase is of real importance to raise the confidence level in the fact that a system behaves properly in a complex environment. We also put a particular emphasis on on-line approaches, in which test and observation are dynamically computed during execution.

URBANET Team

4. Application Domains

4.1. Smart infrastructure

Unlike the communication infrastructure that went through a continuous development in the last decades, the distribution networks in our cities, whether we are talking about water, gas, or electricity, are still based on 19th century infrastructure. With the introduction of new methods for producing renewable but unpredictable energy and with the increased attention towards environmental problems, modernizing distribution networks becomes one of the major concerns in the urban world. An essential component of these enhanced systems is their integration with information and communications technology, the result being a smart distribution infrastructure, with improved efficiency and reliability. This evolution is mainly based on the increased deployment of automatic equipment and the use of machine-to-machine and sensor-to-actuator communications that would allow taking into account the behavior and needs of both consumers and suppliers.

Another fundamental urban infrastructure is the transportation system. The progress achieved by the transportation industry over the last century has been an essential factor in the development of today's urban society, while also triggering the birth and growth of other economic branches. However, the current transportation system has serious difficulties coping with the continuous growth in the number of vehicles, especially in an urban environment. As a major increase in the capacity of a city road infrastructure, already in place for tens or even hundreds of years, would imply dissuasive costs, the more realistic approach is to optimize the use of the existing transportation system. As in the case of distribution networks, the intelligence of the system will be obtained by the integration of information and communication capabilities. However, for smart transportation the challenges are somehow different, because the intelligence is no longer limited to the infrastructure, but propagates to vehicles themselves. Moreover, the degree of automation is reduced in transportation systems, as most actions resulting in reduced road congestion, higher reliability or improved safety must come from the human driver (at least in the foreseeable future).

Finally, smart spaces are becoming an essential component of our cities. The classical architectural tools used to design and shape the urban environment are more and more challenged by the idea of automatically modifying private and public spaces in order to adapt to the requirements and preferences of their users. Among the objectives of this new urban planning current, we can find the transformation of the home in a proactive health care center, fast reconfigurable and customizable workplaces, or the addition of digital content in the public spaces in order to reshape the urban scene. Bringing these changing places in our daily lives is conditioned by a major shift in the construction industry, but it also involves important advancements in digital infrastructure, sensing, and communications

4.2. Urban sensing

Urban sensing can be seen as the same evolution of the environment digitalization as social networking has been for information flows. Indeed, besides dedicated and deployed sensors and actuators, still required for specific sensing operations such as the real-time monitoring of pollution levels, there is a wide range of relevant urban data that can be collected without the need for new communication infrastructures, leveraging instead on the pervasiveness of smart mobile terminals. With more than 80% of the population owning a mobile phone, the mobile market has a deeper penetration than electricity or safe drinking water. Originally designed for voice transmitted over cellular networks, mobile phones are today complete computing, communication and sensing devices, offering in a handheld device multiple sensors and communication technologies.

Mobile devices such as smartphones or tablets are indeed able to gather a wealth of informations through embedded cameras, GPS receivers, accelerometers, and cellular, WiFi and bluetooth radio interfaces. When collected by a single device, such data may have small value per-se, however its fusion over large scales could prove critical for urban sensing to become an economically viable mainstream paradigm.

This is even more true when less traditional mobile terminals are taken into account: privately-owned cars, public transport means, commercial fleets, and even city bikes are starting to feature communication capabilities and the Floating Car Data (FCD) they generate can bring a dramatic contribution to the cause of urban sensing. Indeed, other than enlarging the sensing scope even further, e.g., through Electronic Control Units (ECUs), these mobile terminals are not burdened by strong energy constraints and can thus significantly increase the granularity of data collection.

This data can be used by authorities to improve public services, or by citizens who can integrate it in their choices. However, in order to kindle this hidden information, important problems related to data gathering, aggregation, communication, data mining, or even energy efficiency need to be solved.

4.3. User-centric services

What is the most disobeyed traffic sign in your city? How does the level of pollution on your street compare with the one in other neighborhoods? How long is the queue at that exhibition you were planning to attend today? Combining location awareness and data recovered from multiple sources like social networks or sensing devices can provide answers to all these questions, making visible previously unknown characteristics of the urban environment.

Beyond letting their own devices or vehicles autonomously harvest data from the environment through embedded or onboard sensors, mobile users can actively take part in the participatory sensing process because they can, in return, benefit from citizen-centric services which aim at improving their experience of the urban life. Crowdsourcing applications have the potential to turn citizens into both sources of information and interactive actors of the city. It is not a surprise that emerging services built on live mobile user feedback are rapidly meeting a large success.

In particular, improving everyone's mobility is probably one of the main services that a smart city shall offer to its inhabitants and visitors. This implies providing, through network broadcast data or urban smart-furniture, an accurate and user-tailored information on where people should head in order to find what they are looking for (from a specific kind of shop to a free parking slot), on their current travel time estimates, on the availability of better alternate means of transport to destination. Depending on the context, such information may need to be provided under hard real-time constraints, e.g., in presence of road accidents, unauthorized public manifestations, or delayed public transport schedules.

In some cases, information can also be provided to mobile users so as to bias or even enforce their mobility: drivers can be alerted of the arrival of an emergency vehicle so that they leave the leftmost lane available, or participants leaving vast public events can be directed out of the event venue through diverse routes displayed on their smartphones so as to dynamically balance the pedestrian flows and reduce their waiting times.

ALICE Project-Team (section vide)

ALPAGE Project-Team

4. Application Domains

4.1. Panorama

NLP tools and methods have many possible domains of application. Some of them are already mature enough to be commercialized. They can be roughly classified in three groups:

Human-computer interaction : mostly speech processing and text-to-speech, often in a dialogue context; today, commercial offers are limited to restricted domains (train tickets reservation...);

Language writing aid : spelling, grammatical and stylistic correctors for text editors, controlled-language writing aids (e.g., for technical documents), memory-based translation aid, foreign language learning tools, as well as vocal dictation;

Access to information : tools to enable a better access to information present in huge collections of texts (e.g., the Internet): automatic document classification, automatic document structuring, automatic summarizing, information acquisition and extraction, text mining, question-answering systems, as well as surface machine translation. Information access to speech archives through transcriptions is also an emerging field.

Experimental linguistics : tools to explore language in an objective way (this is related, but not limited to corpus linguistics).

Alpage focuses on some applications included in the three last points, such as information extraction and (linguistic and extra-linguistic) knowledge acquisition (4.2), text mining (4.3), spelling correction (4.5) and experimental linguistics (4.6).

4.2. Information extraction and knowledge acquisition

Participants: Éric Villemonte de La Clergerie, Rosa Stern, François-Régis Chaumartin, Benoît Sagot.

The first domain of application for Alpage parsing systems is information extraction, and in particular knowledge acquisition, be it linguistic or not, and text mining.

Knowledge acquisition for a given restricted domain is something that has already been studied by some Alpage members for several years (ACI Biotim, biographic information extraction from the Maitron corpus, SCRIBO project). Obviously, the progressive extension of Alpage parsing systems or even shallow processing chains to the semantic level increase the quality of the extracted information, as well as the scope of information that can be extracted. Such knowledge acquisition efforts bring solutions to current problems related to information access and take place into the emerging notion of *Semantic Web*. The transition from a web based on data (textual documents,...) to a web based on knowledge requires linguistic processing tools which are able to provide fine grained pieces of information, in particular by relying on high-quality deep parsing. For a given domain of knowledge (say, news or tourism), the extraction of a domain ontology that represents its key concepts and the relations between them is a crucial task, which has a lot in common with the extraction of linguistic information.

In the last years, such efforts have been targeted towards information extraction from news wires in collaboration with the Agence France-Presse (Rosa Stern is a CIFRE PhD student at Alpage and at AFP, and works in relation with the ANR project EDyLex) as well as in the context of the collaboration between Alpage and Proxem, a startup created by François-Régis Chaumartin, PhD student at Alpage (who has defended his PhD in 2012).

These applications in the domain of information extraction raise exciting challenges that require altogether ideas and tools coming from the domains of computational linguistics, machine learning and knowledge representation.

4.3. Processing answers to open-ended questions in surveys: vera

Participants: Benoît Sagot, Valérie Hanoka.

Verbatim Analysis is a startup co-created by Benoît Sagot from Alpage and Dimitri Tcherniak from Towers Watson, a world-wide leader in the domain of employee research (opinion mining among the employees of a company or organization). The aim of its first product, *vera*, is to provide an all-in-one environment for editing (i.e., normalizing the spelling and typography), understanding and classifying answers to open-ended questions, and relating them with closed-ended questions, so as to extract as much valuable information as possible from both types of questions. The editing part relies in part on SxPipe (see section 5.6) and Alexina morphological lexicons. Several other parts of *vera* are co-owned by Verbatim Analysis and by Inria.

4.4. Multilingual terminologies and lexical resources for companies

Participants: Éric Villemonte de La Clergerie, Mickael Morardo, Benoît Sagot.

Lingua et Machina is a small company now headed by François Brown de Colstoun, a former Inria researcher, that provides services for developing specialized multilingual terminologies for its clients. It develops the WEB framework Libellex for validating such terminologies. A formal collaboration with ALPAGE has been set up, with the recruitment of Mikael Morardo as engineer, funded by Inria's DTI. He works on the extension of the web platform *Libellex* for the visualization and validation of new types of lexical resources. In particular, he has integrated a new interface for handling monolingual terminologies, lexical networks, and bilingual wordnet-like structures.

4.5. Automatic and semi-automatic spelling correction in an industrial setting

Participants: Benoît Sagot, Éric Villemonte de La Clergerie, Laurence Danlos.

NLP tools and resources used for spelling correction, such as large n-gram collections, POS taggers and finite-state machinery are now mature and precise. In industrial setting such as post-processing after large-scale OCR, these tools and resources should enable spelling correction tools to work on a much larger scale and with a much better precision than what can be found in different contexts with different constraints (e.g., in text editors). Moreover, such industrial contexts allow for a non-costly manual intervention, in case one is able to identify the most uncertain corrections. An FUI project on this topic has been proposed in collaboration with Diadeis, a company specialized in text digitalization, and two other partners. It has been rerouted to the "Investissements d'avenir" framework, and has been accepted. It started in 2012.

4.6. Experimental linguistics

Participants: Benoît Crabbé, Juliette Thuilier, Luc Boruta.

Alpage is a team that dedicates efforts in producing resources and algorithms for processing large amounts of textual materials. These resources can be applied not only for purely NLP purposes but also for linguistic purposes. Indeed, the specific needs of NLP applications led to the development of electronic linguistic resources (in particular lexica, annotated corpora, and treebanks) that are sufficiently large for carrying statistical analysis on linguistic issues. In the last 10 years, pioneering work has started to use these new data sources to the study of English grammar, leading to important new results in such areas as the study of syntactic preferences [60], [128], the existence of graded grammaticality judgments [83].

The reasons for getting interested for statistical modelling of language can be traced back by looking at the recent history of grammatical works in linguistics. In the 1980s and 1990s, theoretical grammarians have been mostly concerned with improving the conceptual underpinnings of their respective subfields, in particular through the construction and refinement of formal models. In syntax, the relative consensus on a generative-transformational approach [71] gave way on the one hand to more abstract characterizations of the language faculty [71], and on the other hand to the construction of detailed, formally explicit, and often implemented, alternative formulation of the generative approach [59], [94]. For French several grammars have been implemented in this trend, among which the tree adjoining grammars of [63], [73] among others. This general movement led to much improved descriptions and understanding of the conceptual underpinnings of both linguistic competence and language use. It was in large part catalyzed by a convergence of interests of logical, linguistic and computational approaches to grammatical phenomena.

However, starting in the 1990s, a growing portion of the community started being frustrated by the paucity and unreliability of the empirical evidence underlying their research. In syntax, data was generally collected impressionistically, either as ad-hoc small samples of language use, or as ill-understood and little-controlled grammaticality judgements (Schütze 1995). This shift towards quantitative methods is also a shift towards new scientific questions and new scientific fields. Using richly annotated data and statistical modelling, we address questions that could not be addressed by previous methodology in linguistics. In this line, at Alpage we have started investigating the question of choice in French syntax with a statistical modelling methodology. Currently two studies are being led on the position of attributive adjectives w.r.t. the noun and the relative position of postverbal complement. This research has contributed to establish new links with the Laboratoire de Linguistique Formelle (LLF, Paris 7) and the Laboratoire de Psychologie et Neuropsychologie Cognitives (LPNCog, Paris 5).

On the other hand we have also started a collaboration with the Laboratoire de Sciences Cognitives de Paris (LSCP/ENS) where we explore the design of algorithms towards the statistical modelling of language acquisition (phonological acquisition). This is currently supported by one PhD project.

AVIZ Project-Team

4. Application Domains

4.1. Application Domains

AVIZ develops active collaboration with users from various application domains, making sure it can support their specific needs. By studying similar problems in different domains, we can begin to generalize our results and have confidence that our solutions will work for a variety of applications.

Our current application domains include:

- *Genealogy*, in cooperation with North Carolina State University;
- *Biological research*, in cooperation with Institut Pasteur;
- *Digital Libraries*, in cooperation with the French National Archives and the Wikipedia community;
- *Open Data*, in cooperation with Google Open Data and Data Publica;
- *Agrifood Process Modeling*, in cooperation with the DREAM project (see section [8.2.1.1](#));

AXIS Project-Team

3. Application Domains

3.1. Panorama: Living Labs, Smart Cities

AXIS addresses applicative field which has the following features:

a) requiring usage/data storage, preprocessing and analysis tools

- for designing, evaluating and improving huge evolving hypermedia information systems (mainly Web-based ISs), for which end-users are of primary concern,
- for a better understanding of service/product used with data mining techniques and knowledge management
- for social network analysis (for example in Web 2.0 applications, Business Intelligence, Sustainable Development, etc.)

b) requiring user-driven innovation methods.

Even if our know-how, methods and algorithms have a cross domain applicability, our team chooses to focus on **Living Lab projects** (and mainly related to **Sustainable Development for Smart Cities**) (cf. section 5.5.5 which imply user involvement for the generation of future services/products. Indeed, following the Rio Conference (1992) and the Agenda for the 21st Century, local territories are now directly concerned with the set up of actions for a sustainable development. In this frame, ICT tools are supposed to be very efficient to re-engage people in the democratic process and to make decision-making more transparent, inclusive and accessible. So, sustainable development is closely associated with citizen participation. The emerging research field of e-democracy (so called Digital Democracy or eParticipation), concerned with the use of communications technologies such as the Internet to enhance the democratic processes is now a very active field. Though still in its infancy, a lot of literature is already available (see for instance: <http://itc.napier.ac.uk/ITC/publications.asp> or <http://www.demo-net.org/> for a global view of work in Europe) and numerous different topics are addressed in the field.

Our experience particularly stressed on the following applicative domains:

- Transportation systems & Mobility (cf. section 3.2),
- Tourism (cf. section 3.3),
- User Involvement in Energy, Environment, Well Being & Health and e-governement (cf. section 3.4).

3.2. Transportation Systems & Mobility

Major recent evolutions in Intelligent Transportation Systems (ITS) are linked to rapid changes in communication technologies, such as ubiquitous computing, semantic web, contextual design. A strong emphasis is now put on mobility improvements. In addition to development of sustainable transportation systems (better ecological vehicles' performance, reduction of impacts on town planning ...) these improvements concern also mobility management, that is specific measures to encourage people to adopt new mobility behaviour such as public transportation services rather than their personal car. These prompting measures concern for instance the quality of traveller's information systems for trip planning, the ability to provide real time recommendations for changing transportation means according to traffic information, and the quality of embedded services in vehicles to provide enhanced navigation aids with contextualised and personalised information.

Since 2004, AxIS has been concerned with mobility projects :

- PREDIT (2004-2007): The MobiVIP project has been an opportunity to collaborate with local Institutions (Communauté d'Agglomération de Sophia Antipolis - CASA) and SMEs (VU Log) and to apply AxIS' know-how in data and web mining to the field of transportation systems.
- Traveller's information systems and recommender systems have been studied with the evaluation of two CASA web sites : the "Envibus" web site which provides information about a bus network and the "Otto&co" web site support car-sharing.
- Advanced transportation systems has been studied in PREDIT TIC TAC (2010-2012): this project (cf. section 6.1.1) aimed at optimizing travel time by providing in an area with weak transportation services, a just in time on demand shuttle based on real time information. It was for AxIS the opportunity to experiment user implication in the design of a new travel information system called MOBILTIC
- User Experience: in the ELLIOT project (cf. section 6.3.1.1), the mobility scenario is addressed in relation to information on air quality and noise and the use of internet of things.

3.3. Tourism

As tourism is a highly competitive domain, local tourism authorities have developed Web sites in order to offer of services to tourists. Unfortunately, the way information is organised does not necessarily meet Internet users expectations and numerous improvements are necessary to enhance their understanding of visited sites. Thus, even if only for economical reasons, the quality and the diversity of tourism packages have to be improved, for example by highlighting cultural heritage.

Again to illustrate our role in such a domain, let us cite some past projects where AxIS is involved related mainly to **Semantic Web Mining**³. In our case, a) we exploit ontologies and semantic data for improving usage analysis, personalised services, the quality of results of search engines and for checking the content of an IS and also b) we exploit usage data for updating ontologies.) and Information Retrieval.

- Research has been carried out using log files from the city of Metz. This city was chosen because its Web site is in constant development and has been awarded several times, notably in 2003, 2004 and 2005 in the context of the Internet City label. The objective was to extract information about tourists behaviours from this site log files and to identify possible benefits in designing or updating a tourism ontology.
- Providing Tourism Information linked to Transportation information: AxIS has already studied recommender systems in order to provide users with personalised transportation information while looking for tourism information such as cultural information, leisure etc (cf. our recommender Be-TRIP (2006) based on CBR*tools).
- In the context of HOTEL-REF-PACA project (cf. section 6.1.2 , we aimed to better refer the web sites of hotels/campings from the region of TOURVAL in PACA (mainly Vésudie territory), with an approach based on a better understanding of usage from the internauts. To address this, we proposed and adopted a multidisciplinary approach combining various AxIs know-how: knowledge engineering (ontology in tourism), data mining (analysis of Google logs, hotel web site logs and user queries, visual behaviours from eye tracker), Ergonomics (clustering of hotel web sites based on their ergonomical quality).
- Several contacts (PACA, France Living Labs, Island of the Reunion) have been done related to projects in tourism and eco-tourism.

³By Semantic Web Mining, we mean the mutual benefits between two communities Semantic Web and Web Mining

3.4. User Involvement in Energy, Environment, Health and E-gouvernement

Below are some topics where AxIS was or is involved in:

- **Preprocessing and analysing collective usage data and social networks** from group discussions related to design process: cf. ANR Intermed (2009) and FP7 Elliot (cf. section 6.3.1.1) where citizen generate ideas in terms of specific environmental sensors based services according to their needs.
- **Methods and tools for supporting open innovation based on public data:** a first work was made in 2010 with the CDISOD Color action related Public Data in collaboration with Fing (Marseille) and ADEME (Sophia Antipolis). We pursue such a study in the context of FP7 Elliot by providing to citizen environmental data (air quality and noise) issued from citizen and/or territories sensors.

All AxIS topics are relevant for these domains. let us cite: social network analysis, personalization and information retrieval, recommender systems, expert search, design and evaluation of methods and tools for open innovation and user co-creation in the context of Living Labs, usage mining, mining data streams.

We have addressed specific works:

- Energy (cf. section 6.1.3): the main AxIS topic here was usage analysis in the context of an energy challenge in an enterprise (ECOFFICES) taking into account the complex and real situation (installation fo more than 400 sensors, differences between the three concerned teams, differences between the offices). Such an analysis aims to correlate team/office energy consuming, team/office eco-responsible behaviours and team/office profile.
- Health (cf. section 3.4): Axis contributed in 2011 to a Living Lab characterisation in Health domain through the visit of several Living Labs, which operate in the domain of Health and Autonomy, and conducted interviews. This work was done in relation with the CGIET ⁴.
- E-gov: The future Internet will bring a growing number of networked applications (services), devices and individual data (including private ones) to end-users. The important challenges are the organization of their access, and the guarantee of trust and privacy. The objectives of the PIMI ⁵ project (cf. section 6.2.1) are the definition of a design environment and a deployment platform for Personal Information Management system (PIM). The future PIM must provide the end-user personal data access with services that are relevant to his needs. In order to take mobility into account, the PIM will be accessed both by mobile devices (smartphone) and Personal Computers. With the increasing number of services and associated data being accessible through Internet, the number and complexity of PIM will augment dramatically in the near future. This will require strong research investment in a number of topics, all contributing to the expected usability and accessibility of Individual Information Spaces for the end-user.

⁴CGIET: "Conseil Général de l'Economie, de l'Industrie, de l'Energie et des technologies" <http://www.cgeiet.economie.gouv.fr>

⁵Personal Information Management through Internet

AYIN Team

4. Application Domains

4.1. Remote sensing

The first application domain of the AYIN team concerns analysis and classification of remote sensing images. The very high spatial, spectral and temporal resolution of the last generation of imaging sensors (for instance, GeoEye, Ikonos, Pleiades, COSMO-SkyMed, TerraSAR-X, ...) provides rich information about environment and is very useful in a range of applications, such as investigating urban environments, precision agriculture, natural disasters and mineralogy. The development of these applications presents new challenges of high-dimensional and high-volume data analysis. The methods proposed by the AYIN team are applied for analysis of SAR, multi- and hyperspectral remote sensing images. In particular, the team develops approaches for image segmentation and classification, change detection, extraction of structures and object tracking.

4.2. Skin care

The second application domain of the AYIN team is skin care imaging which mainly consists in image analysis and classification for dermatology and cosmetology. Here we also deal with very high spatial, spectral and temporal resolution of the most recent imaging sensors. In dermatology we are particularly interested in hyperpigmentation detection and disorders severity evaluation (for instance, for melasma, acne, ...). In cosmetology our main goals are analysis, modeling and characterization of the condition of human skin, as well as evaluation of means to influence that condition. Some of the changes in skin over time have to do with chronological aging (such as pinheads for teenagers or wrinkles for mature people), others with extrinsic aging, caused for instance by sun exposure and smoking.

COPRIN Project-Team

4. Application Domains

4.1. Application Domains

While the methods developed in the project can be used for a very broad set of application domains (for example we have an activity in CO₂ emission allowances [32]), it is clear that the size of the project does not allow us to address all of them. Hence we have decided to focus our applicative activities on *mechanism theory*, where we focus on *optimal design* and geometrical modeling of mechanisms. Along the same line our focus is *robotics* and especially *service robotics* which includes rescue robotics, rehabilitation and assistive robots for elderly and handicapped people (section 6.2.1.3). Although these topics were new for us in 2008 we have spent two years determining priorities and guidelines by conducting about 200 interviews with field experts (end-users, praticians, family and caregivers, institutes), establishing strong collaboration with them (e.g. with the CHU of Nice-Cimiez) and putting together an appropriate experimental setup for testing our solutions (see 6.2.1.4). A direct consequence of setting up this research framework is a reduction in our publication and contract activities. But this may be considered as an investment as assistance robotics will constitute the major research axis of the project on the long term (COPRIN will reach its twelve years of existence in 2014 but we intend to propose in 2013 a new project on this topic, code-named HEPHAISTOS).

DAHU Project-Team

4. Application Domains

4.1. Application Domains

Databases are pervasive across many application fields. Indeed, most human activities today require some form of data management. In particular, all applications involving the processing of large amounts of data require the use of a database. Increasingly complex Web applications and services also rely on DBMS, and their correctness and robustness is crucial.

We believe that the automated solutions that Dahu aims to develop for verifying such systems will be useful in this context.

DREAM Project-Team

4. Application Domains

4.1. Introduction

The DREAM research applications have been oriented towards surveillance of large networks as telecommunication networks and more recently of web services. During the past few years, we have focussed more and more on agricultural and environmental applications by means of research collaborations with INRA and Agrocampus Ouest.

4.2. Software components monitoring

software components, web services, distributed diagnosis

Web-services, i.e., services that are provided, controlled and managed through Internet, cover nowadays more and more application areas, from travel booking to goods supplying in supermarkets or the management of an e-learning platform. Such applications need to process requests from users and other services on line, and respond accurately in real time. Anyway, errors may occur, which need to be addressed in order to still be able to provide the correct response with a satisfactory quality of service (QoS): on-line monitoring, especially diagnosis and repair capabilities, become then a crucial concern.

We have been working on this problem within the WS-DIAMOND project [68], a large European funded project involving eight partners in Italy, France, Austria and Netherlands <http://wsdiamond.di.unito.it/>. Our own work consisted in two distinct contributions.

The first issue has been to extend the decentralized component-oriented approach, initially developed for monitoring telecommunication networks [4] to this new domain. To this end we have proposed the concept of distributed chronicles, with synchronization events, and the design of an architecture consisting of distributed CRSs (Chronicle Recognition Systems) communicating their local diagnoses to a broker agent which is in charge of merging them to compute a global diagnosis.

Our current work aims at coupling diagnosing and repair, in order to implement *adaptive web services*. We started this study by proposing an architecture inspired from the one developed during the WS-DIAMOND project and dedicated to the adaptive process of a request event when faults occur and propagate through the orchestration.

4.3. Environmental decision making

environment, decision methods

The need of decision support systems in the environmental domain is now well-recognized. It is especially true in the domain of water quality. For instance the program, named “Bretagne Eau Pure”. was launched a few years ago in order to help regional managers to protect this important resource in Brittany. The challenge is to preserve the water quality from pollutants as nitrates and herbicides, when these pollutants are massively used by farmers to weed their agricultural plots and improve the quality and increase the quantity of their crops. The difficulty is then to find solutions which satisfy contradictory interests and to get a better knowledge on pollutant transfer.

In this context, we are cooperating with INRA (Institut National de Recherche Agronomique) and developing decision support systems to help regional managers in preserving the river water quality. The approach we advocate is to rely on a qualitative modeling, in order to model biophysical processes in an explicative and understandable way. The SACADEAU model associates a qualitative biophysical model, able to simulate the biophysical process, and a management model, able to simulate the farmer decisions. One of our main contribution is the use of qualitative spatial modeling, based on runoff trees, to simulate the pollutant transfer through agricultural catchments.

The second issue is the use of learning/data mining techniques to discover, from model simulation results, the discriminant variables and automatically acquire rules relating these variables. One of the main challenges is that we are faced with spatiotemporal data. The learned rules are then analyzed in order to recommend actions to improve a current situation.

This work has been done in the framework of the APPEAU project, funded by ANR and of the ACASSYA project, funded by ANR, having started at the beginning of 2009 and ended at the end of 2012. We were also involved in the PSDR GO CLIMASTER project, that started in september 2008 and end in 2011. CLIMASTER stands for “Changement climatique, systèmes agricoles, ressources naturelles et développement territorial” and is dedicated to the impact of climate changes on the agronomical behaviors in west of France (“Grand Ouest”). PSDR GO stands for “Programme Pour et Sur le Développement Régional Grand Ouest”.

Our main partners are the SAS INRA research group, located in Rennes and the BIA INRA and AGIR INRA research groups in Toulouse.

E-MOTION Project-Team

3. Application Domains

3.1. Introduction

The main applications of our research are those aiming at introducing advanced and secured robotized systems into human environments. In this context, we are focusing onto the following application domains: Future cars and transportation systems, Service and Human assistance robotics, and Potential spin-offs in some other application domains.

3.2. Future cars and transportation systems

Thanks to the introduction of new sensor and ICT technologies in cars and in mass transportation systems, and also to the pressure of economical and security requirements of our modern society, this application domain is quickly changing. Various technologies are currently developed by both research and industrial laboratories. These technologies are progressively arriving at maturity, as it is witnessed by the results of large scale experiments and challenges (e.g., Darpa Urban Challenge 2007) and by the fast development of ambitious projects such as the Google's car project. Moreover, the legal issue starts to be addressed (see for instance the recent laws in Nevada and in California authorizing autonomous vehicles on roads).

In this context, we are interested in the development of *ADAS*¹ systems aimed at improving comfort and safety of the cars users (e.g., ACC, emergency braking, danger warnings), and of *Fully Autonomous Driving* functions for controlling the displacements of private or public vehicles in some particular driving situations and/or in some equipped areas (e.g., automated car parks or captive fleets in downtown centers or private sites).

3.3. Service, intervention, and human assistance robotics

This application domain is currently quickly emerging, and more and more industrial companies (e.g., IS-Robotics, Samsung, LG) are now commercializing service and intervention robotics products such as vacuum cleaner robots, drones for civil or military applications, entertainment robots . . .). One of the main challenges is to propose robots which are sufficiently robust and autonomous, easily usable by non-specialists, and marked at a reasonable cost. A more recent challenge for the coming decade is to develop robotized systems for assisting elderly and/or disabled people. We are strongly involved in the development of such technologies, which are clearly tightly connected to our research work on robots in human environments.

3.4. Potential spin-offs in some other application domains

Our *Bayesian Programming* tools (including the functions for decision making under uncertainty) are also impacting a large spectrum of application domains such as autonomous systems, surveillance systems, preventive maintenance for large industrial plants, fraud detection, video games, etc. These application domains are covered by our start-up *Probayes*.

¹Advanced Driver Assistance Systems

EXMO Project-Team

4. Application Domains

4.1. Application Domains

The main application context motivating our work is the “semantic web” infrastructure (§4.1.1), but it can be applied in any context where semantic technologies are used: semantic social networks, ambient intelligence, linked data, etc. [2].

4.1.1. Semantic web technologies

Internet technologies support organisations and people in accessing and sharing knowledge, often difficult to access in a documentary form. However, these technologies quickly reach their limits: web site organisation is expensive and full-text search inefficient. Content-based information search is becoming a necessity. Content representation will enable computers to manipulate knowledge on a more formal ground and to carry out similarity or generality search. Knowledge representation formalisms are good candidates for expressing content.

The vision of a “semantic web” [29] supplies the web, as we know it (informal) with annotations expressed in a machine-processible form and linked together. In the context where web documents are formally annotated, it becomes necessary to import and manipulate annotations according to their semantics and their use. Taking advantage of this semantic web will require the manipulation of various knowledge representation formats. Exmo concerns are thus central to the semantic web implementation. Our work aims at enhancing content understanding, including the intelligibility of communicated knowledge and formal knowledge transformations.

In addition, Exmo also considers a more specific use of semantic web technologies in semantic peer-to-peer systems, social semantic networks and ambient intelligence (see §6.3). In short, we would like to bring the semantic web to everyone’s pocket. Semantic peer-to-peer systems are made of a distributed network of independent peers which share local resources annotated semantically and locally. This means that each peer can use its own ontology for annotating resources and these ontologies have to be confronted before peers can communicate. In social semantic networks, relationships between people are inferred from relationships between knowledge they use. In ambient intelligence, applications have to reconcile device and sensor descriptions provided by independent sources.

Exmo’s work can be implemented in software: in particular, we have developed an API for expressing ontology alignment (§5.1) and a library of ontology distances and similarities OntoSim (§5.2).

FLOWERS Project-Team

4. Application Domains

4.1. Application Domains

- **Personal robotics.** Many indicators show that the arrival of personal robots in homes and everyday life will be a major fact of the 21st century. These robots will range from purely entertainment or educative applications to social companions that many argue will be of crucial help in our aging society. For example, UNECE evaluates that the industry of entertainment, personal and service robotics will grow from \$5.4Bn to \$17.1Bn over 2008-2010. Yet, to realize this vision, important obstacles need to be overcome: these robots will have to evolve in unpredictable homes and learn new skills while interacting with non-engineer humans after they left factories, which is out of reach of current technology. In this context, the refoundation of intelligent systems that developmental robotics is exploring opens potentially novel horizons to solve these problems.
- **Human-Robot Collaboration.** Robots play a vital role for industry and ensure the efficient and competitive production of a wide range of goods. They replace humans in many tasks which otherwise would be too difficult, too dangerous, or too expensive to perform. However, the new needs and desires of the society call for manufacturing system centered around personalized products and small series productions. Human-robot collaboration could widen the use of robot in this new situations if robots become cheaper, easier to program and safe to interact with. The most relevant systems for such applications would follow an expert worker and works with (some) autonomy, but being always under supervision of the human and acts based on its task models.
- **Video games.** In conjunction with entertainment robotics, a new kind of video games are developing in which the player must either take care of a digital creature (e.g. Neopets), or tame it (e.g. Nintendogs), or raise/accompany them (e.g. Sims). The challenges entailed by programming these creatures share many features with programming personal/entertainment robots. Hence, the video game industry is also a natural field of application for FLOWERS.
- **Environment perception in intelligent vehicles.** When working in simulated traffic environments, elements of FLOWERS research can be applied to the autonomous acquisition of increasingly abstract representations of both traffic objects and traffic scenes. In particular, the object classes of vehicles and pedestrians are of interest when considering detection tasks in safety systems, as well as scene categories ("scene context") that have a strong impact on the occurrence of these object classes. As already indicated by several investigations in the field, results from present-day simulation technology can be transferred to the real world with little impact on performance. Therefore, applications of FLOWERS research that is suitably verified by real-world benchmarks has direct applicability in safety-system products for intelligent vehicles.
- **Automated Tutoring Systems.** Optimal teaching and efficient teaching/learning environments can be applied to aid teaching in schools aiming both at increase the achievement levels and the reduce time needed. From a practical perspective, improved models could be saving millions of hours of students' time (and effort) in learning. These models should also predict the achievement levels of students in order to influence teaching practices.

GRAPHIK Project-Team

4. Application Domains

4.1. Introduction

We currently focus on two application domains: knowledge representation in agronomy, more precisely applied to the quality in agri-food chains, and metadata management, in particular for bibliographic metadata.

The choice of the agronomy domain is motivated both by the local context of GraphIK (UMR IATE) and by its adequation to our research themes. Indeed, the agri-food domain seems to be particularly well-adapted to artificial intelligence techniques: there are no mathematical models available to solve the problems related to the quality of agrifood chains, which need to be stated at a more conceptual level; solving these problems requires an integrated approach that takes into account expert knowledge, which is typically symbolic, as well as numeric data, vague or uncertain information, multi-granularity knowledge, multiple and potentially conflicting viewpoints and actors.

The second area, metadata management, is not strictly speaking an application domain, but rather a cross-cutting axis. Indeed, metadata can be used to describe data in various areas (including for instance scientific publications in agronomy). We have a long experience in this domain, and we currently focus on document metadata.

4.2. Agronomy

Quality control within agri-food chains, but also non-food chains relies on numerous criteria (environmental, economical, functional, sanitary quality, etc.). The objectives of quality are based on several actors. The current structure of chains is questioned as for system perennity, protection of the environment, cost and energy. In all cases, the following questions have to be taken into account:

1. the actors' viewpoints are divergent, hence it is necessary to define reasoning mechanisms able to model and take into account the balance between viewpoints, and the risks and benefits they imply;
2. the successive steps involved in a chain, impacting the quality of end products, have limiting factors. Their improvement is a complex objective that has no simple solution;
3. data from literature are dispersed and scattered, which makes their use difficult.

These questions highlight the need for an integrated approach of agri-food chains, respectively with symbolic reasoning mechanisms, reverse engineering methods, and knowledge organization and modelling.

Our general objective is the conception of a decision support tool for the actors of an agri-food chain, in presence of contradictory viewpoints and priorities, including the concepts of gravity and certainty of a risk or a benefit. The first step is to build a knowledge-based system able to represent the different kinds of knowledge needed, and provided with consistency checking, querying and symbolic simulation mechanisms, which will allow to refine and validate the modelling.

Our results in Sect. 6.1 and Sect. 6.2 can be seen as theoretical requirements towards this objective.

4.3. Document Metadata

Semantic metadata, in particular semantic annotations for multimedia documents, are at the core of the applications we are working on for several years. In the applications we developed in the previous years, mainly with INA (National Institute of Audiovisual) and FMSH (Fondation Maison des Sciences de l'Homme), we have built tools aimed at helping the manual construction of semantic annotations. In these projects, manual construction was unavoidable because semantically rich annotations, not obtainable by automatic processes, had to be built. In our current project Qualinca (see Sect. 8.1), the semantic metadata considered consists of information present in bibliographic databases and authority notices (which respectively describe documents and so-called authorities, such as authors typically). The challenge is not to build these metadata, which have been built by human specialists and already exist, but, for instance, to check their validity, to link or to merge different metadata bases.

IMAGINE Team

4. Application Domains

4.1. Application Domains

This research can be applied to any situation where users need to create new, imaginary, 3D content. Our work should be instrumental, in the long term, for the visual arts, from the creation of 3D films and games to the development of new digital planning tools for theatre or cinema directors. Our models can also be used in interactive prototyping environments for engineering. They can help promoting interactive digital design to scientists, as a tool to quickly express, test and refine models, as well as an efficient way for conveying them to other people. Lastly, we expect our new methodology to put digital modelling within the reach of the general public, enabling educators, media and other practitioners to author their own 3D content.

In practice, fully developing a few specialized interactive systems will be instrumental for testing our models. The multi-disciplinary expertise and professional background of our team members will ease the set up of projects in the domains listed below. The diversity of users these domains bring, from digital experts to other professionals and novices, will be excellent for validating our general methodology. Our ongoing projects in these various application domains are listed in Section 6.

- Visual arts
 - Modeling and animation for 3D films and games (François Faure, Marie-Paule Cani,)
 - Virtual cinematography and tools for theatre directors (Rémi Ronfard)
- Engineering
 - Industrial design (Stéphanie Hahmann, Jean-Claude Léon)
 - Mechanical & civil engineering (Jean-Claude Léon, François Faure)
- Natural Sciences
 - Virtual functional anatomy (Olivier Palombi, François Faure)
 - Virtual plants (Marie-Paule Cani, François Faure)
- Education and Creative tools
 - Sketch-based teaching (Olivier Palombi, Marie-Paule Cani)
 - Creative environments for novice users (Marie-Paule Cani, Jean-Claude Léon)

IMARA Project-Team

4. Application Domains

4.1. Introduction

While the preceding section focused on methodology, in connection with automated guided vehicles, it should be stressed that the evolution of the problems which we deal with remains often guided by the technological developments. We enumerate three fields of application, whose relative importance varies with time and who have strong mutual dependencies: driving assistance, cars available in self-service mode and fully automated vehicles (cybercars).

4.2. Driving assistance

Several techniques will soon help drivers. One of the first immediate goal is to improve security by alerting the driver when some potentially dangerous or dangerous situations arise, i.e. collision warning systems or lane tracking could help a bus driver and surrounding vehicle drivers to more efficiently operate their vehicles. Human factors issues could be addressed to control the driver workload based on additional information processing requirements.

Another issue is to optimize individual journeys. This means developing software for calculating optimal (for the user or for the community) path. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component in account.

4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs.

However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these “lines of desire” or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time.

To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost.

Electric or gas vehicles available in self-service as in the Praxitèle system bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

- ease empty car moves to better distribute them;
- better use of information systems inboard and on ground;
- better integrate this system in the global transportation system.

These systems are now operating (i.e. in La Rochelle). The challenge is to bring them to an industrial phase by transferring technologies to these still experimental projects.

4.4. Cybercars

The long term effort of the project is to put automatically guided vehicles (cybercars) on the road. It seems too early to mix cybercars and traditional vehicles, but data processing and automation now make it possible to consider in the relatively short term the development of such vehicles and the adapted infrastructures. IMARA aims at using these technologies on experimental platforms (vehicles and infrastructures) to accelerate the technology transfer and to innovate in this field.

Other application can be precision docking systems that will allow buses to be automatically maneuvered into a loading zone or maintenance area, allowing easier access for passengers, or more efficient maintenance operations. Transit operating costs will also be reduced through decreased maintenance costs and less damage to the breaking and steering systems.

Regarding technical topics, several aspects of Cybercars have been developed at IMARA this year. First, we have stabilized a generic Cycab architecture involving Inria Syndex tool and CAN communications. The critical part of the vehicle is using a real time Syndex application controlling the actuators via two Motorola's MPC555. Today, we have decided to migrate to the new dsPIC architecture for more efficiency and ease of use.

This application has a second feature, it can receive commands from an external source (Asynchronously this time) on a second CAN bus. This external source can be a PC or a dedicated CPU, we call it high level. To work on the high level, in the past years we have been developing a R&D framework called (Taxi) which used to take control of the vehicle (Cycab and Yamaha) and process data such as gyro, GPS, cameras, wireless communications and so on. Today, in order to rely on a professional and maintained solution, we have chosen to migrate to the RTMAPS SDK development platform. Today, all our developments and demonstrations are using this efficient prototyping platform. Thanks to RT-MAPS we've been able to do all the demonstrations on our cybercars: cycabs, Yamaha AGV and new Cybus platforms. These demonstrations include: reliable SLAMMOT algorithm using 2 to 4 laser sensors simultaneously, automatic line/road following techniques, PDA remote control, multi sensors data fusion, collaborative perception via ad-hoc network.

The second main topic is inter-vehicle communications using ad-hoc networks. We have worked with the HIPERCOM team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications (see Section 3.2). Our goal is to develop a vehicle dedicated communication software suite, running on a specialized hardware. It can be linked also with the Taxi Framework for getting data such GPS information's to help the routing algorithm.

IMEDIA2 Team

4. Application Domains

4.1. Scientific applications

Examples: environmental images databases: fauna and flora; satellite images databases: ground typology; medical images databases: find images of a pathological character for educational or investigation purposes.

We are developing tools enabling multimedia access to biodiversity collections for species identifications inside the PI@ntNet project.

In fact, almost all IMEDIA2 team members are involved in the PI@ntNet project and develop methods that can be used in this context.

4.2. Audio-visual applications

Examples: Look for a specific shot in a movie, documentary or TV news, present a video summary. Help archivists to annotate the contents. Retrieve copies of a given material (photo or video) in a TV stream or on the web.

Our team has a collaboration with AFP press agency in the context of GLOCAL project.

IN-SITU Project-Team

4. Application Domains

4.1. Research Domains

INSITU works on general problems of interaction in multi-surface environments as well as on challenges associated with specific research groups. The former requires a combination of controlled experiments and field studies; the latter involves participatory design with users. We are currently working with highly creative people, particularly designers and music composers, to explore interaction techniques and technologies that support the earliest phases of the design process. We are also working with research scientists, particularly neuroscientists and astrophysicists, in our explorations of interaction in multisurface environments, and with doctors and nurses to support crisis management situations.

LAGADIC Project-Team

4. Application Domains

4.1. Overview

The natural applications of our research are obviously in robotics. In fact, researches undertaken in the Lagadic group can apply to all the fields of robotics implying a vision sensor. They are indeed conceived to be independent of the system considered (and the robot and the vision sensor can even be virtual for some applications).

Currently, we are mostly interested in using visual servoing for aerial and space application, micromanipulation, autonomous vehicle navigation in large urban environments or for disabled or elderly people.

We also address the field of medical robotics. The applications we consider turn around new functionalities of assistance to the clinician during a medical examination: visual servoing on echographic images, needle insertion, compensation of organ motions, etc.

Robotics is not the only possible application field to our researches. In the past, we were interested in applying visual servoing in computer animation, either for controlling the motions of virtual humanoids according to their pseudo-perception, or for controlling the point of view of visual restitution of an animation. In both cases, potential applications are in the field of virtual reality, for example for the design of video games, or virtual cinematography.

Applications also exist in computer vision and augmented reality. It is then a question of carrying out a virtual visual servoing for the 3D localization of a tool with respect to the vision sensor, or for the estimation of its 3D motion. This field of application is very promising, because it is in full rise for the realization of special effects in the multi-media field or for the design and the inspection of objects manufactured in the industrial world.

LEAR Project-Team

4. Application Domains

4.1. Application Domains

A solution to the general problem of visual recognition and scene understanding will enable a wide variety of applications in areas including human-computer interaction, retrieval and data mining, medical and scientific image analysis, manufacturing, transportation, personal and industrial robotics, and surveillance and security. With the ever expanding array of image and video sources, visual recognition technology is likely to become an integral part of many information systems. A complete solution to the recognition problem is unlikely in the near future, but partial solutions in these areas enable many applications. LEAR's research focuses on developing basic methods and general purpose solutions rather than on a specific application area. Nevertheless, we have applied our methods in several different contexts.

Semantic-level image and video access. This is an area with considerable potential for future expansion owing to the huge amount of visual data that is archived. Besides the many commercial image and video archives, it has been estimated that as much as 96% of the new data generated by humanity is in the form of personal videos and images ¹, and there are also applications centering on on-line treatment of images from camera equipped mobile devices (e.g. navigation aids, recognizing and answering queries about a product seen in a store). Technologies such as MPEG-7 provide a framework for this, but they will not become generally useful until the required mark-up can be supplied automatically. The base technology that needs to be developed is efficient, reliable recognition and hyperlinking of semantic-level domain categories (people, particular individuals, scene type, generic classes such as vehicles or types of animals, actions such as football goals, etc). In a collaboration with Xerox Research Center Europe, supported by a CIFRE grant from ANRT, we study cross-modal retrieval of images given text queries, and vice-versa. In the context of the Microsoft-Inria collaboration we concentrate on retrieval and auto-annotation of videos by combining textual information (scripts accompanying videos) with video descriptors. In the EU FP7 project AXES we will further mature such video annotation techniques, and apply them to large archives in collaboration with partners such as the BBC, Deutsche Welle, and the Netherlands Institute for Sound and Vision.

Visual (example based) search. The essential requirement here is robust correspondence between observed images and reference ones, despite large differences in viewpoint or malicious attacks of the images. The reference database is typically large, requiring efficient indexing of visual appearance. Visual search is a key component of many applications. One application is navigation through image and video datasets, which is essential due to the growing number of digital capture devices used by industry and individuals. Another application that currently receives significant attention is copyright protection. Indeed, many images and videos covered by copyright are illegally copied on the Internet, in particular on peer-to-peer networks or on the so-called user-generated content sites such as Flickr, YouTube or DailyMotion. Another type of application is the detection of specific content from images and videos, which can, for example, be used for finding product related information given an image of the product. Transfer of such techniques is the goal of the start-up MilPix, to which our current technologies for image search are licensed. In a collaboration with Technosens we transfer face recognition technology, which they exploit to identify users of a system and adapt the interface to the user.

Automated object detection. Many applications require the reliable detection and localization of one or a few object classes. Examples are pedestrian detection for automatic vehicle control, airplane detection for military applications and car detection for traffic control. Object detection has often to be performed in less common imaging modalities such as infrared and under significant processing constraints. The main challenges are the relatively poor image resolution, the small size of the object regions and the changeable appearance of the objects. Our industrial project with MBDA is on detecting objects under such conditions in infrared images.

¹<http://www.sims.berkeley.edu/research/projects/how-much-info/summary.html>

MAGRIT Project-Team

4. Application Domains

4.1. Augmented reality

We have a significant experience in the AR field especially through the European project ARIS (2001–2004) which aimed at developing effective and realistic AR systems for e-commerce and especially for interior design. Beyond this restrictive application field, this project allowed us to develop nearly real time camera tracking methods for multi-planar environments. Since then, we have amplified our research on multi-planar environments in order to obtain effective and robust AR systems in such environments. We currently investigate both automatic and interactive techniques for scene reconstruction/structure from motion methods in order to be able to consider large and unknown environments.

4.2. Medical Imaging

For 15 years, we have been working in close collaboration with University Hospital of Nancy and GE Healthcare in interventional neuroradiology. Our common aim is to develop a multimodality framework to help therapeutic decisions and interventional gestures. In particular, we aim at developing tools allowing the physicians to take advantage of the various existing imaging modalities on the brain in their clinical practice: 2D subtracted angiography (2DSA), 3D rotational angiography (3DRA), fluoroscopy, MRI,...Recent works concern the use of AR tools for neuronavigation and the development of simulation tools of the interventional act for training or planning. This last project is developed in collaboration with the EPI Shacra.

4.3. Augmented head

Visual information on a speaker, especially jaws and lips but also tongue position, noticeably improves speech intelligibility. Hence, having a realistic augmented head displaying both external and internal articulators could help language learning technology progress in giving the student a feedback on how to change articulation in order to achieve a correct pronunciation. The long term aim of the project is the acquisition of articulatory data and the design of a 3D-t articulatory model from various image modalities: external articulators are extracted from stereovision data, the tongue shape is acquired through ultrasound imaging, 3D images of all articulators can be obtained with MRI for sustained sounds, magnetic sensors are used to recover the tip of the tongue.

MAIA Project-Team

4. Application Domains

4.1. Decision Making

Our group is involved in several applications of its more fundamental work on autonomous decision making and complex systems. Applications addressed include:

- Robotics, where the decision maker or agent is supported by a physical entity moving in the real world;
- Medicine or Personal Assisting Living, where the agent can be an analytic device recommending tests and/or treatments, or able to gather different sources of information (sensors for example) in order to help a final user, detecting for example anormal situation needing the rescue of a person (fall detection of elderly people, risk of hospitalization of a person suffering from chronic disease;
- Computer Security, where the agent can be a virtual attacker trying to identify security holes in a given network;
- and Business Process Management, where the agent can provide an auto-completion facility helping to decide which steps to include into a new or revised process.

4.2. Ambient intelligence

Taking into consideration some scientific strategic choices made by the Nancy – Grand Est Research Center such as developing some platforms around Robotics and Smart Living Apartments, some members of the team have recentered their research toward “ambient intelligence and AI” . This choice has been also comforted by the launch by Inria of a Large-scale initiative project termed PAL (Personal assistant Living) in which we are strongly involved. The regional council of Lorraine also supports this new research line through the CPER, (project "situated computing" or "INFOSITU" (<http://infositu.loria.fr>) whose the coordinator is a member of MAIA Team. Within this new domain of research in MAIA, we are thinking about the design of intelligent environments dedicated to elderly people with loss of autonomy. This domain of research is currently very active, taking up a societal challenge that developed countries have to address. In order to contribute to this research domain, we are currently developing several action lines both at the technological, scientific and service levels:

- Evaluation of the degree of frailty of the elderly : We are currently designing a system whose the objective is twofold : assessing the risk of fall and evaluating the degree of frailty of the elderly. This issue is considered with more or less sophisticated sensors : one kinect, a network of kinects, an heterogeneous sensor network made up of an intelligent floor and a network of kinects. One simple idea which is currently developed is to determine either the center of mass of a person using one or several kinect, or the center of pressure and footsteps localization using an intelligent floor. The idea is to induce from these simple measures, the walking speed, the length of the steps and the position of the monitored persons.
- People activity analysis: Sensitive or intelligent floors have attracted a lot of attention during the last two decades for different applications going from interaction capture in immersive virtual environments to robotics or human tracking, fall detection or activity recognition. Different technologies have been proposed so far either based on optical fiber sensing, pressure sensing or electrical near field. In PAL we envision a more more sophisticate approach in which both computation and sensing is distributed within the floor. This floor is made up of interconnected intelligent tiles with can:
 - communicate with each other,
 - provide some computation,
 - sense the environment activity through four weight sensors, an accelerometer and a magnetometer,
 - interact with users, robots or other sensor networks either by wireless/wire communication or through visual communication (each tiles being equipped be 16 leds).

Several scientific challenges are open to us

- in decentralized spatial computing
- in designing real application for assisting people suffering from loss of autonomy

Concerning the second point, we envision several applications :

- evaluation of the degree of frailty of the elderly, especially evaluating the risk of fall
- activity recognition
- monitoring assistant robots

MANAO Team (section vide)

MAVERICK Team

4. Application Domains

4.1. Introduction

Maverick is part of the research theme “Interaction and Visualization” at Inria. This research theme has historically been very successful inside Inria. It nicely connects industrial applications with fundamental research using advanced mathematics, algorithmic and computer science, and it connects computer science with other sciences such as physics, biology, medicine, environment, psychophysiology.

We envision Maverick at this crossroad. We have several industrial partnerships, with companies making video games (Eden Games), special effects for motion pictures (WetaFX), planetarium (RSA Cosmos), graphical edition software (Adobe), tomography (Digisens) or visualizing simulated data (EDF). The constraints and needs of our partners motivate new problems for us to solve. At the same time, we are looking into fundamental research problems, such as analysis of light transport, human perception, filtering and sampling.

The fundamental research problems we target are not necessarily “long term research”: the computer graphics industry is very dynamic and can adopt (and adapt) a research paper in a matter of months if it sees benefits in it. The research problems we describe as “fundamental” correspond to high-risk, high-benefit research problems. Solving these problems would result in a significant breakthrough for the whole domain of Computer Graphics, both in research and in industry.

4.2. Illustration

Although it has long been recognized that the visual channel is one of the most effective means for communicating information, the use of computer processing to generate effective visual content has been mostly limited to very specific image types: realistic rendering, computer-aided cell animation, etc.

The ever-increasing complexity of available 3d models is creating a demand for improved image creation techniques for general illustration purposes. Recent examples in the literature include computer systems to generate road maps, or assembly instructions, where a simplified visual representation is a necessity.

Our work in expressive rendering and in relevance-guided rendering aims at providing effective tools for all illustration needs that work from complex 3d models. We also plan to apply our knowledge of lighting simulation, together with expressive rendering techniques, to the difficult problem of sketching illustrations for architectural applications.

4.3. Video-games and visualization

Video games represent a particularly challenging domain of application since they require both real-time interaction and high levels of visual quality. Moreover, video games are developed on a variety of platforms with completely different capacities. Automatic generation of appropriate data structures and runtime selection of optimal rendering algorithms can save companies a huge amount of development.

More generally, interactive visualization of complex data (e.g. in scientific engineering) can be achieved only by combining various rendering accelerations (e.g. visibility culling, levels of details, etc.), an optimization task that is hard to perform “by hand” and highly data dependent. One of Maverick’ goals is to understand this dependence and automate the optimization.

4.4. Virtual heritage

Virtual heritage is a recent area which has seen spectacular growth over the past few years. Archeology and heritage exhibits are natural application areas for virtual environments and computer graphics, since they provide the ability to navigate 3D models of environments that no longer exist and can not be recorded on a videotape. Moreover, digital models and 3D renderings give the ability to enrich the navigation with annotations.

Our work on style has proved very interesting to architects who have a long habit of using hand-drawn schemas and wooden models to work and communicate. Wooden models can advantageously be replaced by 3D models inside a computer. Drawing, on the other hand, offers a higher level of interpretation and a richness of expression that are really needed by architects, for example to emphasize that such model is an hypothesis.

By investigating style analysis and expressive rendering, we could “sample” drawing styles used by architects and “apply” them to the rendering of 3D models. The computational power made available by computer assisted drawing can also lead to the development of new styles with a desired expressiveness, which would be harder to produce by hand. In particular, this approach offers the ability to navigate a 3D model while offering an expressive rendering style, raising fundamental questions on how to “animate” a style.

METISS Project-Team

4. Application Domains

4.1. Introduction

This section reviews a number of applicative tasks in which the METISS project-team is particularly active :

- spoken content processing
- description of audio streams
- audio scene analysis
- advanced processing for music information retrieval

The main applicative fields targeted by these tasks are :

- multimedia indexing
- audio and audio-visual content repurposing
- description and exploitation of musical databases
- ambient intelligence
- education and leisure

4.2. Spoken content processing

speaker recognition, user authentication, voice signature, speaker adaptation, spoken document, speech modeling, speech recognition, rich transcription, beam-search, broadcast news indexing, audio-based multimodal structuring

A number of audio signals contain speech, which conveys important information concerning the document origin, content and semantics. The field of speaker characterisation and verification covers a variety of tasks that consist in using a speech signal to determine some information concerning the identity of the speaker who uttered it.

In parallel, METISS maintains some know-how and develops new research in the area of acoustic modeling of speech signals and automatic speech transcription, mainly in the framework of the semantic analysis of audio and multimedia documents.

4.2.1. Robustness issues in speaker recognition

Speaker recognition and verification has made significant progress with the systematical use of probabilistic models, in particular Hidden Markov Models (for text-dependent applications) and Gaussian Mixture Models (for text-independent applications). As presented in the fundamentals of this report, the current state-of-the-art approaches rely on bayesian decision theory.

However, robustness issues are still pending : when speaker characteristics are learned on small quantities of data, the trained model has very poor performance, because it lacks generalisation capabilities. This problem can partly be overcome by adaptation techniques (following the MAP viewpoint), using either a speaker-independent model as general knowledge, or some structural information, for instance a dependency model between local distributions.

METISS also addresses a number of topics related to speaker characterisation, in particular speaker selection (i.e. how to select a representative subset of speakers from a larger population), speaker representation (namely how to represent a new speaker in reference to a given speaker population), speaker adaptation for speech recognition, and more recently, speaker's emotion detection.

4.2.2. *Speech recognition for multimedia analysis*

In multimodal documents, the audio track is generally a major source of information and, when it contains speech, it conveys a high level of semantic content. In this context, speech recognition functionalities are essential for the extraction of information relevant to the tasks of content indexing.

As of today, there is no perfect technology able to provide an error-free speech retranscription and operating for any type of speech input. A current challenge is to be able to exploit the imperfect output of an Automatic Speech Recognition (ASR) system, using for instance Natural Language Processing (NLP) techniques, in order to extract structural (topic segmentation) and semantic (topic detection) information from the audio track.

Along the same line, another scientific challenge is to combine the ASR output with other sources of information coming from various modalities, in order to extract robust multi-modal indexes from a multimedia content (video, audio, textual metadata, etc...).

4.3. Description and structuration of audio streams

audio stream, audio detection, audio tracking, audio segmentation, audio descriptors, multimedia indexing, audiovisual integration, multimodality, information fusion, audio-visual descriptors,

Automatic tools to locate events in audio documents, structure them and browse through them as in textual documents are key issues in order to fully exploit most of the available audio documents (radio and television programmes and broadcasts, conference recordings, etc).

In this respect, defining and extracting meaningful characteristics from an audio stream aim at obtaining a structured representation of the document, thus facilitating content-based access or search by similarity.

Activities in METISS focus on sound class and event characterisation and tracking in audio contents for a wide variety of features and documents.

4.3.1. *Detecting and tracking sound classes and events*

Locating various sounds or broad classes of sounds, such as silence, music or specific events like ball hits or a jingle, in an audio document is a key issue as far as automatic annotation of sound tracks is concerned. Indeed, specific audio events are crucial landmarks in a broadcast. Thus, locating automatically such events enables to answer a query by focusing on the portion of interest in the document or to structure a document for further processing. Typical sound tracks come from radio or TV broadcasts, or even movies.

In the continuity of research carried out at IRISA for many years (especially by Benveniste, Basseville, André-Obrecht, Delyon, Seck, ...) the statistical test approach can be applied to abrupt changes detection and sound class tracking, the latter provided a statistical model for each class to be detected or tracked was previously estimated. For example, detecting speech segments in the signal can be carried out by comparing the segment likelihoods using a speech and a "non-speech" statistical model respectively. The statistical models commonly used typically represent the distribution of the power spectral density, possibly including some temporal constraints if the audio events to look for show a specific time structure, as is the case with jingles or words. As an alternative to statistical tests, hidden Markov models can be used to simultaneously segment and classify an audio stream. In this case, each state (or group of states) of the automaton represent one of the audio event to be detected. As for the statistical test approach, the hidden Markov model approach requires that models, typically Gaussian mixture models, are estimated for each type of event to be tracked.

In the area of automatic detection and tracking of audio events, there are three main bottlenecks. The first one is the detection of simultaneous events, typically speech with music in a speech/music/noise segmentation problem since it is nearly impossible to estimate a model for each event combination. The second one is the not so uncommon problem of detecting very short events for which only a small amount of training data is available. In this case, the traditional 100 Hz frame analysis of the waveform and Gaussian mixture modeling suffer serious limitations. Finally, typical approaches require a preliminary step of manual annotation of a training corpus in order to estimate some model parameters. There is therefore a need for efficient machine learning and statistical parameter estimation techniques to avoid this tedious and costly annotation step.

4.3.2. Describing multi-modal information

Applied to the sound track of a video, detecting and tracking audio events can provide useful information about the video structure. Such information is by definition only partial and can seldom be exploited by itself for multimedia document structuring or abstracting. To achieve these goals, partial information from the various media must be combined. By nature, pieces of information extracted from different media or modalities are heterogeneous (text, topic, symbolic audio events, shot change, dominant color, etc.) thus making their integration difficult. Only recently approaches to combine audio and visual information in a generic framework for video structuring have appeared, most of them using very basic audio information.

Combining multimedia information can be performed at various level of abstraction. Currently, most approaches in video structuring rely on the combination of structuring events detected independently in each media. A popular way to combine information is the hierarchical approach which consists in using the results of the event detection of one media to provide cues for event detection in the other media. Application specific heuristics for decision fusions are also widely employed. The Bayes detection theory provides a powerful theoretical framework for a more integrated processing of heterogeneous information, in particular because this framework is already extensively exploited to detect structuring events in each media. Hidden Markov models with multiple observation streams have been used in various studies on video analysis over the last three years.

The main research topics in this field are the definition of structuring events that should be detected on the one hand and the definition of statistical models to combine or to jointly model low-level heterogeneous information on the other hand. In particular, defining statistical models on low-level features is a promising idea as it avoids defining and detecting structuring elements independently for each media and enables an early integration of all the possible sources of information in the structuring process.

4.3.3. Recurrent audio pattern detection

A new emerging topic is that of motif discovery in large volumes of audio data, i.e. discovering similar units in an audio stream in an unsupervised fashion. These motifs can constitute some form of audio “miniatures” which characterize some potentially salient part of the audio content : key-word, jingle, etc...

This problem naturally requires the definition of a robust metric between audio segments, but a key issue relies in an efficient search strategy able to handle the combinatorial complexity stemming from the competition between all possible motif hypotheses. An additional issue is that of being able to model adequately the collection of instances corresponding to a same motif (in this respect, the HMM framework certainly offers a reasonable paradigm).

4.4. Advanced processing for music information retrieval

audio object, music description, music language modeling, multi-level representations

4.4.1. Music content modeling

Music pieces constitute a large part of the vast family of audio data for which the design of description and search techniques remain a challenge. But while there exist some well-established formats for synthetic music (such as MIDI), there is still no efficient approach that provide a compact, searchable representation of music recordings.

In this context, the METISS research group dedicates some investigative efforts in high level modeling of music content along several tracks. The first one is the acoustic modeling of music recordings by deformable probabilistic sound objects so as to represent variants of a same note as several realisation of a common underlying process. The second track is music language modeling, i.e. the symbolic modeling of combinations and sequences of notes by statistical models, such as n-grams.

4.4.2. Multi-level representations for music information retrieval

New search and retrieval technologies focused on music recordings are of great interest to amateur and professional applications in different kinds of audio data repositories, like on-line music stores or personal music collections.

The METISS research group is devoting increasing effort on the fine modeling of multi-instrument/multi-track music recordings. In this context we are developing new methods of automatic metadata generation from music recordings, based on Bayesian modeling of the signal for multilevel representations of its content. We also investigate uncertainty representation and multiple alternative hypotheses inference.

4.5. Audio scene analysis

source separation, multi-channel audio, source characterization, source localization, compressive sensing

Audio signals are commonly the result of the superimposition of various sources mixed together : speech and surrounding noise, multiple speakers, instruments playing simultaneously, etc...

Source separation aims at recovering (approximations of) the various sources participating to the audio mixture, using spatial and spectral criteria, which can be based either on a priori knowledge or on property learned from the mixture itself.

4.5.1. Audio source separation

The general problem of “source separation” consists in recovering a set of unknown sources from the observation of one or several of their mixtures, which may correspond to as many microphones. In the special case of *speaker separation*, the problem is to recover two speech signals contributed by two separate speakers that are recorded on the same media. The former issue can be extended to *channel separation*, which deals with the problem of isolating various simultaneous components in an audio recording (speech, music, singing voice, individual instruments, etc.). In the case of *noise removal*, one tries to isolate the “meaningful” signal, holding relevant information, from parasite noise.

It can even be appropriate to view audio compression as a special case of source separation, one source being the compressed signal, the other being the residue of the compression process. The former examples illustrate how the general source separation problem spans many different problems and implies many foreseeable applications.

While in some cases –such as multichannel audio recording and processing– the source separation problem arises with a number of mixtures which is at least the number of unknown sources, the research on audio source separation within the METISS project-team rather focusses on the so-called under-determined case. More precisely, we consider the cases of one sensor (mono recording) for two or more sources, or two sensors (stereo recording) for $n > 2$ sources.

We address the problem of source separation by combining spatial information and spectral properties of the sources. However, as we want to resort to as little prior information as possible we have designed self-learning schemes which adapt their behaviour to the properties of the mixture itself [1].

4.5.2. Compressive sensing of acoustic fields

Complex audio scene may also be dealt with at the acquisition stage, by using “intelligent” sampling schemes. This is the concept behind a new field of scientific investigation : compressive sensing of acoustic fields.

The challenge of this research is to design, implement and evaluate sensing architectures and signal processing algorithms which would enable to acquire a reasonably accurate map of an acoustic field, so as to be able to locate, characterize and manipulate the various sources in the audio scene.

MIMETIC Team

4. Application Domains

4.1. Motion Sensing

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intra-variability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics) or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method is used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

4.2. VR and Sports

Sport is characterized by complex displacements and motions. These motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. The perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequential but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understanding sports and interaction between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

4.3. Biomechanics and Motion Analysis

Biomechanics is obviously a very large domain. This large set can be divided regarding to the scale at which the analysis is performed going from microscopic evaluation of biological tissues' mechanical properties to macroscopic analysis and modeling of whole body motion. Our topics in the domain of biomechanics mainly lie within this last scope.

The first goal of such kind of research projects is a better understanding of human motion. The MimeTic team addresses three different situations: everyday motions of a lambda subject, locomotion of pathological subjects and sports gesture.

In the first set, Mimetic is interested in studying how subjects maintain their balance in highly dynamic conditions. Until now, balance have nearly always been considered in static or quasi-static conditions. The knowledge of much more dynamic cases still has to be improved. Our approach has demonstrated that first of all, the question of the parameter that will allow to do this is still open. We have also taken interest into collision avoidance between two pedestrian. This topic includes the research of the parameters that are interactively controlled and the study of each one's role within this interaction.

When patients, in particular those suffering from central nervous system affection, cannot have an efficient walking it becomes very useful for praticians to benefit from an objective evaluation of their capacities. To propose such help to patients following, we have developed two complementary indices, one based on kinematics and the other one on muscles activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous.

Finally, in sports, where gesture can be considered, in some way, as abnormal, the goal is more precisely to understand the determinants of performance. This could then be used to improve training programs or devices. Two different sports have been studied: the tennis serve, where the goal was to understand the contribution of each segments of the body in ball's speed and the influence of the mechanical characteristics of the fin in fin swimming.

After having improved the knowledge of these different gestures a second goal is then to propose modeling solutions that can be used in VR environments for other research topics within MimeTic. This has been the case, for exemple, for the colision avoidance.

4.4. Crowds

Crowd simulation is a very active and concurrent domain. Various disciplines are interested in crowds modeling and simulation: Mathematics, Cognitive Sciences, Physics, Computer Graphics, etc. The reason for this large interest is that crowd simulation raise fascinating challenges.

At first, crowd can be first seen as a complex system: numerous local interactions occur between its elements and results into macroscopic emergent phenomena. Interactions are of various nature and are undergoing various factors as well. Physical factors are crucial as a crowd gathers by definition numerous moving people with a certain level of density. But sociological, cultural and psychological factors are important as well, since crowd behavior is deeply changed from country to country, or depending on the considered situations.

On the computational point of view, crowd push traditional simulation algorithms to their limit. An element of a crowd is subject to interact with any other element belonging the same crowd, a naive simulation algorithm has a quadratic complexity. Specific strategies are set to face such a difficulty: level-of-detail techniques enable scaling large crowd simulation and reach real-time solutions.

MimeTIC is an international key contributor in the domain of crowd simulation. Our approach is specific and based on three axis. First, our modeling approach is founded on human movement science: we conducted challenging experiment on the motion of groups. Second: we developed high-performance solutions for crowd simulation. Third, we develop solutions for realistic navigation in virtual world to enable interaction with crowds in Virtual Reality.

4.5. Interactive Digital Storytelling

Interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computer-generated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools.

Of particular interest for the Mimetic research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the litterature.

4.6. Autonomous characters

Autonomous characters are becoming more and more popular has they are used in an increasing number of application domains. In the field of special effects, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors. In video games and virtual storytelling, autonomous characters play the role of actors that are driven by a scenario. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes.

One of the main challenges in the field of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes).

In the MimeTIC team, we focus on autonomous virtual humans. Our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

MINT Project-Team

4. Application Domains

4.1. Next-generation desktop systems

The term *desktop system* refers here to the combination of a window system handling low-level graphics and input with a window manager and a set of applications that share a distinctive look and feel. It applies not only to desktop PCs but also to any other device or combination of devices supporting graphical interaction with multiple applications. Interaction with these systems currently rely on a small number of interaction primitives such as text input, pointing and activation as well as a few other basic gestures. This limited set of primitives is one reason the systems are simple to use. There is, however, a cost. Most simple combinations being already used, few remain to trigger and control innovative techniques that could facilitate task switching or data management, for example. Desktop systems are in dire need of additional interaction primitives, including gestural ones.

4.2. Ambient Intelligence

Ambient intelligence (AmI) refers to the concept of being surrounded by intelligent systems embedded in everyday objects [33]. Envisioned AmI environments are aware of human presence, adapt to users' needs and are capable of responding to indications of desire and possibly engaging in intelligent dialogue. Ambient Intelligence should be unobtrusive: interaction should be relaxing and enjoyable and should not involve a steep learning curve. Gestural interaction is definitely relevant in this context.

4.3. Serious Games

Serious game refers to techniques extensively used in computer games, that are being used for other purposes than gaming. Fields such as learning, use of Virtual Reality for rehabilitation, 3D interactive worlds for retail, art-therapy, are specific context with which the MINT group has scientific connection, and industrial contacts. This field of application is a good opportunity for us to test and transfer our scientific knowledge and results.

4.4. Interactive Art

The heart of Mint project is about interaction gesture, and aims at making relation between application and user more intimate through the production of tools and methods for application to use more information from user gesture. There seems to be, at first sight, very strong difference of fields, tools, vocabulary, between Science and Art. Up to basic intellectual schemes are classically thought to be different. Yet, a closer look needs to be taken on things. Through time, Art is more and more involved in relation between people and content. For example, *relational art*¹ is centered on inter-human relations and social context. Because of this similar analysis about relation between person and content, research on interactive systems probably has a lot to develop relations with Art, this is also true for research on gestural interaction.

¹http://en.wikipedia.org/wiki/Relational_Art

MORPHEO Team

4. Application Domains

4.1. 4D modeling

Modeling shapes that evolve over time, analyzing and interpreting their motion has been a subject of increasing interest of many research communities including the computer vision, the computer graphics and the medical imaging communities. Recent evolutions in acquisition technologies including 3D depth cameras (Time-of-Light and Kinect), multi-camera systems, marker based motion capture systems, ultrasound and CT scans have made those communities consider capturing the real scene and their dynamics, create 4D spatio-temporal models, analyze and interpret them. A number of applications including dense motion capture, dynamic shape modeling and animation, temporally consistent 3D reconstruction, motion analyses and interpretation have therefore emerged.

4.2. Shape Analysis

Most existing shape analysis tools are local, in the sense that they give local insight about an object's geometry or purpose. The use of both geometry and motion clues makes it possible to recover more global information, in order to get extensive knowledge about a shape. For instance, motion can help to decompose a 3D model of a character into semantically significant parts, such as legs, arms, torso and head. Possible applications of such high-level shape understanding include accurate feature computation, comparison between models to detect defects or medical pathologies, and the design of new biometric models or new anthropometric datasets.

4.3. Human Motion Analysis

The recovery of dense motion information enables the combined analyses of shapes and their motions. Typical examples include the estimation of mean shapes given a set of 3D models or the identification of abnormal deformations of a shape given its typical evolutions. The interest arises in several application domains where temporal surface deformations need to be captured and analysed. It includes human body analyses for which potential applications with are anyway numerous and important, from the identification of pathologies to the design of new prostheses.

4.4. Interaction

The ability to build models of humans in real time allows to develop interactive applications where users interact with virtual worlds. The recent Kinect proposed by Microsoft illustrates this principle with game applications using human inputs perceived with a depth camera. Other examples include gesture interfaces using visual inputs. A challenging issue in this domain is the ability to capture complex scenes in natural environments. Multi-modal visual perception, e.g. depth and color cameras, is one objective in that respect.

MOSTRARE Project-Team

4. Application Domains

4.1. Context

XML transformations are basic to data integration: HTML to XML transformations are useful for information extraction from the Web; XML to XML transformations are useful for data exchange between Web services or between peers or between databases. Doan and Halevy [40] survey novel integration tasks that appear with the Semantic Web and the usage of ontologies. Therefore, the semi-automatic generation of XML transformations is a challenge in the database community and in the semantic Web community.

Also, XML transformations are useful for document processing. For instance, there is need of designing transformations from documents organized w.r.t visual format (HTML, DOC, PDF) into documents organized w.r.t. semantic format (XML according to a DTD or a schema). The semi-automatic design of such transformations is obviously a very challenging objective.

Furthermore, quite some activities of Mostrare concern efficient evaluation of XPath queries on XML documents and XML streams. XPath is fundamental to all XML standards, in particular to XQuery, XSLT, and XProc.

OAK Team

4. Application Domains

4.1. Online content management

This concerns archiving filtered content from online information sources (journals, blogs, ...) with the purpose of recording their perspective on facts involving specific countries, regions or enterprises, key actors etc. We have recently explored such an application in a demonstration proposal, built using our XR model for XML documents with semantic annotations (currently under evaluation). The benefit of XR was to simplify and streamline the specification of document- and semantics-rich data management applications. Support for documents allows to keep an evidence (source) for facts and statements that may be extracted from them. At the same time, support for semantics allows to connect documents to people, ideas, trends etc. and to reason across data sources [9].

4.2. Open data intelligence

Open data intelligence: the goal is to build and efficiently exploit warehouses of Open Data, integrated from several data sources on the Web, in order to produce consolidated rich information to be given to decision makers and to the citizens. Such projects have been started in France notably by the city areas of Rennes (pioneer in Open Data usage), Paris, and more recently by Grenoble, in a project to which we participate; the ICT Labs DataBridges also focused on such topics. Oak competencies required for such projects are related to large-scale RDF data management as well as to the design of innovative data models for semantic-rich content.

4.3. Efficient complex data management in the cloud

In a cloud environment, data catalogs and indices need to be efficiently built and maintained, typically in parallel; queries need to be routed to only those data subsets which are likely to lead to results, and efficiently executed, using parallelism and the available indexes. We work on such topics within the Europa ICT Labs activity. Algorithms for efficiently handling indexes and views in the cloud for heterogeneous, complex-structure data are also at the core of our work proposal in the Datalyse project (see below). Our 2012 work in this context has led to [10], [12], [13].

4.4. 360 degree customer view

Companies seek to gather as much information as possible about their customers, for instance for more targeted publicity campaigns, market analysis, offer personalization, etc. That is, they want to consider data beyond the data they collected about their customers in their proprietary databases. Complementary data include for instance social data extracted from customers' activities in social networks, public data related to their place of residence (e.g., crime rate or housing price evolution). To achieve this goal, data integration on highly heterogeneous and massive data is necessary. Furthermore, as one means to assess both the correctness of the integration result and the quality (in this application most notably trustworthiness), we can resort to data provenance. These topics are explored in the project Datalyse which we submitted in 2012 to the French national "AAP Cloud 3: Big Data" call, a project headed by the "Business & Decision" company and whose ongoing evaluation will continue in early 2013.

ORPAILLEUR Project-Team

4. Application Domains

4.1. Life Sciences

Participants: Yasmine Assess, Thomas Bourquard, Emmanuel Bresso, Marie-Dominique Devignes, Elias Egho, Anisah Ghoorah, Renaud Grisoni, Nicolas Jay, Bernard Maigret, Amedeo Napoli, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone, Yannick Toussaint.

knowledge discovery in life sciences, bioinformatics, biology, chemistry, gene

Knowledge discovery in life sciences is a process for extracting knowledge units from large biological databases, e.g. collection of genes.

One major application domain which is currently investigated by Orpailleur team is related to life sciences, with particular emphasis on biology, medicine, and chemistry. The understanding of biological systems provides complex problems for computer scientists, and, when they exist, solutions bring new research ideas for biologists and for computer scientists as well. Accordingly, the Orpailleur team includes biologists, chemists, and a physician, making Orpailleur a very original EPI at Inria.

Knowledge discovery is gaining more and more interest and importance in life sciences for mining either homogeneous databases such as protein sequences and structures, or heterogeneous databases for discovering interactions between genes and environment, or between genetic and phenotypic data, especially for public health and pharmacogenomics domains. The latter case appears to be one main challenge in knowledge discovery in biology and involves knowledge discovery from complex data and thus KDDK. The interactions between researchers in biology and researchers in computer science improve not only knowledge about systems in biology, chemistry, and medicine, but knowledge about computer science as well. Solving problems for biologists using KDDK methods involves the design of specific modules that, in turn, leads to adaptations of the KDDK process, especially in the preparation of data and in the interpretation of the extracted units.

4.2. Knowledge Management in Medicine

Participants: Nicolas Jay, Jean Lieber, Thomas Meilender, Amedeo Napoli.

knowledge representation, description logics, classification-based reasoning, case-based reasoning, formal concept analysis, semantic web

The Kasimir research project holds on decision support and knowledge management for the treatment of cancer [103]. This is a multidisciplinary research project in which participate researchers in computer science (Orpailleur), experts in oncology (“Centre Alexis Vautrin” in Vandœuvre-lès-Nancy), Oncolor (a healthcare network in Lorraine involved in oncology), and A2Zi (a company working in Web technologies and involved in several projects in the medical informatics domain, <http://www.a2zi.fr/>). For a given cancer localization, a treatment is based on a protocol similar to a medical guideline, and is built according to evidence-based medicine principles. For most of the cases (about 70%), a straightforward application of the protocol is sufficient and provides a solution, i.e. a treatment, that can be directly reused. A case out of the 30% remaining cases is “out of the protocol”, meaning that either the protocol does not provide a treatment for this case, or the proposed solution raises difficulties, e.g. contraindication, treatment impossibility, etc. For a case “out of the protocol”, oncologists try to *adapt* the protocol. Actually, considering the complex case of breast cancer, oncologists discuss such a case during the so-called “breast cancer therapeutic decision meetings”, including experts of all specialties in breast oncology, e.g. chemotherapy, radiotherapy, and surgery.

The semantic Web technologies have been used and adapted in the Kasimir project for several years. Recently, a semantic wiki has been deployed in its production version (<http://www.oncologik.fr>) It allows the management of decision protocols [48] More precisely, the migration from the static HTML site of Oncolor to a semantic wiki (with limited editing rights and unlimited reading rights) has been done [49]. This has consequences on the editorial chain of the published protocols which is more collaborative. A decision tree editor that has been integrated into the wiki and that has an export facility to formalized protocols in OWL DL has also been developed [61].

4.3. Cooking

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Laura Infante Blanco, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

cooking, knowledge representation, knowledge discovery, case-based reasoning, semantic wiki

The origin of the Taaable project is the Computer Cooking Contest (CCC). A contestant of the CCC is a system that answers queries of recipes, using a recipe base; if no recipe exactly matches the query, then the system adapts another recipe. Taaable is a case-based reasoning system that uses various technologies used and developed in the Orpailleur team, such as technologies of the semantic web, knowledge discovery techniques, knowledge representation and reasoning techniques, etc. From a research viewpoint it enables to test the scientific results on an application domain that is at the same time simple to understand and raising complex issues, and to study the complementarity of various research domains. Taaable has been at the origin of the project Kolflow of the ANR CONTINT program, whose application domain is WikiTaaable, the semantic wiki of Taaable. It is also used for other projects under submission.

PAROLE Project-Team

4. Application Domains

4.1. Application Domains

Our research is applied in a variety of fields from ASR to paramedical domains. Speech analysis methods will contribute to the development of new technologies for language learning (for hearing-impaired persons and for the teaching of foreign languages) as well as for hearing aids. In the past, we developed a set of teaching tools based on speech analysis and recognition algorithms of the group (cf. the ISAEUS [43] project of the EU that ended in 2000). We are continuing this effort towards the diffusion of a course on Internet.

Speech is likely to play an increasing role in man-machine communication. Actually, speech is a natural mean of communication, particularly for non-specialist persons. In a multimodal environment, the association of speech and designation gestures on touch screens can, for instance, simplify the interpretation of spatial reference expressions. Besides, the use of speech is mandatory in many situations where a keyboard is not available: mobile and on-board applications (for instance in the framework of the HIWIRE European project for the use of speech recognition in a cockpit plane), interactive vocal servers, telephone and domestic applications, etc. Most of these applications will necessitate to integrate the type of speech understanding process that our group is presently studying. Furthermore, speech to speech translation concerns all multilingual applications (vocal services, audio indexing of international documents). The automatic indexing of audio and video documents is a very active field that will have an increasing importance in our group in the forthcoming years, with applications such as economic intelligence, keyword spotting and automatic categorization of mails.

PERCEPTION Team

4. Application Domains

4.1. Human action recognition

We are particularly interested in the analysis and recognition of human actions and gestures. The vast majority of research groups concentrate on isolated action recognition. We address continuous recognition. The problem is difficult because one has to simultaneously address the problems of recognition and segmentation. For this reason, we adopt a per-frame representation and we develop methods that rely on dynamic programming and on hidden Markov models. We investigate two type of methods: one-pass methods and two-pass methods. One-pass methods enforce both within-action and between-action constraints within sequence-to-sequence alignment algorithms such as dynamic time warping or the Viterbi algorithm. Two-pass methods combine a per-action representation with a discriminative classifier and with a dynamic programming post-processing stage that find the best sequence of actions. These algorithms were well studied in the context of large-vocabulary continuous speech recognition systems. We investigate the modeling of various per-frame representations for action and gesture analysis and we devise one-pass and two-pass algorithms for recognition.

4.2. 3D reconstruction using TOF and color cameras

TOF cameras are active-light range sensors. An infrared beam of light is generated by the device and depth values can be measured by each pixel, provided that the beam travels back to the sensor. The associated depth measurement is accurate if the sensed surface sends back towards the sensor a fair percentage of the incident light. There is a large number of practical situations where the depth readings are erroneous: specular and bright surfaces (metal, plastic, etc.), scattering surfaces (hair), absorbing surfaces (cloth), slanted surfaces, e.g., at the bounding contours of convex objects which are very important for reconstruction, mutual reflections, limited range, etc. The resolution of currently available TOF cameras is of 0.3 to 0.5MP. Modern 2D color cameras deliver 2MP images at 30FPS or 5MP images at 15FPS. It is therefore judicious to attempt to combine the active-range and the passive-stereo approaches within a mixed methodology and system. Standard stereo matching methods provide an accurate depth map but are often quite slow because of the inherent complexity of the matching algorithms. Moreover, stereo matching is ambiguous and inaccurate in the presence of weakly textured areas. We develop TOF-stereo matching and reconstruction algorithms that are able to combine the advantages of the two types of depth estimation technologies.

4.3. Sound-source separation and localization

We explore the potential of binaural audition in conjunction with modern machine learning methods in order to address the problems of sound source separation and localization. We exploit the spectral properties of interaural cues, namely the interaural level difference (ILD) and the interaural phase difference (IPD). We have started to develop a novel supervised framework based on a training stage. During this stage, a sound source emits a broadband random signal which is perceived by a microphone pair embedded into a dummy head with a human-like head related transfer function (HRTF). The source emits from a location parameterized by azimuth and elevation. Hence, a mapping between a high-dimensional interaural spectral representation and a low-dimensional manifold can be estimated from these training data. This allows the development of various single-source localization methods as well as multiple-source separation and localization methods.

4.4. Audio-visual fusion for human-robot interaction

Modern human-robot interaction systems must be able to combine information from several modalities, e.g., vision and hearing, in order to allow high-level communication via gesture and vocal commands, multimodal dialogue, and recognition-action loops. Auditory and visual data are intrinsically different types of sensory data. We have started the development of a audio-visual mixture model that takes into account the heterogenous nature of visual and auditory observations. The proposed multimodal model uses modality specific mixtures (one mixture model for each modality). These mixtures are tied through latent variables that parameterize the joint audiovisual space. We thoroughly investigate this novel kind of mixtures with their associated efficient parameter estimation procedures.

POTIOC Team

4. Application Domains

4.1. Application domains

Since our project aims at providing 3D digital worlds to all, including the general public, in order to stimulate understanding, learning, communication and creation, our scope of applications will naturally be the following one:

- **Culture and education:** We are convinced that a 3D digital world is a powerful media that may contribute to enhance understanding processes. For example, a museum would benefit from new 3D user interfaces allowing visitors to better understand complex content. Similarly, at school, this media has an extraordinary potential for enhancing learning. For example, a child being able to navigate in archaeological sites, or being able to manipulate by himself 3D molecules as described previously, will probably understand and learn things while having pleasure in interacting with the content.
- **Art:** We believe that 3D digital worlds may stimulate creativity, too. Our first investigations with music and drawings have shown that this media opens new possibilities for creation. The challenge here will be to design good interfaces that will allow artists to explore new dimensions. The user may be an experienced artist, or a three years old child who would express his creativity thought tools that go beyond papers and pens.
- **Healthcare:** People suffering from motor or cognitive impairments are one of the target populations of the Potioc project. Indeed, we believe that new interfaces that exploit 3D digital worlds may help people to overcome their disabilities. For example, someone with very reduced motor capabilities could benefit from BCI to explore a virtual museum, or a children having difficulties for concentrating may benefit from new 3D interactive systems.
- **Entertainment:** The objective of Potioc is to open 3D digital worlds to everyone by designing innovative interfaces driven by enjoyment. Consequently, the entertainment industry will be an obvious application domain where Potioc can contribute. This can be in the scope of videogames, entertainment parks, Web and TV of the future, applications for mobile devices, and so on.

Naturally, we will not necessarily address all these applications, and certainly not all at the same time. These areas just define the applicative scope of our work. As an example, much of our current work is targeted at artistic and entertainment applications, with VR-based musical performances, augmented paper-based drawing or BCI-based video games. We are also currently starting to conduct research on digital cities, in order to provide ordinary citizens suitable tools and UI to explore 3D content related to their city, such as 3D maps, information about population density or sound nuisance, among other. It should also be noted that our work might find applicative connexions outside these main application domains and benefit to a large range of academic and industrial areas, with which we could build relationships. For example, in the scope of medicine, new and easy to adopt user interfaces designed in Potioc could prove valuable for medical professionals as well, to better access and interact with biological 3D content (e.g., X-rays or MRI scans).

PRIMA Project-Team

4. Application Domains

4.1. The Smart Spaces Research Platform

Participants: Rémi Barraquand, James Crowley [correspondant], Augustin Lux, Amaury Nègre, Patrick Reignier, Lukas Rummelhard, Dominique Vaufreydaz.

Smart Spaces, Observation of human activity, context aware systems and services.

Project PRIMA has recently moved to a new Smart Spaces Research Platform in order to develop and test components and services for context aware human centered services. The Smart Spaces Research Platform is a 50 Square Meter space equipped with a microphone array, wireless lapel microphones, wide angle surveillance cameras, panoramic cameras, steerable cameras, scanning range sensors and two camera-projector video-interaction devices, and a KNX smart electrical system. The microphone array is used as an acoustic sensor to detect, locate and classify acoustic signals for recognizing human activities. The wide-angle and panoramic cameras provide fields of view that cover the entire room, and allows detection and tracking of individuals. Steerable cameras are used to acquire video of activities from any viewing direction.

Context aware human centered services may be categorized as tools, advisors, or media. Tool services are designed to perform a specific task or function as robustly as possible. If any adaptation is involved, it should serve to adapt the function to a changing environment. The user interface, and any interaction with users should be perfectly predictable. The degree to which the operation of a tool should be transparent, visible or hidden from the user is an open research question. Advisor services observe the users actions and environment in order to propose information on possible courses of actions. Advisors should be completely obedient and non-disruptive. They should not take initiatives or actions that cannot be overridden or controlled by the user. Media services provide interpersonal communications, entertainment or sensorial extension.

Examples of human centered tool services include:

- An activity log recording system that records the events and activities of an individual's daily activities.

- A service that integrates control of heating, air-conditioning, lighting, windows, window-shades, exterior awnings, etc to provide an optimum comfort level defined in terms of temperature, humidity, CO₂, acoustic noise and ambient light level.

- A service that manages the available stock of supplies in a home and orders supplies over the Internet to assure that the appropriate level of supplies are always available.

- A service to measure the walking rate, step size and posture of an elderly person to estimate health and predict the likelihood of a fall.

Some examples of advisor services include:

- A service that provides shopping advice about where and when to shop.

- A service that can propose possible menus based on the available food stuffs in the kitchen.

- A service that observes the activities of humans and appliances within the home and can suggest ways to reduce the cost of heating, electricity or communications.

- A service that observes lifestyle and can offer advice about improving health.

Some examples of media services include

- A service that maintains a sense of informal non-disruptive presence with distant family members.

- A robot device that communicates affection.

- A device that renders the surface temperature of wall, floors and windows to show energy consumption and loss within a house.

- Services that enable seamless tele-presence for communication with others

4.2. Ambient Services for Assisted Living

Participants: Rémi Barraquand, James Crowley, Frédéric Devernay, Amaury Nègre.

Ambient Assisted Living, Monitoring Services, Presence awareness.

The continued progress in extending life-span, coupled with declining birth rates have resulted in a growing number of elderly people with varying disabilities who are unable to conduct a normal life at home, thereby becoming more and more isolated from society. Governmental agencies including hospitals, healthcare institutions and social care institutions are increasingly overburdened with care of this growing population. Left unchecked, economic and man-power requirements for care of the elderly could well trigger a societal and economic crisis. There is an urgent societal need for technologies and services that allow elderly people to live autonomously in their own environments for longer periods. Smart environments provide a promising new enabling technology for such services.

Adapting smart environments to enhance the autonomy and quality of life for elderly require:

- Robust, plug-and-play sensor technologies monitor the activities and health of elderly in their own home environments.

- Easy to use communications services that allow people to maintain a sense of presence to avoid isolation without disrupting privacy or distracting attention from normal daily activities.

- Architectural frameworks that allow ad hoc composition of services from distributed heterogeneous components scattered throughout the environment.

- Distributed system architectures which allow the cooperation of independent emergency services to work together to provide emergency care,

- Technologies interpret activity to warn of loss of mobility or cognitive function.

- Engineering approaches for the customization/personalization/adaptation of living assistance systems at installation and run time,

- Social, privacy, ethical and legal safeguards for privacy and control of personal data.

4.3. 3-D video processing

Participants: Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

3-D display, Stereoscapy, View Interpolation, Auto-calibration

Stereoscopic cinema has seen a surge of activity in recent years, and for the first time all of the major Hollywood studios released 3-D movies in 2009. This is happening alongside the adoption of 3-D technology for sports broadcasting, and the arrival of 3-D TVs for the home. Two previous attempts to introduce 3-D cinema in the 1950s and the 1980s failed because the contemporary technology was immature and resulted in viewer discomfort. But current technologies such as accurately-adjustable 3-D camera rigs with onboard computers to automatically inform a camera operator of inappropriate stereoscopic shots, digital processing for post-shooting rectification of the 3-D imagery, digital projectors for accurate positioning of the two stereo projections on the cinema screen, and polarized silver screens to reduce cross-talk between the viewers left- and right-eyes mean that the viewer experience is at a much higher level of quality than in the past. Even so, creation of stereoscopic cinema is an open, active research area, and there are many challenges from acquisition to post-production to automatic adaptation for different-sized display [41], [43].

Until recently, in order to view stereoscopic 3-D video, the user had to wear special glasses. Recent advances in 3-D displays provide true 3-D viewing experience without glasses. These screens use either a micro-lenticular network or a parallax barrier placed in front of a standard LCD, plasma, or LED display, so that different viewpoints provide different images. If the characteristics of the network and the screen are carefully chosen, the user will perceive two different images from the viewpoints of the left and right eyes. Such glasses-free 3-D screens usually display between 8 and a few dozen different viewpoints.

When the 3-D scene which has to be displayed is computer-generated, it is usually not a problem to generate a few dozen viewpoints. But when a real scene has to be displayed, one would have to shoot it through the same number of synchronized cameras as there are viewpoints in order to display it properly. This makes 3-D shooting of real scenes for glasses-free 3-D displays mostly unpractical. For this reason, we are developing high-quality view-interpolation techniques, so that the many different viewpoints can be generated from only a few camera positions [42].

Our research focuses on algorithms derived from Computer Vision and Computer Graphics, applied to live-action stereoscopic 3-D content production or post-production, including [40]:

- Live monitoring of stereoscopic video: geometric image misalignment, depth budget (i.e. limits on horizontal disparity), left-right color balance, left-right depth-of-field consistency [59].

- Live correction of stereoscopic video: correct the above defects in real-time when it is possible, with the help of GPU-based architectures.

- Adaptation of the stereoscopic content to the display size and distance, to avoid divergence or geometric deformations [42].

- Novel camera setups and algorithms for unconstrained stereoscopic shooting (especially when using long focal length).

- Novel camera setups and algorithms for glasses-free 3D displays.

- Stereoscopic inpainting.

- Stereoscopic match-moving.

- Compositing stereoscopic video and matte painting without green screen.

- Relighting of stereoscopic video, especially when videos are composited.

4.4. User localization in large-scale Smart Spaces

Participants: Amaury Nègre, Lukas Rummelhard, Dominique Vaufreydaz [correspondant].

Multi-modal perception, Smart Spaces, Localisation.

Ad-hoc assemblies of mobile devices embedding sensing, display, computing, communications, and interaction provide an enabling technology for smart environments. In the PRIMA project we have adopted a component oriented programming approach to compose smart services for such environments. Common services for smart spaces include

- Services to manage energy in building, including regulating temperature, illumination, and acoustic noise,

- Ambient assisted living services to extend the autonomy of elderly and infirm,

- Logistics management for daily living,

- Communication services and tools for collaborative work,

- Services for commercial environments,

- Orientation and information services for public spaces, and

- Services for education and training.

We are pursuing development of components based on the concept of "large-scale" smart space that is an intelligent environment which will be deployed on a large surface containing several buildings (as a university campus for example). We also define the "augmented man" concept as a human wearing one or many mobile intelligent wireless devices (telephone, Smartphone, pda, notebook). Using all these devices, one can use many different applications (read emails, browse the Internet, file exchange, etc.). By combining the concepts of large-scale perceptive environments and mobile computing, we can create intelligent spaces, it becomes possible to propose services adapted to individuals and their activities. We are currently focussing on two aspects of this problem: the user profile and the user location within a smart space.

A fundamental requirement for such services is the ability to perceive the current state of the environment. Depending on the nature of the service, environment state can require sensing and modeling the physical properties of the environment, the location, identity and activity of individuals within the environment, as well as the set of available computing devices and software components that compose the environment. All of these make up possible elements for context modeling.

Observing and tracking people in smart environments remains a challenging fundamental problem. Whether it is at the scale of a campus, of a building or more simply of a room, we can combine several additional localization levels (and several technologies) to allow a more accurate and reliable user perception system. Within the PRIMA project, we are currently experimenting with a multi-level localization system allowing variable granularity according to the available equipment and the precision required for the targeted service.

REVES Project-Team (section vide)

SEMAGRAMME Team

4. Application Domains

4.1. Introduction

Our applicative domains concern natural language processing applications that rely on a deep semantic analysis. For instance, one may cite the following ones:

- textual entailment and inference;
- dialogue systems;
- semantic-oriented query systems;
- content analysis of unstructured documents;
- text transformation and automatic summarization;
- (semi) automatic knowledge acquisition.

However, if the need for semantics seems to be ubiquitous. There is a challenge in finding applications for which a deep semantic analysis results in a real improvement over non semantic-based techniques.

4.2. Text Transformation

Text transformation is an application domain featuring two important sub-fields of computational linguistics:

- parsing: from surface form to abstract representation;
- generation: from abstract representation to surface form.

Text simplification or automatic summarization belong to that domain.

We aim at using the framework of Abstract Categorical Grammars we develop to this end. It is indeed a *reversible* framework that allows both parsing and generation. Its underlying mathematical structure of λ -calculus makes it fit with our type-theoretic approach to discourse dynamics modeling. The ANR project POLYMNIE(see section 7.2.1.1) is especially dedicated to this aim.

SIROCCO Project-Team

4. Application Domains

4.1. Introduction

The application domains addressed by the project are:

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images)
- Networked multimedia applications via their various needs in terms of image and 2D and 3D video compression, or in terms of network adaptation (e.g., resilience to channel noise)
- Content editing and post-production

4.2. Compression with advanced functionalities

Compression of images and of 2D video (including High Definition and Ultra High Definition) remains a widely-sought capability for a large number of applications. The continuous increase of access network bandwidth leads to increasing numbers of networked digital content users and consumers which in turn triggers needs for higher core bandwidth and higher compression efficiencies. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

In medical imaging, the large increase of medical analysis using various image sources for clinical purposes and the necessity to transmit or store these image data with improved performances related to transmission delay or storage capacities, command to develop new coding algorithms with lossless compression algorithms or *almost* lossless compression characteristics with respect to the medical diagnosis.

4.3. Networked visual applications

3D and Free Viewpoint TV: The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. It also allows visualizing the scene from any viewpoint, for scene navigation and free-viewpoint TV (FTV) applications. However, the large volumes of data associated to multi-view video plus depth content raise new challenges in terms of compression and communication.

Internet and mobile video: Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

Mobile video retrieval: The Internet has changed the ways of interacting with content. The user is shifting its media consumption from a passive to a more interactive mode, from linear broadcast (TV) to on demand content (YouTubes, iTunes, VoD), and to user-generated, searching for relevant, personalized content. New mobility and ubiquitous usage has also emerged. The increased power of mobile devices is making content search and retrieval applications using mobile phones possible. Quick access to content in mobile environments with restricted bandwidth resources will benefit from rate-efficient feature extraction and description.

Wireless multi-camera vision systems: Our activities on scene modelling, on rate-efficient feature description, distributed coding and compressed sensing should also lead to algorithmic building blocks relevant for wireless multi-camera vision systems, for applications such as visual surveillance and security.

4.4. Medical Imaging (CT, MRI, Virtual Microscopy)

The use of medical imaging has greatly increased in recent years, especially with *magnetic resonance images (MRI) and computed tomography (CT)*. In the medical sector, lossless compression schemes are in general used to avoid any signal degradation which could mask a pathology and hence disturb the medical diagnosis. Nevertheless, some discussions are on-going to use near-lossless coding of medical images, coupled with a detection and segmentation of region-of interest (ROIs) guided by a modeling stage of the image sensor, a precise knowledge of the medical imaging modalities and by the diagnosis and expertise of practitioners. New application domains using these new approaches of telemedicine will surely increase in the future. The second aspect deals with the legal need of biomedical images storage. The legacy rules of such archives are changing and it could be interesting to propose adaptive compression strategies, i.e to explore reversible lossy-to-lossless coding algorithms and new storage modalities which use, in a first stage, the lossless representation and continuously introduce controlled lossy degradations for the next stages of archives. Finally, it seems promising to explore new representation and coding approaches for 3D biological tissue imaging captured by *3D virtual microscopy*. These fields of interest and scientific application domains commonly generate terabytes of data. Lossless schemes but also lossy approaches have to be explored and optimized, and interactive tools supporting scalable and interactive access to large-sized images such as these virtual microscopy slides need to be developed.

4.5. Editing and post-production

Video editing and post-production are critical aspects in the audio-visual production process. Increased ways of “consuming” video content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content captured at very high resolutions may need to be repurposed in order to be adapted to the requirements of actual users, to the transmission channel or to the terminal. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools,

more visual effects, with more realism become possible. Other applications such as video annotation/retrieval, video restoration/stabilization, augmented reality, can also benefit from the proposed research.

SMIS Project-Team

4. Application Domains

4.1. Application Domains

data privacy, personal information management, healthcare, ambient intelligence

Our work addresses varied application domains. Typically, data management techniques on chip are required each time data-driven applications have to be embedded in ultra-light computing devices. This situation occurs for example in healthcare applications where medical folders are embedded into smart tokens (e.g., smart cards, secured USB keys), in telephony applications where personal data (address book, agenda, etc.) is embedded into cellular phones, in sensor networks where sensors log raw measurements and perform local computation on them, in smart-home applications where a collection of smart appliances gather information about the occupants to provide them a personalized service, and more generally in most applications related to ambient intelligence.

Safeguarding data confidentiality has become a primary concern for citizens, administrations and companies, broadening the application domains of our work on access control policies definition and enforcement. The threat on data confidentiality is manifold: external and internal attacks on the data at rest, on the data on transit, on the data hosted in untrusted environments (e.g., Database Service Providers, Web-hosting companies) and subject to illegal usage, insidious gathering of personal data in an ambient intelligence surrounding. Hence, new access control models and security mechanisms are required to accurately declare and safely control who is granted access to which data and for which purpose.

While the application domain mentioned above is rather large, one application is today more specifically targeted by the SMIS project. This application deals with privacy preservation in EHR (Electronic Health Record) systems. Several countries (including France) launched recently ambitious EHR programs where medical folders will be centralized and potentially hosted by private Database Service Providers. Centralization and hosting increase the risk of privacy violation. In 2007, we launched two projects (PlugDB and DMSP) tackling precisely this issue, with the final objective to experiment our technologies in the field. In 2011, we launched a new project (KISS) capitalizing on the previous ones and extending their scope towards the protection of any personal data delivered to individuals in an electronic form.

STARS Team

4. Application Domains

4.1. Introduction

While in our research the focus is to develop techniques, models and platforms that are generic and reusable, we also make effort in the development of real applications. The motivation is twofold. The first is to validate the new ideas and approaches we introduce. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, Stars focuses on two main domains: **video analytics** and **healthcare monitoring**.

4.2. Video Analytics

Our experience in video analytics [7], [1], [9] (also referred to as visual surveillance) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

For instance, the Keeneo start-up was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in video analytics (VSIP library, which was a previous version of SUP). Keeneo has been bought by Digital Barriers in August 2011 and is now independent from Inria. However, Stars continues to maintain a close cooperation with Keeneo for impact analysis of VSIP and for exploitation of new results.

Moreover new challenges are arising from the visual surveillance community. For instance, people detection and tracking in a crowded environment are still open issues despite the high competition on these topics. Also detecting abnormal activities may require to discover rare events from very large video data bases often characterized by noise or incomplete data.

4.3. Healthcare Monitoring

We have initiated a new strategic partnership (called CobTek) with Nice hospital [62], [81] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longer analysis. Finer activity description is needed for instance, to discriminate the activities (e.g. sitting, walking, eating) of Alzheimer patients from the ones of healthy older people. It is essential to be able to pre-diagnose dementia and to provide a better and more specialised care. Longer analysis is required when people monitoring aims at measuring the evolution of patient behavioural disorders. Setting up such long experimentation with dementia people has never been tried before but is necessary to have real-world validation. This is one of the challenge of the European FP7 project Dem@Care where several patient homes should be monitored over several months.

For this domain, a goal for Stars is to allow people with dementia to continue living in a self-sufficient manner in their own homes or residential centers, away from a hospital, as well as to allow clinicians and caregivers remotely proffer effective care and management. For all this to become possible, comprehensive monitoring of the daily life of the person with dementia is deemed necessary, since caregivers and clinicians will need a comprehensive view of the person's daily activities, behavioural patterns, lifestyle, as well as changes in them, indicating the progression of their condition.

The development and ultimate use of novel assistive technologies by a vulnerable user group such as individuals with dementia, and the assessment methodologies planned by Stars are not free of ethical, or even legal concerns, even if many studies have shown how these Information and Communication Technologies (ICT) can be useful and well accepted by older people with or without impairments. Thus one goal of Stars team is to design the right technologies that can provide the appropriate information to the medical carers while preserving people privacy. Moreover, Stars will pay particular attention to ethical, acceptability, legal and privacy concerns that may arise, addressing them in a professional way following the corresponding established EU and national laws and regulations, especially when outside France.

As presented in 3.1, Stars aims at designing cognitive vision systems with perceptual capabilities to monitor efficiently people activities. As a matter of fact, vision sensors can be seen as intrusive ones, even if no images are acquired or transmitted (only meta-data describing activities need to be collected). Therefore new communication paradigms and other sensors (e.g. accelerometers, RFID, and new sensors to come in the future) are also envisaged to provide the most appropriate services to the observed people, while preserving their privacy. To better understand ethical issues, Stars members are already involved in several ethical organizations. For instance, F. Bremond has been a member of the ODEGAM - "Commission Ethique et Droit" (a local association in Nice area for ethical issues related to older people) from 2010 to 2011 and a member of the French scientific council for the national seminar on "La maladie d'Alzheimer et les nouvelles technologies - Enjeux éthiques et questions de société" in 2011. This council has in particular proposed a chart and guidelines for conducting researches with dementia patients.

For addressing the acceptability issues, focus groups and HMI (Human Machine Interaction) experts, will be consulted on the most adequate range of mechanisms to interact and display information to older people.

TEXMEX Project-Team

4. Application Domains

4.1. Copyright protection of images and videos

With the proliferation of high-speed Internet access, piracy of multimedia data has developed into a major problem and media distributors, such as photo agencies, are making strong efforts to protect their digital property. Today, many photo agencies expose their collections on the web with a view to selling access to the images. They typically create web pages of thumbnails, from which it is possible to purchase high-resolution images that can be used for professional publications. Enforcing intellectual property rights and fighting against copyright violations is particularly important for these agencies, as these images are a key source of revenue. The most problematic cases, and the ones that induce the largest losses, occur when “pirates” steal the images that are available on the Web and then make money by illegally reselling those images.

This applies to photo agencies, and also to producers of videos and movies. Despite the poor image quality, thousands of (low-resolution) videos are uploaded every day to video-sharing sites such as YouTube, eDonkey or BitTorrent. In 2005, a study conducted by the Motion Picture Association of America was published, which estimated that their members lost 2,3 billion US\$ in sales due to video piracy over the Internet. Due to the high risk of piracy, movie producers have tried many means to restrict illegal distribution of their material, albeit with very limited success.

Photo and video pirates have found many ways to circumvent even the protection mechanisms. In order to cover up their tracks, stolen photos are typically cropped, scaled, their colors are slightly modified; videos, once ripped, are typically compressed, modified and re-encoded, making them more suitable for easy downloading. Another very popular method for stealing videos is cam-cording, where pirates smuggle digital camcorders into a movie theater and record what is projected on the screen. Once back home, that goes to the web.

Clearly, this environment calls for an automatic content-based copyright enforcement system, for images, videos, and also audio as music gets heavily pirated. Such a system needs to be effective as it must cope with often severe attacks against the contents to protect, and efficient as it must rapidly spot the original contents from a huge reference collection.

4.2. Video database management

The existing video databases are generally little digitized. The progressive migration to digital television should quickly change this point. As a matter of fact, the French TV channel TF1 switched to an entirely digitized production, the cameras remaining the only analogical spot. Treatment, assembly and diffusion are digital. In addition, domestic digital decoders can, from now on, be equipped with hard disks allowing a storage initially modest, of ten hours of video, but larger in the long term, of a thousand of hours.

One can distinguish two types of digital files: private and professional files. On one hand, the files of private individuals include recordings of broadcasted programs and films recorded using digital camcorders. It is unlikely that users will rigorously manage such collections; thus, there is a need for tools to help the user: Automatic creation of summaries and synopses to allow finding information easily or to have within few minutes a general idea of a program. Even if the service is rustic, it is initially evaluated according to the added value brought to a system (video tape recorder, decoder), must remain not very expensive, but will benefit from a large diffusion.

On the other hand, these are professional files: TV channel archives, cineclubs, producers... These files are of a much larger size, but benefit from the attentive care of professionals of documentation and archiving. In this field, the systems can be much more expensive and are judged according to the profits of productivity and the assistance which they bring to archivists, journalists and users.

A crucial problem for many professionals is the need to produce documents in many formats for various terminals from the same raw material without multiplying the editing costs. The aim of such a *repurposing* is for example to produce a DVD, a web site or an alert service by mobile phone from a TV program at the minimum cost. The basic idea is to describe the documents in such a way that they can be easily manipulated and reconfigured easily.

4.3. Textual database management

Searching in large textual corpora has already been the topic of many researches. The current stakes are the management of very large volumes of data, the possibility to answer requests relating more on concepts than on simple inclusions of words in the texts, and the characterization of sets of texts.

We work on the exploitation of scientific bibliographical bases. The explosion of the number of scientific publications makes the retrieval of relevant data for a researcher a very difficult task. The generalization of document indexing in data banks did not solve the problem. The main difficulty is to choose the keywords, which will encircle a domain of interest. The statistical method used, the factorial analysis of correspondences, makes it possible to index the documents or a whole set of documents and to provide the list of the most discriminating keywords for these documents. The index validation is carried out by searching information in a database more general than the one used to build the index and by studying the retrieved documents. That in general makes it possible to still reduce the subset of words characterizing a field.

We also explore scientific documentary corpora to solve two different problems: to index the publications with the help of meta-keys and to identify the relevant publications in a large textual database. For that, we use factorial data analysis, which allows us to find the minimal sets of relevant words that we call meta-keys and to free the bibliographical search from the problems of noise and silence. The performances of factorial correspondence analysis are sharply greater than classic search by logical equation.

VR4I Team

4. Application Domains

4.1. Panorama

The research topics of the VR4i team are related to applications of the industrial, training and education domains.

The applications to the industrial domain are very promising. For instance, the PSA Automotive Design Network, which is a new design center, groups all the tools used for automotive design, from classical CAD systems to Virtual Reality applications. The coupling of virtual reality and simulation algorithms is a key point and is the core of VR4i simulation activities. Major issues in which industrials are strongly involved are focussing on collaborative tasks between multiple users in digital mockups and for scientific visualization (ANR Part@ge and ANR Collaviz 7.1.2), tackling the challenging problem of training in Virtual Reality by providing interactive scenario languages with realists actions and reactions within the environment (GVT Project, ANR Corvette 7.1.3 and FUI SIFORAS 7.1.1). In this context, we are tackling the problem of using Virtual Reality environments for improving the ergonomics of workstations. Collaborative work is now a hot issue for facing the question of sharing expertise of distant experts for project review, for collaborative design or for analysis of data resulting from scientific computations (FP7-Infra VISIONAIR project 7.2.1) where we propose new software architectures ensuring the data distribution and the synchronization of the users (Figure 1).

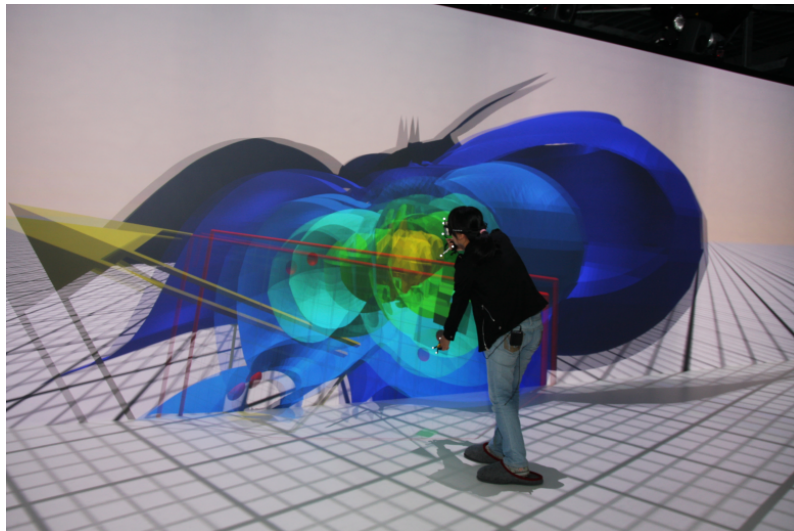


Figure 1. Collaboration between VR4i team in Immersia Room 6.4 and UCL on shared analysis of earthquake simulation within VISIONAIR project 7.2.1

WAM Project-Team

4. Application Domains

4.1. Introduction

Broadly speaking, the main application domain of our research is the web and its numerous applications. This includes the recent evolutions of the web, with a special attention paid to the mobile web, the multimedia web, and the web as a platform for applications. The goal of our research is to enable new multimedia and mobile applications that can be deployed easily on the web, taking advantage of the existing infrastructure and the latest advances in web technology.

More specifically, our work this year has focused on two main application domains: web development and pedestrian navigation.

4.2. Web Development

Current content representation practices and programming methods for the web remain severely limited. Designing web applications is becoming increasingly complex as it relies more and more on a jungle of programming languages, tools and data formats, each targeted toward the different application layers (presentation, application and storage). This often yields complex and opaque applications organized in silos, which are costly, inefficient, hard to maintain and evolve, and vulnerable to errors. In addition, the communication aspects are often handled independently via remote service invocations and represent another source of complexity and vulnerability.

Most research activities in WAM address these issues and try to cope with the fundamentals aspects of web applications (advanced content, data and communication) by studying rich document formats, data models and communication patterns, to offer correction guarantees and flexibility in the application architecture. For instance, applications need to be checked, optimized and managed as a whole while leveraging on the consistency of their individual components and data fragments.

4.3. Pedestrian Navigation

A number of factors are changing our thinking of an accessible town, namely the open data movement exemplified by OpenStreetMap, MEMS sensors embedded in mobile phones (accelerometers, magnetometers, gyroscopes), web and Augmented Reality technologies, increase in processing power of mobile phones. All these changes are allowing us to build energy efficient urban pedestrian navigation systems. These systems can now be based on Inertial Measurement Unit (IMU) and will run on mobile phones with customized embedded geographical data, a routing engine, and real time queries to urban information systems.

Richer and more precise AR mobile applications in such fields as cultural heritage visits, outdoor games, or guidance of people with disabilities will be enabled by MMG navigation, i.e. the join use of micro, macro and global navigation.

Micro-navigation builds upon embedded software ability to create a greater awareness of the immediate environment, using texture-based tracking or vision algorithms and relating this information to map and IMU data. Micro-navigation includes avoiding obstacles, locating a clear path in the proximate surroundings or at a complex crossing, finding objects and providing absolute positioning using known landmarks or beacons. Micro-navigation works at a precision level of a few centimeters by using predefined landmarks.

Macro-navigation refers to the actions required to find a route in a larger, not immediately perceptible environment, and builds upon carefully designed pedestrian ways incorporating speech instructions, audio guidance, environmental queries and IMU instructions among other things. Macro-navigation works at a precision level of one step using carefully designed routes with map-matching instructions.

Global navigation is based on an absolute global localization system like the GPS. Its precision is that of a few meters if used in a adequate geographical environment where data from external sensors is accessible. It can be used to bootstrap macro-navigation.

There is a duality relation between micro-navigation and macro-navigation. Micro-navigation is based on a localization system giving an absolute position which makes it possible to compute a relative position with respect to the planned route. Macro-navigation is based on a localization system giving a relative position which allows to compute an absolute position on the route through a process called map-matching. As a consequence, this two kinds of navigation complement and enhance each other.

WILLOW Project-Team

4. Application Domains

4.1. Introduction

We believe that foundational modeling work should be grounded in applications. This includes (but is not restricted to) the following high-impact domains.

4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering—that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. We are engaged in a project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. This effort is part of the MSR-Inria project mentioned earlier and that will be discussed further later in this report. Application of our 3D reconstruction technology is now being explored in the field of cultural heritage and archeology by the start-up Iconem, founded by Y. Ubelmann, a Willow collaborator.

4.3. Video Annotation, Interpretation, and Retrieval

Both specific and category-level object and scene recognition can be used to annotate, augment, index, and retrieve video segments in the audiovisual domain. The Video Google system developed by Sivic and Zisserman (2005) for retrieving shots containing specific objects is an early success in that area. A sample application, suggested by discussions with Institut National de l'Audiovisuel (INA) staff, is to match set photographs with actual shots in film and video archives, despite the fact that detailed timetables and/or annotations are typically not available for either medium. Automatically annotating the shots is of course also relevant for archives that may record hundreds of thousands of hours of video. Some of these applications will be pursued in our MSR-Inria project, in which INA is one of our partners.

WIMMICS Team

4. Application Domains

4.1. Introduction

A number of evolutions have changed the face of information systems in the past decade but the advent of the Web is unquestionably a major one and it is here to stay. From an initial wide-spread perception of a public documentary system, the Web as an object turned into a social virtual space and, as a technology, grew as an application design paradigm (services, data formats, query languages, scripting, interfaces, reasoning, etc.). The universal deployment and support of its standards led the Web to take over nearly all of our information systems. As the Web continues to evolve, our information systems are evolving with it.

Today in organizations, not only almost every internal information system is a Web application, but these applications also more and more often interact with external Web applications. The complexity and coupling of these Web-based information systems call for specification methods and engineering tools. From capturing the needs of users to deploying a usable solution, there are many steps involving computer science specialists and non-specialists.

We defend the idea of relying on Semantic Web formalisms to capture and reason on the models of these information systems supporting the design, evolution, interoperability and reuse of the models and their data as well as the workflows and the processing.

4.2. Linked Data on the Web and on Intranets

With billions of triples online (see Linked Open Data initiative), the Semantic Web is providing and linking open data at a growing pace and publishing and interlinking the semantics of their schemas. Information systems can now tap into and contribute to this Web of data, pulling and integrating data on demand. Many organisations also started to use this approach on their intranets leading to what is called linked enterprise data.

A first application domain for us is the publication and linking of data and their schemas through Web architectures. Our results provide software platforms to publish and query data and their schemas, to enrich these data in particular by reasoning on their schemas, to control their access and licences, to assist the workflows that exploit them, to support the use of distributed datasets, to assist the browsing and visualization of data, etc.

Examples of collaboration and applied projects include: Corese/KGRAM, Datalift, DBpedia, ALU/BLF Convention, ADT SeGViz.

4.3. Assisting Web-based Epistemic Communities

In parallel to linked open data on the Web, social Web applications also spread virally (e.g. Facebook growing toward 800 million users) first giving the Web back its status of a social read-write media and then leading it to its full potential of a virtual place where to act, react and interact. In addition, many organizations are now considering deploying social Web applications internally to foster community building, expert cartography, business intelligence, technological watch and knowledge sharing in general.

Reasoning on the linked data and the semantics of the schemas used to represent social structures and Web resources, we intend to provide applications supporting communities of practice and interest and fostering their interactions.

We use typed graphs to capture and mix: social networks with the kinds of relationships and the descriptions of the persons; compositions of Web services with types of inputs and outputs; links between documents with their genre and topics; hierarchies of classes, thesauri, ontologies and folksonomies; recorded traces and suggested navigation courses; submitted queries and detected frequent patterns; timelines and workflows; etc.

Our results assist epistemic communities in their daily activities such as biologists exchanging results, business intelligence and technological watch networks informing companies, engineers interacting on a project, conference attendees, students following the same course, tourists visiting a region, mobile experts on the field, etc. Examples of collaboration and applied projects include: Kolflow, OCKTOPUS, ISICIL, SAP Convention.

ZENITH Project-Team

4. Application Domains

4.1. Data-intensive Scientific Applications

The application domains covered by Zenith are very wide and diverse, as they concern data-intensive scientific applications, i.e. most scientific applications. Since the interaction with scientists is crucial to identify and tackle data management problems, we are dealing primarily with application domains for which Montpellier has an excellent track record, i.e. agronomy, environmental science, life science, with scientific partners like INRA, IRD, CIRAD and IRSTEA (prev. CEMAGREF). However, we are also addressing other scientific domains (e.g. astronomy, oil extraction) through our international collaborations (e.g. in Brazil).

Let us briefly illustrate some representative examples of scientific applications on which we have been working on.

- **Management of astronomical catalogs.** An example of data-intensive scientific applications is the management of astronomical catalogs generated by the Dark Energy Survey (DES) project on which we are collaborating with researchers from Brazil. In this project, huge tables with billions of tuples and hundreds of attributes (corresponding to dimensions, mainly double precision real numbers) store the collected sky data. Data are appended to the catalog database as new observations are performed and the resulting database size is estimated to reach 100TB very soon. Scientists around the globe can query the database with queries that may contain a considerable number of attributes. The volume of data that this application holds poses important challenges for data management. In particular, efficient solutions are needed to partition and distribute the data in several servers. An efficient partitioning scheme should try to minimize the number of fragments accessed in the execution of a query, thus reducing the overhead associated to handle the distributed execution.
- **Pesticide reduction.** In a pesticide reduction application, with CEMAGREF, we plan to work on sensor data for plant monitoring. Sensors are used to observe the development of diseases and insect attacks in the agricultural farms, aiming at using pesticides only when necessary. The sensors periodically send to a central system their data about different measures such as plants contamination, temperature or moisture level. A decision support system analyzes the sent data, and triggers a pesticide treatment only when needed. However, the data sent by sensors are not entirely certain. The main reasons for uncertainty are the effect of climate events on sensors, e.g. rain, unreliability of the data transmission media, fault in sensors, etc. This requires to deal with uncertain data in modeling and querying to be used for data analysis and data mining.
- **Botanical data sharing.** Botanical data is highly decentralized and heterogeneous. Each actor has its own expertise domain, hosts its own data, and describes them in a specific format. Furthermore, botanical data is complex. A single plant's observation might include many structured and unstructured tags, several images of different organs, some empirical measurements and a few other contextual data (time, location, author, etc.). A noticeable consequence is that simply identifying plant species is often a very difficult task; even for the botanists themselves (the so-called taxonomic gap). Botanical data sharing should thus speed up the integration of raw observation data, while providing users an easy and efficient access to integrated data. This requires to deal with social-based data integration and sharing, massive data analysis and scalable content-based information retrieval. We address this application in the context of the French initiative PI@ntNet, with CIRAD and IRD.
- **Deepwater oil exploitation.** An important step in oil exploitation is pumping oil from ultra-deepwater from thousand meters up to the surface through long tubular structures, called risers. Maintaining and repairing risers under deep water is difficult, costly and critical for the environment. Thus, scientists must predict risers fatigue based on complex scientific models and observed data for the risers. Risers fatigue analysis requires a complex workflow of data-intensive activities which

may take a very long time to compute. A typical workflow takes as input files containing riser information, such as finite element meshes, winds, waves and sea currents, and produces result analysis files to be further studied by the scientists. It can have thousands of input and output files and tens of activities (e.g. dynamic analysis of risers movements, tension analysis, etc.). Some activities, e.g. dynamic analysis, are repeated for many different input files, and depending on the mesh refinements, each single execution may take hours to complete. To speed up risers fatigue analysis requires parallelizing workflow execution, which is hard to do with existing SWfMS. We address this application in collaboration with UFRJ, and Petrobras.

These application examples illustrate the diversity of requirements and issues which we are addressing with our scientific application partners (CIRAD, INRA, CEMAGREF, etc.). To further validate our solutions and extend the scope of our results, we also want to foster industrial collaborations, even in non scientific applications, provided that they exhibit similar challenges.