



RESEARCH CENTER

FIELD

Activity Report 2014

Section Application Domains

Edition: 2015-03-24

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

| | |
|--|----|
| 1. ALF Project-Team | 9 |
| 2. ANTIQUE Team | 10 |
| 3. AOSTE Project-Team | 12 |
| 4. ARIC Project-Team | 14 |
| 5. ATEAMS Project-Team (section vide) | 15 |
| 6. CAIRN Project-Team | 16 |
| 7. CAMUS Team | 17 |
| 8. CAMEL Project-Team | 18 |
| 9. CARTE Project-Team | 20 |
| 10. CASCADE Project-Team | 22 |
| 11. CASSIS Project-Team | 24 |
| 12. CELTIQUE Project-Team (section vide) | 26 |
| 13. COMETE Project-Team | 27 |
| 14. COMPSYS Project-Team | 28 |
| 15. CONVECS Project-Team | 29 |
| 16. CRYPT Team (section vide) | 30 |
| 17. DEDUCTEAM Exploratory Action | 31 |
| 18. DICE Team (section vide) | 32 |
| 19. DREAMPAL Team (section vide) | 33 |
| 20. ESTASYS Exploratory Action | 34 |
| 21. GALAAD2 Team | 35 |
| 22. GALLIUM Project-Team | 36 |
| 23. GCG Team | 38 |
| 24. GEOMETRICA Project-Team | 39 |
| 25. GRACE Project-Team | 40 |
| 26. HYCOMES Team | 41 |
| 27. LFANT Project-Team (section vide) | 42 |
| 28. MARELLE Project-Team | 43 |
| 29. MEXICO Project-Team | 44 |
| 30. MUTANT Project-Team | 46 |
| 31. PAREO Project-Team | 48 |
| 32. PARKAS Project-Team | 49 |
| 33. PARSIFAL Project-Team | 50 |
| 34. PI.R2 Project-Team (section vide) | 52 |
| 35. POLSYS Project-Team | 53 |
| 36. POSTALE Team (section vide) | 54 |
| 37. PRIVATICS Project-Team | 55 |
| 38. PROSECCO Project-Team | 57 |
| 39. SECRET Project-Team | 58 |
| 40. SPADES Team | 59 |

| | |
|--------------------------|----|
| 41. SPECFUN Project-Team | 60 |
| 42. SUMO Project-Team | 61 |
| 43. TASC Project-Team | 63 |
| 44. TEA Project-Team | 64 |
| 45. TEMPO Team | 65 |
| 46. TOCCATA Project-Team | 66 |
| 47. VEGAS Project-Team | 67 |
| 48. VERIDIS Project-Team | 68 |

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

| | |
|---|-----|
| 49. APICS Project-Team | 69 |
| 50. ASPI Project-Team | 74 |
| 51. BACCHUS Team (section vide) | 75 |
| 52. BIPOP Project-Team | 76 |
| 53. CAGIRE Team | 78 |
| 54. CLASSIC Project-Team (section vide) | 80 |
| 55. COMMANDS Project-Team | 81 |
| 56. CORIDA Team | 82 |
| 57. CQFD Project-Team | 84 |
| 58. DEFI Project-Team | 85 |
| 59. DISCO Project-Team | 88 |
| 60. DOLPHIN Project-Team | 89 |
| 61. ECUADOR Project-Team | 91 |
| 62. GAMMA3 Project-Team (section vide) | 94 |
| 63. GECO Project-Team | 95 |
| 64. GEOSTAT Project-Team | 99 |
| 65. I4S Project-Team | 100 |
| 66. IPSO Project-Team | 102 |
| 67. MATHERIALS Team | 104 |
| 68. MATHRISK Project-Team | 107 |
| 69. Maxplus Project-Team | 108 |
| 70. MC2 Team | 112 |
| 71. MCTAO Project-Team | 114 |
| 72. MEPHYSTO Team | 117 |
| 73. MISTIS Project-Team | 119 |
| 74. MODAL Project-Team | 120 |
| 75. MOKAPLAN Team | 121 |
| 76. NACHOS Project-Team | 124 |
| 77. NANO-D Project-Team | 130 |
| 78. NECS Project-Team | 134 |
| 79. NON-A Project-Team | 136 |
| 80. OPALE Project-Team | 139 |

| | |
|--|-----|
| 81. POEMS Project-Team | 141 |
| 82. QUANTIC Team | 142 |
| 83. REALOPT Project-Team | 143 |
| 84. REGULARITY Project-Team | 147 |
| 85. SELECT Project-Team | 149 |
| 86. SEQUEL Project-Team | 151 |
| 87. SIERRA Project-Team (section vide) | 154 |
| 88. TAO Project-Team | 155 |
| 89. TOSCA Project-Team | 156 |

DIGITAL HEALTH, BIOLOGY AND EARTH

| | |
|---|-----|
| 90. ABS Project-Team (section vide) | 159 |
| 91. AMIB Project-Team (section vide) | 160 |
| 92. ANGE Project-Team | 161 |
| 93. ARAMIS Project-Team | 163 |
| 94. ASCLEPIOS Project-Team (section vide) | 165 |
| 95. ATHENA Project-Team | 166 |
| 96. BAMBOO Project-Team | 168 |
| 97. BEAGLE Project-Team (section vide) | 169 |
| 98. BIGS Project-Team (section vide) | 170 |
| 99. BIOCORE Project-Team | 171 |
| 100. BONSAI Project-Team | 174 |
| 101. CARMEN Team | 175 |
| 102. CASTOR Project-Team | 176 |
| 103. CLIME Project-Team | 177 |
| 104. COFFEE Project-Team | 179 |
| 105. DEMAR Project-Team (section vide) | 181 |
| 106. DRACULA Project-Team | 182 |
| 107. DYLISS Project-Team | 187 |
| 108. FLUMINANCE Project-Team | 189 |
| 109. GALEN Project-Team | 190 |
| 110. GENSCALE Project-Team | 191 |
| 111. IBIS Project-Team (section vide) | 192 |
| 112. KALIFFE Project-Team | 193 |
| 113. LEMON Team | 196 |
| 114. LIFEWARE Team | 197 |
| 115. M3DISIM Team | 198 |
| 116. MAGIQUE-3D Project-Team | 199 |
| 117. MAGNOME Project-Team | 201 |
| 118. MAMBA Team | 203 |
| 119. MASAIE Project-Team | 206 |
| 120. MNEMOSYNE Project-Team | 207 |

| | |
|---|-----|
| 121. MODEMIC Project-Team | 208 |
| 122. MOISE Project-Team | 209 |
| 123. MORPHEME Project-Team (section vide) | 212 |
| 124. MYCENAE Project-Team | 213 |
| 125. NEUROMATHCOMP Project-Team (section vide) | 215 |
| 126. NEUROSYS Team | 216 |
| 127. NUMED Project-Team (section vide) | 218 |
| 128. PARIETAL Project-Team | 219 |
| 129. POMDAPI Project-Team | 221 |
| 130. POPIX Team | 222 |
| 131. REO Project-Team | 225 |
| 132. SAGE Project-Team | 227 |
| 133. SERPICO Project-Team | 228 |
| 134. SHACRA Project-Team (section vide) | 230 |
| 135. SISTM Team | 231 |
| 136. SISYPHE Project-Team | 232 |
| 137. STEEP Team | 234 |
| 138. TONUS Team | 238 |
| 139. VIRTUAL PLANTS Project-Team (section vide) | 240 |
| 140. VISAGES Project-Team | 241 |

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

| | |
|---|-----|
| 141. ALGORILLE Project-Team | 243 |
| 142. ALPINES Project-Team | 244 |
| 143. ASAP Project-Team | 246 |
| 144. ASCOLA Project-Team | 247 |
| 145. ATLANMOD Project-Team | 249 |
| 146. AVALON Project-Team | 250 |
| 147. CIDRE Project-Team | 252 |
| 148. COAST Team (section vide) | 253 |
| 149. COATI Project-Team | 254 |
| 150. CTRL-A Exploratory Action | 255 |
| 151. DANTE Team | 257 |
| 152. DIANA Team (section vide) | 259 |
| 153. DIONYSOS Project-Team (section vide) | 260 |
| 154. DIVERSE Project-Team | 261 |
| 155. DYOGENE Project-Team | 262 |
| 156. FOCUS Project-Team | 263 |
| 157. FUN Project-Team (section vide) | 264 |
| 158. GANG Project-Team | 265 |
| 159. HIEPACS Project-Team | 266 |
| 160. HIPERCOM2 Team | 269 |

| | |
|--|-----|
| 161. INDES Project-Team | 271 |
| 162. INFINE Team | 272 |
| 163. KerData Project-Team | 274 |
| 164. MADYNES Project-Team | 276 |
| 165. MAESTRO Project-Team | 277 |
| 166. MESCAL Project-Team | 278 |
| 167. MIMOVE Team | 279 |
| 168. MOAIS Project-Team (section vide) | 281 |
| 169. MUSE Team | 282 |
| 170. MYRIADS Project-Team (section vide) | 285 |
| 171. PHOENIX Project-Team | 286 |
| 172. RAP Project-Team (section vide) | 288 |
| 173. REGAL Project-Team (section vide) | 289 |
| 174. RMOD Project-Team | 290 |
| 175. ROMA Team | 291 |
| 176. RUNTIME Team | 292 |
| 177. SCALE Team | 293 |
| 178. SOCRATE Project-Team | 295 |
| 179. SPIRALS Team | 298 |
| 180. TACOMA Team | 299 |
| 181. TYREX Project-Team | 301 |
| 182. URBANET Team | 302 |
| 183. WHISPER Team | 304 |

PERCEPTION, COGNITION AND INTERACTION

| | |
|-----------------------------|-----|
| 184. ALICE Project-Team | 305 |
| 185. ALPAGE Project-Team | 306 |
| 186. AVIZ Project-Team | 309 |
| 187. AYIN Team | 310 |
| 188. DAHU Project-Team | 311 |
| 189. DREAM Project-Team | 312 |
| 190. E-MOTION Project-Team | 315 |
| 191. EXMO Project-Team | 316 |
| 192. FLOWERS Project-Team | 317 |
| 193. GRAPHIK Project-Team | 318 |
| 194. HEPHAISTOS Team | 319 |
| 195. HYBRID Project-Team | 320 |
| 196. IMAGINE Project-Team | 321 |
| 197. IN-SITU Project-Team | 322 |
| 198. LAGADIC Project-Team | 323 |
| 199. LEAR Project-Team | 324 |
| 200. LINKMEDIA Project-Team | 325 |

| | |
|---|-----|
| 201. LINKS Team | 326 |
| 202. MAGNET Team | 327 |
| 203. MAGRIT Project-Team | 328 |
| 204. MAIA Project-Team | 329 |
| 205. MANAO Project-Team (section vide) | 330 |
| 206. MAVERICK Project-Team (section vide) | 331 |
| 207. MIMETIC Project-Team | 332 |
| 208. MINT Project-Team | 336 |
| 209. MORPHEO Project-Team | 337 |
| 210. MULTISPEECH Team | 338 |
| 211. OAK Project-Team | 340 |
| 212. ORPAILLEUR Project-Team | 341 |
| 213. PANAMA Project-Team | 343 |
| 214. PERCEPTION Project-Team (section vide) | 345 |
| 215. POTIOC Project-Team | 346 |
| 216. PRIMA Project-Team (section vide) | 347 |
| 217. REVES Project-Team | 348 |
| 218. RITS Team | 349 |
| 219. SEMAGRAMME Project-Team | 351 |
| 220. SIROCCO Project-Team | 352 |
| 221. SMIS Project-Team | 354 |
| 222. STARS Project-Team | 355 |
| 223. TITANE Project-Team | 357 |
| 224. WILLOW Project-Team | 358 |
| 225. WIMMICS Project-Team | 359 |
| 226. ZENITH Project-Team | 361 |

ALF Project-Team

4. Application Domains

4.1. Any computer usage

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.

ANTIQUE Team

4. Application Domains

4.1. Verification of safety critical embedded software

The verification of safety critical embedded software is a very important application domain for our group. First, this field requires a high confidence in software, as a bug may cause disastrous events. Thus, it offers an obvious opportunity for a strong impact. Second, such software usually have better specifications and a better design than many other families of software, hence are an easier target for developing new static analysis techniques (which can later be extended for more general, harder to cope with families of programs). This includes avionics, automotive and other transportation systems, medical systems...

For instance, the verification of avionics systems represent a very high percentage of the cost of an airplane (about 30 % of the overall airplane design cost). The state of the art development processes mainly resort to testing in order to improve the quality of software. Depending on the level of criticality of a software (at highest levels, any software failure would endanger the flight) a set of software requirements are checked with test suites. This approach is both costly (due to the sheer amount of testing that needs to be performed) and unsound (as errors may go unnoticed, if they do not arise on the test suite).

By contrast, static analysis can ensure higher software quality at a lower cost. Indeed, a static analyzer will catch all bugs of a certain kind. Moreover, a static analysis run typically lasts a few hours, and can be integrated in the development cycle in a seamless manner. For instance, **ASTRÉE** successfully verified the absence of runtime error in several families of safety critical fly-by-wire avionic software, in at most a day of computation, on standard hardware. Other kinds of synchronous embedded software have also been analyzed with good results.

In the future, we plan to greatly extend this work so as to verify *other families of embedded software* (such as communication, navigation and monitoring software) and *other families of properties* (such as security and liveness properties).

Embedded software in charge of communication, navigation, monitoring typically rely on a *parallel* structure, where several threads are executed in parallel, and manage different features (input, output, user interface, internal computation, logging...). This structure is also often found in automotive software. An even more complex case is that of *distributed* systems, where several separate computers are run in parallel and take care of several sub-tasks of a same feature, such as braking. Such a logical structure is not only more complex than the synchronous one, but it also introduces new risks and new families of errors (deadlocks, data-races...). Moreover, such less well designed, and more complex embedded software often utilizes more complex data-structures than synchronous programs (which typically only use arrays to store previous states) and may use dynamic memory allocation, or build dynamic structures inside static memory regions, which are actually even harder to verify than conventional dynamically allocated data structures. Complex data-structures also introduce new kinds of risks (the failure to maintain structural invariants may lead to runtime errors, non termination, or other software failures). To verify such programs, we will design additional abstract domains, and develop new static analysis techniques, in order to support the analysis of more complex programming language features such as parallel and concurrent programming with threads and manipulations of complex data structures. Due to their size and complexity, the verification of such families of embedded software is a major challenge for the research community.

Furthermore, embedded systems also give rise to novel security concerns. It is in particular the case for some aircraft-embedded computer systems, which communicate with the ground through untrusted communication media. Besides, the increasing demand for new capabilities, such as enhanced on-board connectivity, e.g. using mobile devices, together with the need for cost reduction, leads to more integrated and interconnected systems. For instance, modern aircrafts embed a large number of computer systems, from safety-critical cockpit avionics to passenger entertainment. Some systems meet both safety and security requirements.

Despite thorough segregation of subsystems and networks, some shared communication resources raise the concern of possible intrusions. Because of the size of such systems, and considering that they are evolving entities, the only economically viable alternative is to perform automatic analyses. Such analyses of security and confidentiality properties have never been achieved on large-scale systems where security properties interact with other software properties, and even the mapping between high-level models of the systems and the large software base implementing them has never been done and represents a great challenge. Our goal is to prove empirically that the security of such large scale systems can be proved formally, thanks to the design of dedicated abstract interpreters.

The long term goal is to make static analysis more widely applicable to the verification of industrial software.

4.2. Static analysis of software components and libraries

An important goal of our work is to make static analysis techniques easier to apply to wider families of software. Then, in the longer term, we hope to be able to verify less critical, yet very commonly used pieces of software. Those are typically harder to analyze than critical software, as their development process tends to be less rigorous. In particular, we will target operating systems components and libraries. As of today, the verification of such programs is considered a major challenge to the static analysis community.

As an example, most programming languages offer Application Programming Interfaces (API) providing ready-to-use abstract data structures (e.g., sets, maps, stacks, queues, etc.). These APIs, are known under the name of containers or collections, and provide off-the-shelf libraries of high level operations, such as insertion, deletion and membership checks. These container libraries give software developers a way of abstracting from low-level implementation details related to memory management, such as dynamic allocation, deletion and pointer handling or concurrency aspects, such as thread synchronization. Libraries implementing data structures are important building bricks of a huge number of applications, therefore their verification is paramount. We are interested in developing static analysis techniques that will prove automatically the correctness of large audience libraries such as Glib and Threading Building Blocks.

4.3. Biological systems

Computer Science takes a more and more important role in the design and the understanding of biological systems such as signaling pathways, self assembly systems, DNA repair mechanisms. Biology has gathered large data-bases of facts about mechanistic interactions between proteins, but struggles to draw an overall picture of how these systems work as a whole. High level languages designed in Computer Science allow to collect these interactions in integrative models, and provide formal definitions (i.e., semantics) for the behavior of these models. This way, modelers can encode their knowledge, following a bottom-up discipline, without simplifying *a priori* the models at the risk of damaging the key properties of the system. Yet, the systems that are obtained this way suffer from combinatorial explosion (in particular, in the number of different kinds of molecular components, which can arise at run-time), which prevents from a naive computation of their behavior.

We develop various abstract interpretation-based analyses, tailored to different phases of the modeling process. We propose automatic static analyses in order to detect inconsistencies in the early phases of the modeling process. These analyses are similar to the analysis of classical safety properties of programs. They involve both forward and backward reachability analyses as well as causality analyses, and can be tuned at different levels of abstraction. We also develop automatic static analyses so as to identify the key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify the models. The correctness of this simplification process is proved by the means of abstract interpretation: this ensures formally that the simplification preserves the quantitative properties that have been specified beforehand by the modeler. The whole pipeline is parameterized by a large choice of abstract domains which exploits different features of the high level description of models.

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 (NoC) 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. New parallel architecture paradigms, such as GPU co-processors or Massively Parallel Processor Arrays (MPPA) form natural targets as NoC-based platforms.

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 Project-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;
- cryptography.

4.2. Floating-point and Validated Numerics

Our expertise on validated numerics is useful to analyze and improve, and guarantee the quality of numerical results in a wide range of applications including:

- scientific simulation;
- global optimization;
- 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.

Another interesting field of application 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 take advantage of information technology advances. While wireless transmissions allow a car to communicate with another one 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. It is noteworthy that analysis documents from governmental agencies (see e.g., [31]) use cryptanalysis results as their key material.

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 [4]. 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.

Important objects related to the structure of genus-2 curves are the isogenies between their Jacobians. Computing such isogenies is a key point in understanding important underlying objects such as the endomorphism ring, and can be useful in various situations, including for cryptographic or cryptanalytic applications. The team has produced important results in this context [6], [2].

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 (as was done by the team by factoring RSA-768 [5]) and discrete-logarithm computations (as was done by the team in 2013 for the field $\text{GF}(2^{809})$ [15]). 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. To this regard the breakthrough provided by the new quasi-polynomial discrete logarithm [17] is of course of utmost importance.

4.2. 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.

4.2.1. 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 — several 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.

4.2.2. 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.

4.2.3. 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 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.

4.3. Standardization

4.3.1. 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.

CARTE Project-Team

4. Application Domains

4.1. Computer Virology

4.1.1. *The theoretical track.*

It is rightful to wonder why there are 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 [45] in order to understand the limits of this method. The second one consists in analysing the behavior of a program by monitoring it. Following [47], 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 [67].

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.

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., [29]), control theory (see e.g., [37]), neural networks (see e.g., [68]), and so on. We are interested in the formal decidability of properties of dynamical systems, such as reachability [58], the Skolem-Pisot problem [33], the computability of the ω -limit set [57]. 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, e.g., 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 [69], on recursive analysis [74], on the algebraic approach [65] and on Markov computability [60]. 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, used to describe dynamic systems, in particular in the adversary context. As they were initially developed in the context of automated deduction, the rewriting proof techniques, although now numerous, are not yet adapted to the complex framework of modelization and programming. An important stake in the domain is then 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 [48], [49], [50], to weak termination [51], sufficient completeness [53] and probabilistic termination [55]. The last three results take place in the context of adversary computations, since they allow for proving that even a divergent program, in the sense where it does not terminate, can give the expected results. A common mechanism has been extracted from the above works, providing a generic inductive proof framework for properties of reduction relations, which can be parametrized by the property to be proved [54], [56]. Provided 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 years [62], [63], [64]. 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. Privacy for the Cloud

Many companies have already started the migration to the Cloud and many individuals share their personal informations on social networks. While some of the data are public information, many of them are personal and even quite sensitive. Unfortunately, the current access mode is purely right-based: the provider first authenticates the client, and grants him access, or not, according to his rights in the access-control list. Therefore, the provider itself not only has total access to the data, but also knows which data are accessed, by whom, and how: privacy, which includes secrecy of data (confidentiality), identities (anonymity), and requests (obliviousness), should be enforced. Moreover, while high availability can easily be controlled, and thus any defect can immediately be detected, failures in privacy protection can remain hidden for a long time. The industry of the Cloud introduces a new implicit trust requirement: nobody has any idea at all of where and how his data are stored and manipulated, but everybody should blindly trust the providers. The providers will definitely do their best, but this is not enough. Privacy-compliant procedures cannot be left to the responsibility of the provider: however strong the trustfulness of the provider may be, any system or human vulnerability can be exploited against privacy. This presents too huge a threat to tolerate. *The distribution of the data and the secrecy of the actions must be given back to the users. It requires promoting privacy as a global security notion.*

In order to protect the data, one needs to encrypt it. Unfortunately, traditional encryption systems are inadequate for most applications involving big, complex data. Recall that in traditional public key encryption, a party encrypts data to a single known user, which lacks the expressiveness needed for more advanced data sharing. In enterprise settings, a party will want to share data with groups of users based on their credentials. Similarly, individuals want to selectively grant access to their personal data on social networks as well as documents and spreadsheets on Google Docs. Moreover, the access policy may even refer to users who do not exist in the system at the time the data is encrypted. Solving this problem requires an entirely new way of encrypting data.

A first natural approach would be **fully homomorphic encryption** (FHE, see above), but a second one is also **functional encryption**, that is an emerging paradigm for public-key encryption: it enables more fine-grained access control to encrypted data, for instance, the ability to specify a decryption policy in the ciphertext so that only individuals who satisfy the policy can decrypt, or the ability to associate keywords to a secret key so that it can only decrypt documents containing the keyword. Our work on functional encryption centers around two goals:

1. to obtain more efficient pairings-based functional encryption;
2. and to realize new functionalities and more expressive functional encryption schemes.

Another approach is **secure multi-party computation protocols**, where interactivity might provide privacy in a more efficient way. Recent implicit interactive proofs of knowledge can be a starting point. But stronger properties are first expected for improving privacy. They can also be integrated into new ad-hoc broadcast systems, in order to distribute the management among several parties, and eventually remove any trust requirements.

Strong privacy for the Cloud would have a huge societal impact since it would revolutionize the trust model: users would be able to make safe use of outsourced storage, namely for personal, financial and medical data, without having to worry about failures or attacks of the server.

4.2. Hardware Security

Cryptography is only one component of information security, but it is a crucial component. Without cryptography, it would be impossible to establish secure communications between users over insecure networks like the Internet. In particular, public-key cryptography (invented by Diffie and Hellman in 1976) enables to establish secure communications between users who have never met physically before. One can argue that companies like E-Bay or Amazon could not exist without public-key cryptography. Since 30 years the theory of cryptography has developed considerably. However cryptography is not only a theoretical science; namely at some point the cryptographic algorithms must be implemented on physical devices, such as PCs, smart cards or RFIDs. Then problems arise: in general smart cards and RFIDs have limited computing power and leak information through power consumption and electro-magnetic radiations. Similarly a PC can be exposed to various computer viruses which can leak private informations to a remote attacker. Such information leakage can be exploited by an attacker; this is called a **side-channel attack**. It is well known that a cryptographic algorithm which is perfectly secure in theory can be completely insecure in practice if improperly implemented.

In general, countermeasures against side-channel attacks are heuristic and can only make a particular implementation resist particular attacks. Instead of relying on ad-hoc security patches, a better approach consists in working in the framework of **provable security**. The goal is to prove that a cryptosystem does not only resist specific attacks but can resist any possible side-channel attack. As already demonstrated with cryptographic protocols, this approach has the potential to significantly increase the security level of cryptographic products. Recently the cryptography research community has developed new security models to take into account these practical implementation attacks; the most promising such model is called the **leakage-resilient model**.

Therefore, our goal is to define new security models that take into account any possible side-channel attack, and then to design new cryptographic schemes and countermeasures with a proven security guarantee against side-channel attacks.

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 and AVANTSSAR platforms.

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 [75] and Java Card Virtual Machine Transaction mechanism [77]), information system and for embedded software [85].

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., [81]): 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 adapt the method to 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: Nicolas Bordenabe, Konstantinos Chatzikokolakis, Catuscia Palamidessi, Marco Stronati.

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 processors, hardware accelerators (programmable or not), embedded systems, and those (the same plus other) that 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.

4.2. Users of HPC Platforms and Scientific Computing

The convergence between embedded computing systems and high-performance computing (HPC) technologies offers new computing platforms and tools for the users of scientific computing (e.g., people working in numerical analysis, in simulation, modeling, etc.). The proliferation of “cheap” hardware accelerators and multicores makes the “small HPC” (as opposed to computing centers with more powerful computers, grid computing, and exascale computing) accessible to a larger number of users, even though it is still difficult to exploit, due to the complexity of parallel programming, code tuning, interaction with compilers, which result from the multiple levels of parallelism and of memories in the recent architectures. The link between compiler and code optimization research (as in Compsys) and such users are still to be reinforced, both to guarantee the relevance of compiler research efforts with respect to application needs, and to help users better interact with compiler choices and understand performance issues.

The support of Labex MILYON (through its thematic quarters, such as the thematic quarter on compilation we organized in 2013 ⁰, or the upcoming 2016 thematic quarter on high-performance computing) and the activities of the LyonCalcul initiative ⁰ are means to get closer to users of scientific computing, even if it is too early to know if Compsys will indeed be directly helpful to them.

⁰Thematic quarter on compilation: <http://labexcompilation.ens-lyon.fr/>

⁰Lyon Calcul federation: <http://lyoncalcul.univ-lyon1.fr>

CONVECS Project-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,
- *Component-based systems*: Web services, peer-to-peer networks,
- *Databases*: transaction protocols, distributed knowledge bases, stock management,
- *Distributed systems*: virtual shared memory, dynamic reconfiguration algorithms, fault tolerance algorithms, cloud computing,
- *Embedded systems*: air traffic control, avionic systems, medical devices,
- *Hardware architectures*: multiprocessor architectures, systems on chip, 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.

CRYPT Team (section vide)

DEDUCTEAM Exploratory Action

4. Application Domains

4.1. 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.

4.2. B-set theory

Set theory appears to be an appropriate theory for automated theorem provers based on Deduction modulo, in particular the several extensions of *Zenon* (*Super Zenon* and *Zenon Modulo*). Modeling techniques using set theory are therefore good candidates to assess these tools. This is what we have done with the *B* method whose formalism relies on set theory. A collaboration with *Siemens* has been developed to automatically verify the *B* proof rules of *Atelier B* [42]. From this work presented in the Doctoral dissertation of Mélanie Jacquél, the *Super Zenon* tool [13] [35] has been designed in order to be able to reason modulo the *B* set theory. As a sequel of this work, we contribute to the *BWare* project whose aim is to provide a mechanized framework to support the automated verification of *B* proof obligations coming from the development of industrial applications. In this context, we have recently designed *Zenon Modulo* [33], [34] (Pierre Halmagrand's PhD thesis, which has started on October 2013) to deal with the *B* set theory. In this work, the idea is to manually transform the *B* set theory into a theory modulo and provide it to *Zenon Modulo* in order to verify the proof obligations of the *BWare* project.

4.3. Termination certificate verification

Termination is an important property to verify, especially in critical applications. Automated termination provers use more and more complex theoretical results and external tools (e.g. sophisticated SAT solvers) that make their results not fully trustable and very difficult to check. To overcome this problem, a language for termination certificates, called **CPF**, has been developed since several years now. Deducteam develops a formally certified tool, **Rainbow**, based on the Coq library **CoLoR**, that is able to automatically verify the correctness of such termination certificates.

DICE Team (section vide)

DREAMPAL Team (section vide)

ESTASYS Exploratory Action

4. Application Domains

4.1. Application Domains

In two years, ESTASYS should lead to the creation of a top class research team at Inria as well as to an interdisciplinary community of researchers and practitioners at the world level.

ESTASYS sets the foundations for an engineering domain dedicated to SoS that will benefit the European software industry. This is achieved by creating mathematical models that capture the computational power, autonomous decisions and complex stochastic and real-time dynamics of SoS. ESTASYS produces new decidability and complexity results, simulation-based techniques, and algorithms with correctness arguments. All aim at efficient reasoning about SoS and are traced back to case studies. **Our strategy to work in close collaboration with contact in industry will guarantee their wider adoption by the european software industry.**

In the near future, The ESTASYS-PLASMA toolset will be distributed as open source whenever possible, but will create a new market of tools for SoS.

GALAAD2 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. Optimizing 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 development 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 a compact geometric model. We are working in collaboration with groups in agronomy on the problem 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 (B-spline) 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, program proof, 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 [62] and enforcement of data confidentiality through type-based inference of information flows and noninterference properties [67].

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. 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, the USA, and Japan.

GCG Team

4. Application Domains

4.1. Transfer

The main industrial sector related to the research activities of GCG is the one of semi-conductor (programmable architectures spanning from embedded systems to servers). Obviously any computing application which has the objective of exploiting as much as possible the resources (in terms of high-performance but also low energy consumption) of the host architecture is intended to take advantage of advances in compiler and runtime technology. These applications are based over numerical kernels (linear algebra, FFT, convolution...) that can be adapted on a large spectrum of architectures. Members of GCG already maintain fruitful and strong collaborations with several companies such as STMicroelectronics, Kalray, plus a recent (not yet formal) collaboration with Intel.

GEOMETRICA Project-Team

4. Application Domains

4.1. Application Domains

- Medical Imaging
- Numerical simulation
- Geometric modeling
- Geographic information systems
- Visualization
- Data analysis
- Astrophysics
- Material physics

GRACE Project-Team

4. Application Domains

4.1. Cryptography and Cryptanalysis

In the twenty-first century, cryptography plays two essential roles: it is used to ensure *security* and *integrity* of communications and communicating entities. Contemporary cryptographic techniques can be used to hide private data, and to prove that public data has not been modified; to provide anonymity, and to assert and prove public identities. The creation and testing of practical cryptosystems involves

1. The design of provably secure protocols;
2. The design and analysis of compact and efficient algorithms to implement those protocols, and to attack their underlying mathematical and computational problems;
3. The robust implementation of those algorithms in low-level software and hardware, and their deployment in the wild.

While these layers are interdependent, GRACE’s cryptographic research is focused heavily on the middle layer: we design, implement, and analyze the most efficient algorithms for fundamental tasks in contemporary cryptography. Our “clients”, in a sense, are protocol designers on the one hand, and software and hardware engineers on the other.

F. Morain and B. Smith work primarily on the number-theoretic algorithms that underpin the current state-of-the-art in public-key cryptography (which is used to establish secure connections, and create and verify digital signatures, among other applications). For example, their participation in the ANR CATREL project aims to give a realistic assessment of the security of systems based on the Discrete Logarithm Problem, by creating a free, open, algorithmic package implementing the fastest known algorithms for attacking DLP instances. This will have an extremely important impact on contemporary pairing-based cryptosystems, as well as legacy finite field-based cryptosystems. On a more constructive note, F. Morain’ elliptic curve point counting and primality proving algorithms are essential tools in the everyday construction of strong public-key cryptosystems, while B. Smith’s recent work on elliptic curves aims to improve the speed of curve-based cryptosystems (such as Elliptic Curve Diffie–Hellman key exchange, a crucial step in establishing secure internet connections) without compromising their security.

D. Augot, F. Levy-dit-Vehel, and A. Couvreur’s research on codes has far-reaching applications in *code-based cryptography*. This is a field which is growing rapidly in importance—partly due to the supposed resistance of code-based cryptosystems to attacks from quantum computing, partly due to the range of new techniques on offer, and partly because the fundamental problem of parameter selection is relatively poorly understood. For example, A. Couvreur’s work on filtration attacks on codes has an important impact on the design of code-based systems using wild Goppa codes or algebraic geometry codes, and on the choice of parameter sizes for secure implementations.

Coding theory also has important practical applications in the improvement of conventional symmetric cryptosystems. For example, D. Augot’s recent work on MDS matrices via BCH codes gives a more efficient construction of optimal diffusion layers in block ciphers. Here we use combinatorial, non-algorithmic properties of codes, in the internals of designs of block ciphers.

While coding theory brings tools as above for the classical problems of encryption, authentication, and so on, it can also provide solutions to new cryptographic problems. This is classically illustrated by the use of Reed-Solomon codes in secret sharing schemes. Grace is involved in the study, construction and implementation of locally decodable codes, which have applications in quite a few cryptographic protocols : *Private Information Retrieval, Proofs of Retrievability, Proofs of Ownership*, etc.

HYCOMES Team

4. Application Domains

4.1. Modeling cyber-physical systems

Hybrid systems modeling plays a particular role in the design of cyber-physical systems, eg. systems mixing physical devices, computing platforms, communication buses and control and diagnosis software. A faithful modeling of the physical environment is a key element in a successful design of a cyber-physical system.

Several types of physical components can be found in a system, for example: mechanical, hydraulic or electrical. Component models should cover several viewpoints. For instance, the three viewpoints of an electronic device would be its electrical, thermal and reliability models. All these viewpoints interact, and it is not possible to analyze any of them in isolation. Let alone these complex cross-viewpoint interactions, modeling physics requires refined mathematics. For instance, it is a misconception to assume that physical laws result in smooth dynamics that can be captured by systems of ordinary differential equations. On the contrary, physics is often nonsmooth, meaning that trajectories may be discontinuous — consider the example of colliding billiard balls. Physical systems are networks of elementary components. The dynamics of each component can often be captured by a simple (differential) equation. However, these (differential) equations are coupled by network equations (Kirchhoff laws, mechanical couplings, ...) resulting from the structure of the system. The end result, is a system mixing differential equations with linear or algebraic constraints: a system of differential algebraic equations (DAE).

The Hycomes team is focusing on the design of hybrid systems modeling languages with DAE and nonsmooth dynamics (Fillipov differential inclusions, or complementarity systems), with applications in the energy industry (power plants, smart grids), and in the railway, automotive and aeronautic industries — see section 3.1 for a deeper insight on the research program.

4.2. Supporting requirements engineering for aeronautic systems design

The design of embedded systems onboard certified civil aircrafts, for instance navigation, fly-by-wire and FADEC (Full Authority Digital Engine Control) applications, has to follow a stringent discipline imposed by civil aviation authorities. Designers have to provide evidence that both the design process they used and the system under design meet several industry standards, including the well-known ED-79/ARP-4754A and the DO-178 A/B regarding hardware and software artifacts.

These standards prescribe that every feature of a design can be traced back to one or several system-level requirements. Conversely, evidence shall be provided that every requirement has been accounted for. Correctness, consistency, compatibility and completeness of requirements are four key properties described in the ED-79/ARP-4754A standard that should also be assessed every time requirements are transformed. This puts a high burden on designers, especially on the system architect: requirements capture and analysis is by large a heuristic and manual process.

Formal requirements engineering is in its infancy: mathematical models, formal analysis techniques and links to system implementation must be developed. We advocate the use of contract-based reasoning techniques (see section 3.3) to support requirements engineering activities, during the early stages of the design process [4].

LFANT Project-Team (section vide)

MARELLE Project-Team

4. Application Domains

4.1. Reliability of embedded software

Software embedded in physical devices performs computations where the inputs are provided by measures and the outputs are transformed into actions performed by actuators. To improve the quality of these devices, we expect that all the computations performed in this kind of software will need to be made more and more reliable. We claim that formal methods can serve this purpose and we develop the libraries and techniques to support this claim. This implies that we take a serious look at how mathematics can be included in formal methods, especially concerning geometry and calculus.

4.2. Security and Cryptography

The modern economy relies on the possibility for every actor to trust the communications they perform with their colleagues, customers, or providers. We claim that this trust can only be built by a careful scrutiny of the claims made by all public protocols and software that are reproduced in all portable devices, computers, and internet infrastructure systems. We advocate the use of formal methods in these domains and we provide easy-to-use tools for cryptographers so that the formal verification of cryptographic algorithms can become routine and amenable to public scrutiny.

4.3. Mathematics and Education

As libraries for theorem provers evolve, they tend to cover an ever increasing proportion of the mathematical background expected from engineers and scientists of all domains. Because the content of a formally verified library is extremely precise and explicit, we claim that this will provide a new kind of material for teaching mathematics, especially useful in remote education.

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

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.

We have participated in the Univerself Project (see below) on self-aware networks, and will be searching new cooperations.

4.2. Transport Systems

Participants: Stefan Haar, Simon Theissing.

We participate in the project MIC on multi-modal transport systems with in the IRT *System X*, with academic partners UPMC, IFSTTAR and CEA, and several industrial partners including Alstom (project leader), COSMO and Renault. Transportation operators in an urban area need to plan, supervise and steer different means of transportation with respect to several criteria:

- Maximize capacity;
- guarantee punctuality and robustness of service;
- minimize energy consumption.

The systems must achieve these objectives not only under ideal conditions, but also be robust to perturbations (such as a major cultural or sport event creating additional traffic), modifications of routes (roadwork, accidents, demonstrations, ...) and tolerant to technical failures. Therefore, systems must be enabled to raise appropriate alarms upon detection of anomalies, diagnose the type of anomaly and select the appropriate response.

While the above challenges belong already to the tasks of individual operators in the unimodal setting, the rise of and increasing demand for *multi-modal* transports forces to achieve these planning, optimization and control goals not in isolation, but in a cooperative manner, across several operators. The research task here is first to analyze the transportation system regarding the available means, capacities and structures, and so as to identify the impacting factors and interdependencies of the system variables. Based on this analysis, the task is to derive and implement robust planning, with tolerance to technical faults; diagnosis and control strategies that are optimal under several, possibly different, criteria (average case vs worst case performance, energy efficiency, etc.) and allow to adapt to changes e.g. from nominal mode to reduced mode, sensor failures, etc.

4.3. Biological Systems

Participants: Stefan Haar, Serge Haddad, Stefan Schwoon, Thomas Chatain, Loïg Jezequel.

We have begun in 2014 to examine concurrency issues in systems biology, and are currently enlarging the scope of our research's applications in this direction. To see the context, note that in recent years, a considerable shift of biologists' interest can be observed, from the mapping of *static* genotypes to *gene expression*, i.e. the processes in which genetic information is used in producing functional products. These processes are far from being uniquely determined by the gene itself, or even jointly with static properties of the environment; rather, *regulation* occurs throughout the expression processes, with specific mechanisms increasing or decreasing the production of various products, and thus modulating the outcome. These regulations are central in understanding cell fate (how does the cell differentiate ? Do mutations occur ? etc), and progress there hinges on our capacity to analyse, predict, monitor and control complex and variegated processes. Our first step in this domain is related in the conference contribution [33], where we apply Petri net unfolding techniques for the efficient computation of *attractors* in a regulatory network; that is, to identify strongly connected reachability components that correspond to stable evolutions, e.g. of a cell that differentiates into a specific functionality (or mutation). This constitutes the starting point of a broader research with Petri net unfolding techniques in regulation. In fact, the use of *ordinary* Petri nets for capturing regulatory network (RN) dynamics overcomes the limitations of traditional RN models : those impose e.g. Monotonicity properties in the influence that one factor had upon another, i.e. always increasing or always decreasing, and were thus unable to cover all actual behaviours (see [76]). Rather, we follow the more refined model of boolean networks of automata, where the local states of the different factors jointly determine which state transitions are possible. For these connectors, ordinary PNs constitute a first approximation, improving greatly over the literature but leaving room for improvement in terms of introducing more refined logical connectors. Future work thus involves transcending this class of PN models. Via unfoldings, one has access – provided efficient techniques are available – to all behaviours of the model, rather than over-or under-approximations as previously. This opens the way to efficiently searching in particular for determinants of the cell fate : which attractors are reachable from a given stage, and what are the factors that decide in favor of one or the other attractor, etc. The list of potential applications in biology and medicine of such a methodology would be too long to reproduce here.

MUTANT Project-Team

4. Application Domains

4.1. Authoring and Performing Interactive Music

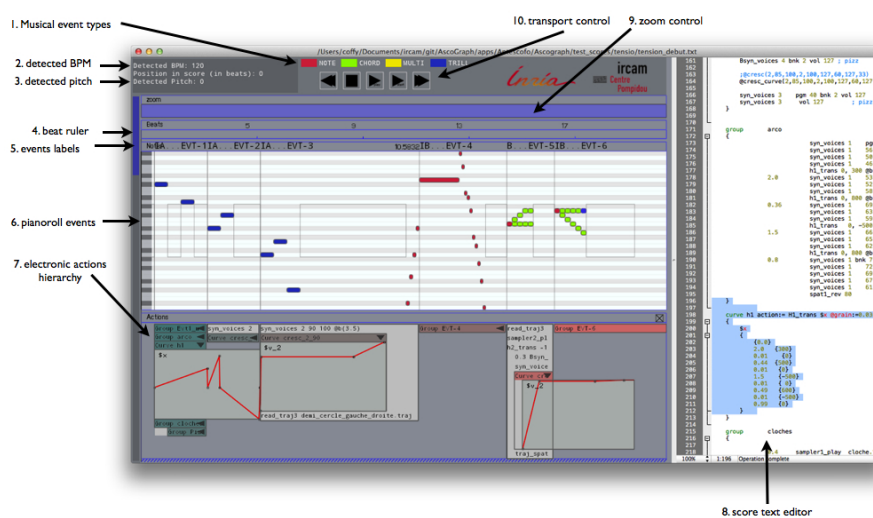


Figure 2. Screenshot of Ascograph, the Antescofo graphical score editor

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 40 prestigious public performances yearly. The outcomes of the teams's research will enhance the interactive and reactive aspects of this emerging paradigm as well as creating novel authoring tool for such purposes.

The *AscoGraph* authoring environment, started in 2013 and shown in Figure 2, is the first step towards such authoring environments. 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.

4.2. Realtime Music Information Retrieval

Realtime Music Information Retrieval is used as front-end for various applications requiring sonic interaction between software/hardware and the physical worlds. MuTant has focused on realtime machine listening since its inception and holds state-of-the-art algorithms for realtime alignment of audio to symbolic score, realtime tempo detection, realtime multiple-pitch extraction. Recent results have pushed our application to more generalised listening schemes beyond music signals as reported in [26].

4.3. Automatic Accompaniment/Creative Tools for Entertainment Industry

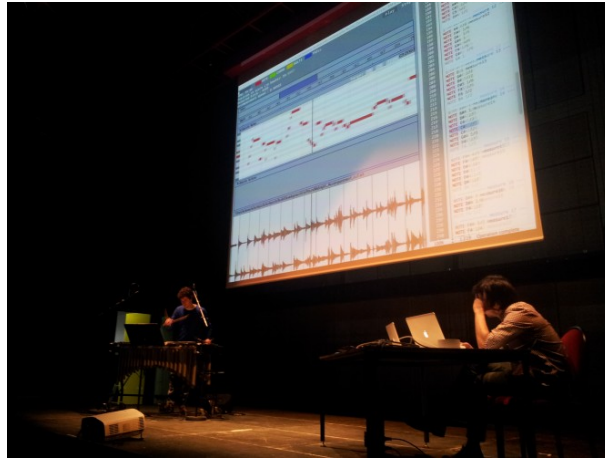


Figure 3. Automatic Accompaniment Session with Antescofo during ACM CHI 2013 Conference

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*. Technologies developed around *Antescofo* can enhance interaction between user and the computer for such large public applications.

Highlights in 2014 include collaborations with Orchestre de Paris Archives that resulted in prototype **demonstrated to public** in June 2014.

We will 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* project-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. Provably safe and efficient computing systems

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. Integrating a model checker and a theorem prover

The goal of combining model checking with inductive and co-inductive theorem in a rather appealing one. The strengths of systems in these two different systems are strikingly different. A model checker is capable of exploring a finite space automatically: such a tool can repeatedly explores all possible cases for how a computational space can be explored. On the other hand, a theorem prover might be able to prove clever things about a search space. For example, a model checker could attempt to discover whether or not there exists a winning strategy for, say, tic-tac-toe while an inductive theorem prover might be able to prove that if there is a winning strategy from one board then there is a winning strategy from any symmetric version of that board. Of course, being about to combine proofs from these system could drastically reduce the state exploration and proof certificate that needs to be produced to prove the existence of winning strategies.

Our first step to providing an integration of model checking and (inductive) theorem proving was to develop a strong logic, we call \mathcal{G} , that extends intuitionistic logic with notions of least and greatest fixed points. We have developed the proof theory of this logic in earlier papers [3] [52]. We have now recently converted the Bedwyr system so that it formally accepts almost all definitions and statements of theorems that are accepted by the inductive theorem prover Abella. Thus, these two systems are proving theorems in the same logic and their theorems can now be shared.

The tabling mechanism of Bedwyr has been extended so that its it can make use of previously proved lemmas. Thus, when a goal to prove that some board position has a winning strategy, the lemma can to conclude yes if some symmetric board position is already in the table.

For more about recent progress on providing checkable proof certificates for model checking, see the web site for Bedwyr <http://slimmer.gforge.inria.fr/bedwyr/>.

4.2. Implementing trusted proof checkers

Traditionally, theorem provers—whether interactive or automatic—are usually monolithic: if any part of a formal development was to be done in a particular theorem prover, then all parts of it would need to be done in that prover. Increasingly, however, formal systems are being developed to integrate the results returned from several, independent and high-performing, specialized provers: see, for example, the integration of Isabelle with an SMT solver [51] as well as the Why3 and ESC/Java systems.

Within the Parsifal team, we have been working on foundational aspects of this problem of integrating different provers. As we have described above, we have been developing a formal framework for defining the semantics of proof evidence. We have also been working on building prototype checkers of proof evidence which are capable to executing such formal definitions. The proof definition language described in the papers [47], [46] is currently given an implementation in the λ Prolog programming language [69]. This initial implementation will be able to serve as a “reference” proof checker: others developing proof evidence definitions will be able to use this reference checker to make sure that they are getting their definitions to do what they expect.

Using λ Prolog as an implementation language has both good and bad points. The good points are that it is rather simple to confirm that the checker is, in fact, sound. The language also supports a rich set of abstracts which make it impossible to interfere with the code of the checker (no injection attacks are possible). On the negative side, however, the performance of our λ Prolog interpreters is lower than specially written checkers and kernels.

4.3. Trustworthy implementations of theorem proving techniques

Instead of integrating different provers by exchanging proof evidence and relying on a back-end proof-checker, another approach to integration consists in re-implementing the theorem proving techniques as proof-search strategies, on an architecture that guarantees correctness. Focused systems can serve as the basis of such an architecture, identifying points of choice and backtrack and providing primitives for the exploration of the search space. These form a trusted *Application Programming Interface* that can be used to program and experiment various proof-search heuristics without worrying about correctness. No proof-checking is needed if one trusts the implementation of the API.

Following the description, in this framework, of quantifier-free techniques such as DPLL(T) [2], we are now exploring how the architecture can be adapted to accommodate techniques that handle quantifiers. In particular, unification-based or triggers-based techniques [37], [49].

This approach has led to the development of the Psyche engine.

PI.R2 Project-Team (section vide)

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

POSTALE Team (section vide)

PRIVATICS Project-Team

3. Application Domains

3.1. Domain 1: Privacy in smart environments.

One illustrative example is our latest work on privacy-preserving smart-metering [2]. Several countries throughout the world are planning to deploy smart meters in house-holds in the very near future. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in real-time, consumption and possibly adjust generation and prices according to the demand. Billing customers by how much is consumed and at what time of day will probably change consumption habits to help matching energy consumption with production. In the longer term, with the advent of smart appliances, it is expected that the smart grid will remotely control selected appliances to reduce demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart-meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases. Analysing high-resolution consumption data, Non-intrusive Appliance Load Monitoring (NALM) can be used to identify a remarkable number of electric appliances (e.g., water heaters, well pumps, furnace blowers, refrigerators, and air conditioners) employing exhaustive appliance signature libraries. We developed DREAM, Differentially privatE smArt Metering, a scheme that is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart-meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user's trace when he watched TV or turned on heating.

3.2. Domain 2: Big Data and Privacy

We believe that another important problem will be related to privacy issues in big data. Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. It is now an urgent need to develop Privacy-Preserving Data Analytics (PPDA) systems that collect and transform raw data into a version that is immunized against privacy attacks but that still preserves useful information for data analysis. This is one of the objectives of Privatics. There exists two classes of PPDA according to whether the entity that is collecting and anonymizing the data is trusted or not. In the trusted model, that we refer to as Privacy-Preserving Data Publishing (PPDP), individuals trust the publisher to which they disclose their data. In the untrusted model, that we refer to as Privacy-Preserving Data Collection (PPDC), individuals do not trust the data publisher. They may add some noise to their data to protect sensitive information from the data publisher.

Privacy-Preserving Data Publishing: In the trusted model, individuals trust the data publisher and disclose all their data to it. For example, in a medical scenario, patients give their true information to hospitals to receive proper treatment. It is then the responsibility of the data publisher to protect privacy of the individuals' personal data. To prevent potential data leakage, datasets must be sanitized before possible release. Several proposals have been recently proposed to release private data under the Differential Privacy model [25, 56, 26, 57, 50]. However most of these schemes release a "snapshot" of the datasets at a given period of time. This release often consists of histograms. They can, for example, show the distributions of some pathologies (such as cancer, flu, HIV, hepatitis, etc.) in a given population. For many analytics applications, "snapshots" of data are not enough, and sequential data are required. Furthermore, current work focusses on rather simple data structures, such as numerical data. Release of more complex data, such as graphs, are often also very useful. For example, recommendation systems need the sequences of visited websites or bought items. They also need to analyse people connection graphs to identify the best products to recommend. Network trace analytics also rely on sequences of events to detect anomalies or intrusions. Similarly, traffic analytics applications typically need sequences of visited places of each user. In fact, it is often essential for these applications to know that user A moved from position 1 to position 2, or at least to learn the probability of a move from position 1 to position 2. Histograms would typically represent the number of users in position 1 and position 2, but would not provide the number of users that moved from position 1 to position 2. Due to the inherent sequentiality and high-dimensionality of sequential data, one major challenge of applying current data sanitization solutions on sequential data comes from the uniqueness of sequences (e.g., very few sequences are identical). This fact makes existing techniques result in poor utility. Schemes to privately release data with complex data structures, such as sequential, relational and graph data, are required. This is one the goals of Privatics. In our current work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n-grams [15]. We then intend to extend this approach to more complex data structures.

Privacy-Preserving Data Collection: In the untrusted model, individuals do not trust their data publisher. For example, websites commonly use third party web analytics services, such as Google Analytics to obtain aggregate traffic statistics such as most visited pages, visitors' countries, etc. Similarly, other applications, such as smart metering or targeted advertising applications, are also tracking users in order to derive aggregated information about a particular class of users. Unfortunately, to obtain this aggregate information, services need to track users, resulting in a violation of user privacy. One of our goals is to develop Privacy-Preserving Data Collection solutions. We propose to study whether it is possible to provide efficient collection/aggregation solutions without tracking users, i.e. without getting or learning individual contributions.

PROSECCO Project-Team

4. Application Domains

4.1. Cryptographic Protocol Libraries

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 and SSH, 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 serve pages over HTTPS, authenticate users with a single sign-on protocol such as OAuth, encrypt user files on the server-side using XML encryption, and deploy client-side cryptographic mechanisms using a JavaScript cryptographic library. The security of these applications depends on the public key infrastructure (X.509 certificates), web browsers' implementation of HTTPS and the same origin policy (SOP), the semantics of JavaScript, HTML5, and their various associated security standards, as well as the correctness of the specific web application code of interest. We build analysis tools to find bugs in all these artifacts and verification tools that can analyze commercial web applications and evaluate their security against sophisticated web-based attacks.

SECRET Project-Team

4. Application Domains

4.1. Domain

Our main application domains are:

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

SPADES 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, we are looking to propose 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/ dSPACE⁰) and execution platforms (OS such as VXWORKS, QNX, real-time versions of LINUX ...) start 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: SCADE⁰ (based on LUSTRE), CONTROLBUILD and RT-BUILDER (based on SIGNAL) from GEENYSYS⁰ (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.

⁰<http://www.dspaceinc.com>

⁰<http://www.esterel-technologies.com>

⁰<http://www.geensoft.com>

SPECFUN Project-Team

4. Application Domains

4.1. Experimental mathematics with special functions

Applications in combinatorics and mathematical physics frequently involve equations of so high orders and so large sizes, that computing or even storing all their coefficients is impossible on existing computers. Making this tractable is another challenge of our project. The approach we believe in is to design algorithms of good, ideally quasi-optimal, complexity in order to extract precisely the required data from the equations, while avoiding the computationally intractable task of completely expanding them into an explicit representation.

Typical applications with expected high impact are the automatic discovery and proof of results in combinatorics and mathematical physics for which human proofs are currently unattainable.

SUMO Project-Team

4. Application Domains

4.1. Telecommunication network management

The domain of autonomic network management, will remain an important playground for SUMO. It covers a wide variety of problems, ranging from distributed (optimal) control to distributed diagnosis, optimization, re-configuration, provisioning, etc. We have a long experience in model-based diagnosis, in particular distributed (active) diagnosis, and have recently proposed promising techniques for self-modeling. It consists in building the model of the managed network on the fly, guided by the needs of the diagnosis algorithm. This approach allows one to deal with potentially huge models, that are only described by their construction grammar, and discovered at runtime. Another important research direction concerns the management of “multi-resolution” models, that can be considered at different granularity levels. This feature is central to network design, but has no appropriate modeling formalism nor management approaches. This is a typical investigation field for abstraction techniques. Technology is ahead of theory in this domain since networks are already driven or programmed through management policies, that assign high level objectives to an abstract view of the network, leaving open the question of their optimal implementation. As a last topic of investigation, today management issues are no longer isolated within one operator, but range across several of them, up to the supported services, which brings game theory aspects into the picture.

4.2. Control of data centers

Data centers are another example of a large scale reconfigurable and distributed system: they are composed of thousands of servers on which Virtual Machines (VM) can be (de)activated, migrated, etc. depending on the requests of the customers, on the load of the servers and on the power consumption. Autonomic management functionalities already exist to deploy and configure applications in such a distributed environment. They can also monitor the environment and react to events such as failures or overloads and reconfigure applications and/or infrastructures accordingly and autonomously. To supervise these systems, Autonomic Managers (AM) can be deployed in order to apply administration policies of specific aspects to the different entities of a data center (servers, VM, web services, power supply, etc). These AMs may be implemented in different layers: the hardware level, the operating system level or the middleware level. Therefore several control loops may coexist, and they have to take globally consistent decisions to manage the trade-off between availability, performance, scalability, security and energy consumption. This leads to multi-criteria optimization and control problems in order to automatically derive controllers in charge of the coordination of the different AMs. We are relatively new on this topic, that will require more technical investment from us. But we are driven to it by both the convergence of IT and networking, by virtualization techniques that reach networks (see the growing research effort about network operating systems), and by the call for more automation in the management of clouds. We believe our experience in network management can help. Some members of SUMO are already involved in the ANR Ctrl-Green, which addresses the controller coordination problem. We are also in contact with the Myriads team, which research interests moved from OS for grids/clouds to autonomic methods. This is supported as well by the activities of b<>com, the local IRT, where some projects in cloud management and in networking may start joint activities.

4.3. Web services and distributed active documents

Data centric systems are already deployed, and our goal is not to design new languages, architectures, or standards for them, but rather to propose techniques for the verification and monitoring of the existing systems. A bottleneck is the complexity and heterogeneity of web-based systems, that make them difficult to model and analyze. However, one can still hope for some lightweight verification or monitoring techniques for some specific aspects, for example to check the absence of conflict of interest in a transaction system, to verify

(off line) and maintain (on line) the QoS, to prevent security breaches, etc. Safety aspects of WS are little addressed; any progress in that area would be useful. Besides, modeling issues are central for some applications of data centric systems. Collaborative work environments with shared active documents can be found in many domains ranging from banking, maintenance of critical systems, webstores... We consider that models for data driven systems can find applications in most of these application areas. Our approach will be to favor purely declarative approaches for the specification of such collaborative environments. We have contacts with Centre Pasteur in Yaoundé on the design of diseases monitoring systems in developing countries. Diseases monitoring systems can be seen as a collaborative edition work, where each actor in the system reports and aggregates information about cases he or she is aware of. This collaboration is an opportunity to confront our models to real situations and real users needs. Formally modeling such a large distributed system can be seen as a way to ensure its correctness. We also envision to promote this approach as a support for maintenance operations in complex environments (train transportation, aeronautics,...). We believe this framework can be useful both for the specification of distributed maintenance procedures, for circulating information and sharing processes across teams, but also for the analysis of the correctness of procedures, possibly for their optimization or redesign, and finally to automatically elaborate logs of maintenance operations. We are in contact with several major companies on these topics, for the maintenance application side. Other industrial contacts need to be built: we have preliminary contact with IBM (leader in business artifacts), and would like to establish relations with SAP (leader in service architectures).

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 2014 the **TASC** team was involved in the following application domains:

- *Replanning* in industrial timetabling problems in a Labcom project with **Eurodécision**.
- *Planning and replanning* in Data Centres taking into account energy consumption in the EPOC (Energy Proportional and Opportunistic Computing system) 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**.

TEA Project-Team

4. Application Domains

4.1. Application Domains

From our continuous collaboration with major academic and industrial partners through projects TOPCASED, OPENEMBEDD, SPACIFY, CESAR, OPEES, P and CORAIL, our experience has primarily focused on the aerospace domain. The topics of time and architecture of team TEA extend to both avionics and automotive, as demonstrated from this section to section 8. Yet, the research focus on time in team TEA is central in any aspect of, cyber-physical, embedded system design in automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

Nonetheless, the application domains of our two direct collaborations with industry, avionics with Thales and automotive Toyota, are perfectly in line with the research objectives of team TEA and will allow us to quickly stream our theoretical results onto software and standards, which we will continue to distribute in open-source.

Multi-scale, multi-aspect time modelling, analysis and software synthesis will greatly contribute to architecture modelling in these domains, with applications to optimised (distributed, parallel, multi-core) code generation for avionics (our project with Thales avionics, section 8) as well as modelling standards, real-time simulation and virtual integration in automotive (our project with Toyota, section 8).

Together with the importance of open-source software, one of these project, the FUI Project P, demonstrated that a centralised model for system design could not just be a domain-specific programming language, such as discrete Simulink data-flows or a synchronous language. Synchronous languages implement a fixed model of time using logical clocks that are abstraction of time as sensed by software. They correspond to a fixed viewpoint in system design, and in a fixed hardware location in the system, which is not adequate to our purpose and must be extended.

In project P, we first tried to define a centralised model for importing discrete-continuous models onto a simplified implementation of SIMULINK: P models. Certified code generators would then be developed from that format. Because this does not encompass all aspects being translated to P, the P meta-model is now being extended to architecture description concepts (of the AADL) in order to become better suited for the purpose of system design. Another example is the development of System Modeller on top of SCADE, which uses the more model-engineering flavoured formalism SysML to try to unambiguously represent architectures around SCADE modules.

An abstract specification formalism, capable of representing time, timing relations, with which heterogeneous models can be abstracted, from which programs can be synthesised, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all onboard electronics of PSA cars. This ‘hardware in the loop’ simulator was used to test equipments supplied by other manufacturers with respect to virtual cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

RT-Builder is the commercial implementation of Signal, whose industrial transfer with TNI was realised in the 90s by Paul Le Guernic and Albert Benveniste. As its actual industry usage has demonstrated, it is clear that the synchronous multi-clocked, or polychronous MoCC of Signal is an appropriate semantic core for the design of embedded software architectures.

TEMPO Team

4. Application Domains

4.1. Cyber Physical Systems

The overall project is geared towards the development of complex heterogeneous cyber physical systems that require high reliability such as nuclear power, energy distribution, industry automation and transportation, where formal verification methods are necessary.

4.2. 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.

4.3. Automated Test Generation

Manual testing is tedious. Automated testing makes it possible to increase test coverage while also minimizing the amount of redundancy created by manual testing.

TOCCATA Project-Team

4. Application Domains

4.1. Mission-Critical Software

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 the following. For each of them we refer to our past or current actions, in particular in relations with projects, contracts and industrial partners. Currently our industrial collaborations mainly belong to the first of these domains, transportation.

- **Transportation** including aeronautics, railroad, space flight, automotive.

These domains were considered in the context of the ANR U3CAT project, led by CEA, in partnership with Airbus France, Dassault Aviation, Sagem Défense et Sécurité. It included proof of C programs via *Frama-C/Jessie/Why*, proof of floating-point programs, the use of the *Alt-Ergo* prover via CAVEAT tool (CEA) or *Frama-C/WP*. This action is continued in the new project Soprano.

Aeronautics is the main target of the Verasco project, led by Verimag, on the development of certified static analyzers, in partnership with Airbus.

The former FUI project Hi-Lite, led by Adacore company, uses *Why3* and *Alt-Ergo* as back-end to SPARK2014, an environment for verification of Ada programs. This is applied to the domain of aerospace (Thales, EADS Astrium). This action is continued in the new joint laboratory ProofInUse. A recent paper [71] provides an extensive list of applications of SPARK, a major one being the British air control management.

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

- **Energy** is naturally an application in particular with our long-term partner CEA, in the context of U3CAT and Soprano projects.
- **Communications and Data** in particular in contexts with a particular need for security or confidentiality: smart phones, Web applications, health records, electronic voting, etc.

Part of the applications of SPARK [71] include verification of security-related properties, including cryptographic algorithms.

Our new AJACS project addresses issues related to security and privacy in web applications written in Javascript, also including correctness properties.

The Cubicle model checker modulo theories based on the *Alt-Ergo* SMT prover, in collaboration with Intel Strategic Cad Labs (Hillsboro, OR, USA) is particularly targeted to the verification of concurrent programs and protocols (<http://cubicle.lri.fr/>).

- **Medicine**, including diagnostic devices, computer-assisted surgery

Such applications involve techniques for control and command close to what is done in transportation. Moreover, in this context, there is a need for modeling using differential equations, finite elements, hybrid systems, which are considered in other projects of us: FastRelax, ELFIC, Cafein.

- **Financial applications, banking**

We add projects in the past about safety and security of smart cards, in collaboration with Gemalto (European project VerifiCard, two CIFRE theses). Banking is naturally a domain of application of techniques dealing with security and confidentiality already mentioned above.

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.

VERIDIS Project-Team

4. Application Domains

4.1. Application Domains

Distributed algorithms and protocols are 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. Computing infrastructure must be highly available and is ideally invisible to the end user, therefore correctness is crucial. 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, the cost of unavailable services due to malfunctioning system components and software provides wider incentives for verification. For example, we are working on modeling and verifying medical devices that require closed-loop models of both the system and its environment.

APICS Project-Team

4. Application Domains

4.1. Introduction

Application domains are naturally linked to the problems described in Sections 3.2.1 and 3.2.2. By and large, they split into a systems-and-circuits part and an inverse-source-and-boundary-problems part, united under a common umbrella of function-theoretic techniques as described in Section 3.3.

4.2. Inverse source problems in EEG

Participants: Laurent Baratchart, Juliette Leblond.

This work is performed in collaboration with Maureen Clerc and Théo Papadopoulo from the Athena Project-Team, and Jean-Paul Marmorat (Centre de mathématiques appliquées - CMA, École des Mines de Paris).

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. Indeed, the latter involves propagating the initial conditions through several layers of different conductivities, from the boundary shell down to the center of the domain where the singularities (*i.e.* the sources) lie. Once propagated to the innermost sphere, it turns out that traces of the boundary data on 2-D cross sections coincide with analytic functions with branched singularities in the slicing plane [3]. The singularities are related to the actual location of the sources, namely their moduli reach in turn a maximum when the plane contains one of the sources. Hence we are back to the 2-D framework of Section 3.3.3, and recovering these singularities can be performed *via* best rational approximation. The goal is to produce a fast and sufficiently accurate initial guess on the number and location of the sources in order to run heavier descent algorithms on the direct problem, which are more precise but computationally costly and often fail to converge if not properly initialized.

Numerical experiments give very good results on simulated data and we are now engaged in the process of handling real experimental data (see Sections 5.6 and 6.1), in collaboration with the Athena team at Inria Sophia Antipolis, neuroscience teams in partner-hospitals (la Timone, Marseille), and the BESA company (Munich).

4.3. Inverse magnetization problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Dmitry Ponomarev.

Generally speaking, inverse potential problems, similar to the one appearing in Section 4.2, occur naturally in connection with systems governed by Maxwell's equation in the quasi-static approximation regime. In particular, they arise in magnetic reconstruction issues. A specific application is to geophysics, which led us to form the Inria Associate Team "IMPINGE" (Inverse Magnetization Problems IN GEosciences) together with MIT and Vanderbilt University. A recent collaboration with Cerege (CNRS, Aix-en-Provence), in the framework of the ANR-project MagLune, completes this picture, see Section 8.2.2.

To set up the context, recall that the Earth's geomagnetic field is generated by convection of the liquid metallic core (geodynamo) and that rocks become magnetized by the ambient field as they are formed or after subsequent alteration. Their remanent magnetization provides records of past variations of the geodynamo, which is used to study important processes in Earth sciences like motion of tectonic plates and geomagnetic reversals. Rocks from Mars, the Moon, and asteroids also contain remanent magnetization which indicates the past presence of core dynamos. Magnetization in meteorites may even record fields produced by the young sun and the protoplanetary disk which may have played a key role in solar system formation.

For a long time, paleomagnetic techniques were only capable of analyzing bulk samples and compute their net magnetic moment. The development of SQUID microscopes has recently extended the spatial resolution to sub-millimeter scales, raising new physical and algorithmic challenges. This associate team aims at tackling them, experimenting with the SQUID microscope set up in the Paleomagnetism Laboratory of the department of Earth, Atmospheric and Planetary Sciences at MIT. Typically, pieces of rock are sanded down to a thin slab, and the magnetization has to be recovered from the field measured on a parallel plane at small distance above the slab.

Mathematically speaking, both inverse source problems for EEG from Section 4.2 and inverse magnetization problems described presently 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:

$$\int_{\Omega} \frac{\operatorname{div} m(x') dx'}{|x-x'|}, \quad (1)$$

outside the volume Ω of the object. The difference is that the distribution m is located in a volume in the case of EEG, and on a plane in the case of rock magnetization. This results in quite different identifiability properties, see [38] and Section 6.1.2.

4.4. Free boundary problems

Participants: Laurent Baratchart, Juliette Leblond.

This work is conducted in part with Yannick Privat, CNRS, Lab. J.-L. Lions, Paris.

The team has engaged in the study of problems with variable conductivity σ , governed by a 2-D equation of the form $\operatorname{div}(\sigma \nabla u) = 0$. Such equations are in one-to-one correspondence with real parts of solutions to conjugate-Beltrami equations $\bar{\partial} f = \nu \partial \bar{f}$, so that complex analysis is a tool to study them, see [4], [14], [34]. This research was prompted by issues in 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 [58]. Investigations started in collaboration with CEA-IRFM (Cadarache) and the Laboratoire J.-A. Dieudonné at the Univ. of Nice-SA. Within the team, they now expand to cover Dirichlet-Neumann problems for larger classes of conductivities, *cf.* in particular [34] (see Section 6.2).

4.5. Identification and design of microwave devices

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

This is joint work 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 function-theoretic applications by the team is the identification and design of physical systems whose performance is assessed frequency-wise. 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 an essentially discrete set of wave vectors is selected. In the considered range of frequency, the electrical field in each cavity can be 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 screws are conductors, they behave 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 an iris is opposite to that of a screw: no condition is imposed on 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 Maxwell's equations is given by the solution of a second order differential equation. Thus, one obtains an electrical model of the filter as a sequence of electrically-coupled resonant circuits, each circuit being modeled by two resonators, one per mode, the resonance frequency of which represents the frequency of a mode, and whose resistance accounts for electric losses (current on the surface) of the cavities.

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. One is now interested in the power which is transmitted and reflected. This leads one to define a scattering matrix S , which may 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 on Figure 2), and the key step consists in expressing the components of the equivalent electrical circuit as functions 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 quality of design, in particular 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 linearizing near the central frequency, which is no longer conjugate symmetric (*i.e.* the underlying system may no longer have real coefficients) but whose degree is divided by 2 (8 in the example).

In short, the strategy for identification 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).

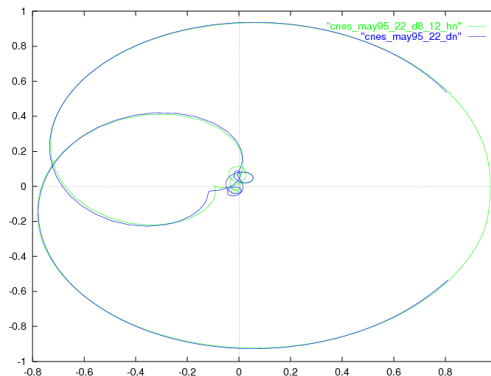


Figure 2. Nyquist Diagram. Rational approximation (degree 8) and data - S_{22} .

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, which are 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 proposed, in order to derive guaranteed optimal multi-band filter responses subject to modulus constraints [11]. This generalizes the classical single band design techniques based on Chebyshev polynomials and elliptic functions. The approach relies on the fact that the modulus of the scattering parameter $|S_{1,2}|$ admits a simple expression in terms of the filtering function $D = |S_{1,1}|/|S_{1,2}|$, namely

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

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 unitary polynomial solving the classical Feldtkeller spectral equation:

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

The relative simplicity of the derivation of a filter's response, under modulus constraints, owes much to the possibility of forgetting about Feldtkeller's equation and express all design constraints in terms of the filtering function. This is 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 among the team's recent investigation, where techniques based on constrained Nevanlinna-Pick interpolation problems are being considered (see Section 6.3.1).

Through contacts with CNES (Toulouse) and UPV (Bilbao), Apics got further involved three years ago with the design of amplifiers which, unlike filters, are active devices. A prominent issue here is stability. A twenty years back, it was not possible to simulate unstable responses, and only after building a device could one detect instability. The advent of so-called *harmonic balance* techniques, which compute steady state responses of linear elements in the frequency domain and look for a periodic state in the time domain of a network connecting these linear elements *via* static nonlinearities made it possible to compute the harmonic response of a (possibly nonlinear and unstable) device [82]. This has had tremendous impact on design, and there is a growing demand for software analyzers.

There are two types of stability involved. The first is stability of a fixed point around which the linearized transfer function accounts for small signal amplification. The second is stability of a limit cycle which is reached when the input signal is no longer small and truly nonlinear amplification is attained (*e.g.* because of saturation). Work by the team so far is concerned with the first type of stability, and emphasis is put on defining and extracting the "unstable part" of the response, see Section 6.4.

ASPI Project-Team

4. Application Domains

4.1. Localisation, navigation and tracking

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 [39], [33], 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 [38], [52], 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 , 5.2 , and 5.3 .

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 inefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [46], 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 (section vide)

BIPOP Project-Team

4. Application Domains

4.1. 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.

4.2. 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 the question to simulate a complete 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.

4.3. Walking robots

As compared to rolling robots, the walking ones – for example hexapods – possess definite advantages whenever the ground is not flat 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.

4.4. Optimization

Optimization exists in virtually all economic sectors. Simulation tools can be used to optimize the simulated system. 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 involved 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.

4.5. Computer graphics animation

Computer graphics animation is dedicated to the numerical modeling and simulation of physical phenomena featuring a high visual impact. Typically, deformable objects prone to strong deformation, large displacements, complex and nonlinear or even nonsmooth behavior, are of interest for this community. We are interested in two main mechanical phenomena: on the one hand, the behavior of slender (nonlinear) structures such as rods, plates and shells; on the other hand, the effect of frictional contact between rigid or deformable bodies. In both cases the goal is to design realistic, efficient, robust, and controllable computational models. 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. 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. Finally, we are interested in studying certain discrepancies (inexistence of solution) due to the combination of incompatible models such as contacting rigid bodies subject to Coulomb friction.

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 .

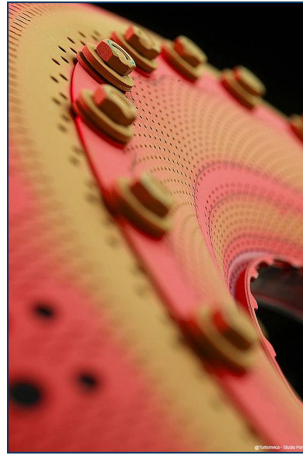


Figure 2. Effusion cooling of aeronautical combustion chambers: close view of a typical perforated chamber wall

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.

An important aspect that we began to address in this project 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 be particularly useful to develop one or multidimensional models of wall-fluid interaction [27]. The large eddy simulation performed by Florenciano [26] clearly put into evidence the strong effect of the presence of an acoustic wave in the crossflow on the dynamics of the heat transfer coefficient at the wall.

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.

CLASSIC Project-Team (section vide)

COMMANDS Project-Team

4. Application Domains

4.1. Energy production planning

We work with colleagues from U. Chile, in the framework of Inria Chile, on the management of electricity production and storage for a microgrid.

4.2. Fuel saving by optimizing airplanes trajectories

We have a collaboration with the startup Safety Line on the optimization of trajectories for civil aircrafts.

4.3. Hybrid vehicles

We have a collaboration with IFPEN on the energy management for hybrid vehicles.

CORIDA Team

4. Application Domains

4.1. Biology and Medicine

4.1.1. 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. [63]. 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. [53], 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.

4.2. Simulation of viscous fluid-structure interactions

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

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

4.3. 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 in collaboration with SAGEM.

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) in collaboration with St Antoine Hospital in Paris.

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 Benchmark Problems

- ρMNK -landscapes constitute a problem-independent model used for constructing multiobjective multimodal landscapes with objective correlation. They extend single-objective NK-landscapes [59] and multiobjective NK-landscapes with independent objective functions [54]. The four parameters defining a ρMNK -landscape are: (i) the size of (binary string) solutions N , (ii) the variable correlation $K < N$, (iii) the number of objective functions M , and (iv) the correlation coefficient ρ . A number of problem instances and an instance generator are available at the following URL: <http://mocobench.sf.net/>.
- The Unconstrained Binary Quadratic Programming (UBQP) problem is known to be a unified modeling and solution framework for many combinatorial optimization problems [60]. Given a collection of n items such that each pair of items is associated with a profit value that can be positive, negative or zero, UBQP seeks a subset of items that maximizes the sum of their paired values. We proposed an extension of the single-objective UBQP to the multiobjective case (mUBQP), where multiple objectives are to be optimized simultaneously. We showed that the mUBQP problem is both NP-hard and intractable. Some problem instances with different characteristics and an instance generator are also available at the following URL: <http://mocobench.sf.net/>.

4.2. Transportation and logistics

- **Scheduling problems under uncertainty:** The flow-shop scheduling problem is one of the most well-known problems from scheduling. However, most of the works in the literature use a deterministic single-objective formulation. 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 bi-objective flow-shop problem with uncertainty on the duration, minimizing in addition the maximum tardiness, is also studied. It allows us to develop and test multi-objective (and not only bi-objective) optimization methods under uncertainty.
- **Routing problems under uncertainty:** The vehicle routing problem (VRP) is a well-known problem and it has been studied since the end of the fifties. 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. Uncertainty on the demands has also been introduced in the model.

4.3. Bioinformatics and Health care

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 at analyzing 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 data mining 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 amount of data increases rapidly. Within the collaboration with Genes Diffusion, specialized in genetics and animal reproduction for bovine, swine, equine and rabbit species, we study combinations of Single Nucleotide Polymorphisms (SNP) that can explain some phenotypic characteristics. Therefore feature selection for regression is addressed using metaheuristics.

4.3.2. Optimization for health care

The collaboration with the Alicante company, a major actor in the hospital decision making, deals 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 multi-objective combinatorial optimization are implemented, requiring extensive work to model the problem, to define criteria optimization and to design specific optimization methods.

ECUADOR Project-Team

4. Application Domains

4.1. Algorithmic Differentiation

Algorithmic Differentiation of programs gives sensitivities or gradients, useful for instance for :

- 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. On 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 the most efficient 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* [28] 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 not well known. Only some measurements at arbitrary places and times are available. A good initial state is found by solving a least squares problem between the measurements 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 [31]. Figure 1 shows an example of a data assimilation exercise using the oceanography code OPA [29] 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 measurements 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 is more complex, 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.

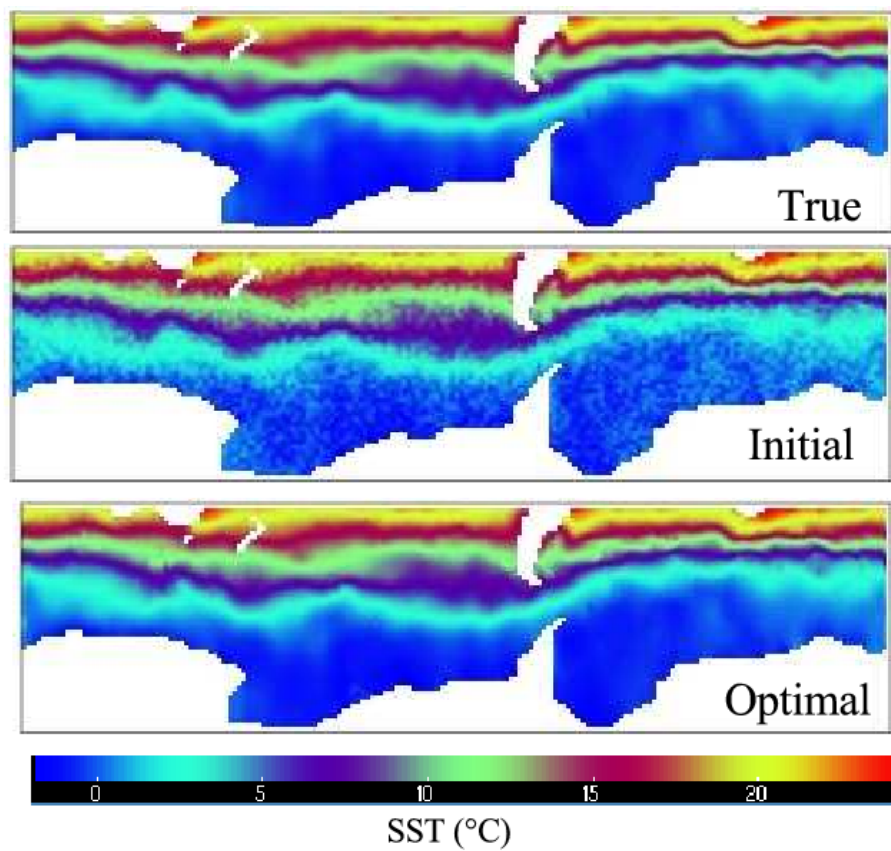


Figure 1. Twin experiment using the adjoint of OPA. Random noise was added to a simulation of the sea surface temperature around the Antarctic, and we remove this noise by minimizing the discrepancy with the physical model

4.4. Linearization

Simulating a complex system often requires solving a system of Partial Differential Equations. This can be 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 an efficient approximation: 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.

GAMMA3 Project-Team (section vide)

GECO Project-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, [91], [96]). The rapid evolution of the domain is driven by a multitude of experiments getting more and more precise and complex (see the recent review [52]). 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 [34],
- bounds on the controllability time [30],
- STIRAP processes [101],
- simultaneous control [74],
- optimal control ([70], [43], [54]),
- numerical simulations [80].

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, [35], [71], [92], [49]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [39], [50], [67], [40]). 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 [100]. Moreover, it has been shown that a relevant model, the quantum oscillator, is not even approximately controllable [93], [83]. These negative results have been more recently completed by positive ones. In [41], [42] 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 [56]). 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 [82], [87], [68]. While [82] studies a controlled Schrödinger equation in \mathbb{R} for which the uncontrolled Schrödinger operator has mixed spectrum, [87], [68] 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 [59], [46].

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 [65]. Optimal control approaches have also been considered [38], [51]. 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 [89]) 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 ([90]) and then modified by Citti and Sarti ([55]). The model is based on experimental observations, and in particular on the fundamental work by Hubel and Wiesel [64] 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 robotics and neurophysiology. It could help to design better control strategies for robots and artificial limbs, yielding smoother and more progressive movements. Another 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 [37]), 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 [99] 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 [69] or for Markov processes [88]. 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 ([76]).

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 [97], [77]).

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 [78]. 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 [84] and, provided that the admissible modes are finitely many, the Lyapunov function can be taken polyhedral or polynomial [44], [45], [57]. 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 [58] and references therein). Algebraic approaches to prove the stability of switched systems under arbitrary switching, not relying on Lyapunov techniques, have been proposed in [75], [31].

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 [63].

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 [72]. It is known that, in this context, stability cannot be tested on periodic signals alone [47].

Finally, let us mention that little is known about infinite-dimensional switched system, with the exception of some results on uniform asymptotic stability ([81], [94], [95]) and some recent papers on optimal control ([62], [102]).

GEOSTAT Project-Team

4. Application Domains

4.1. Application domains

As mentioned above, applicative aspects in GEOSTAT encompass biomedical data (heartbeat signal analysis with IHU LIRYC, biomedical applications in speech signal analysis) and the study of universe science datasets. GEOSTAT's objectives in analysis of biomedical data hinge on the following observations:

- The analysis and detection of cardiac arrhythmia and pathological voice disorders is a paradigm in nonlinear methodologies applied to these types of signals.
- The classical hypothesis under linear approaches are confronted with strong nonlinearities, aperiodicity and chaotic phenomena present in these signals.
- Existing nonlinear approaches are lacking physiological interpretation.

Our objective in this part is to propose new measures based on low-level transition characteristics, these transition phenomena being related to general concepts associated to predictability in complex systems.

I4S Project-Team

4. Application Domains

4.1. Civil Engineering

For at least three decades, monitoring the integrity of the civil infrastructure has been an active research topic because of major economical and societal issues, such as durability and safety of infrastructures, buildings and networks. Control of civil structures began a century ago. At stake is the 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 is very ancient since for example early developments of optical methods to monitor civil structures began in the 70s and SHM practice can be traced back to the 50s with the vibrating wire sensors as strain gauges for dams. Still the number of sensors actually placed on civil structures is kept to a minimum, mainly for cost reasons, but also because the return on investment sensing and data processing technologies is not properly established for civil structures. One of the current thematic priorities of the C2D2 governmental initiative is devoted to construction monitoring and diagnostics. The picture in Asia (Japan, and also China) is somewhat different, in that 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. However, the actual use of available data for operational purpose remains unclear.

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. 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 can be significant w.r.t. the deviations to be monitored.

Thermomechanical prestress states affect the dynamic and the static behavior of most bridges, not only of very long and flexible ones. So, the reliable and fast determination of the state of prestress and prestrain associated with a temperature field becomes a crucial step in several engineering processes such as the health monitoring of civil structures. The best possible reconstruction of the temperature field could then become part of a complete process including massively distributed sensing of thermomechanical information on the structure, modeling and algorithms for the on-line detection of damages in the sense of abnormalities with regard to a nominal state, the whole chain being encapsulated in professional tools used by engineers in charge of real-life structural monitoring. For lack of an adequate mobilization of the useful multidisciplinary skills, this way remains about unexplored today.

4.2. Electrical cable and network monitoring

The fast development of electronic devices in modern engineering systems comes with more and more connections through cables, and consequently, the reliability of electric connections becomes a crucial issue. For example, in a modern automotive vehicle, the total length of onboard cables has tremendously increased during the last decades and is now up to 4km. These wires and connectors are subject to aging or degradation because of severe environmental conditions. In this area, reliability becomes a safety issue. In some other domains, cable defects may have catastrophic consequences. It is thus a crucial challenge to design smart embedded diagnosis systems able to detect wired connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in electric transmission lines and wired networks. Original methods have been recently developed by Inria, notably based on the inverse scattering theory, for cable and network monitoring. Further developments concern both theoretic study and industrial applications.

4.3. Aeronautics

Improved safety and performance and reduced aircraft development and operating costs are major concerns in aeronautics industry. One critical design objective is to clear the aircraft from unstable aero-elastic vibrations (flutter) in all flight conditions. Opening of flight domain 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). Today, structural flight tests require controlled excitation by ailerons or other devices, stationary flight conditions (constant elevation and speed), and no turbulence. As a consequence, flight domain opening requires a lot of test flights and its costly. This is even worse for aircrafts having a large number of variants (business jets, military aircrafts). A key challenge is therefore to allow for exploiting more data under more conditions during flight tests: uncontrolled excitation, nonstationary conditions.

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.

MATERIALS Team

4. Application Domains

4.1. Homogenization and related problems

Over the years, the 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.

4.2. 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.

4.3. 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 $\mathcal{N}_A \sim 10^{23}$, 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 \mathcal{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 μ 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?”

MATHRISK Project-Team

4. Application Domains

4.1. Application Domains

Risk management, Quantitative finance, Computational Finance, Market Microstructure analysis, Systemic risk, Portfolio optimization, Risk modeling, Option pricing and hedging in incomplete markets, insurance.

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 [87], des problèmes de calcul de débit pour des ateliers [138], pour des réseaux ferroviaires [86] ou routiers, et l'évaluation de performance des réseaux de communication [77]. 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" [81], [147]. Le projet Maxplus collabore avec le projet Metalau, qui étudie particulièrement les applications des modèles max-plus à la modélisation microscopique du trafic routier [155], [151], [116].

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 [87], or computing the throughput of a workshop [138] 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 [86] and [77]. 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 [81], [147]. The Maxplus team collaborates with the Metalau team, working particularly on the applications of max-plus models to the microscopic modelling of road traffic [155], [151], [116].

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 [89]. Ces liens conduisent parfois à de nouveaux algorithmes pour les problèmes de recherche opérationnelle classiques,

comme le problème de circuit de poids moyen maximum [96]. Certains problèmes combinatoires, comme des problèmes de programmation disjonctive, peuvent être décomposés par des méthodes de type max-plus [186]. 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 [125] pour le cas du “jobshop”, une représentation plus simple a été obtenue dans [148] 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 [148]. Finalement, en regardant l’algèbre max-plus comme une limite de l’algèbre classique, on peut utiliser des outils algébriques en optimisation combinatoire [145].

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 [89]. Sometimes, these relations lead to new algorithms for classical Operations Research problems, like the maximal circuit mean [96]. There are also special combinatorial problems, like certain problems of disjunctive programming, which can be decomposed by max-plus type methods [186]. 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 [125] for the jobshop case, a simpler representation was given in [148] 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 [148]. Finally, viewing max-plus algebra as a limit of classical algebra allows to use algebraic tools in combinatorial optimisation [145].

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

L’interprétation abstraite est une technique, introduite par P. et R. Cousot [100], 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:

| | | | |
|-----------------|------|---------|---------------------------------------|
| void main() { | | $x_1 =$ | $[0, 0]$ |
| int x=0; | // 1 | $x_2 =$ | $] - \infty, 99] \cap (x_1 \cup x_3)$ |
| while (x<100) { | // 2 | $x_3 =$ | $x_2 + [1, 1]$ |
| x=x+1; | // 3 | $x_4 =$ | $[100, +\infty[\cap (x_1 \cup x_3)$ |
| } | // 4 | | |
| } | | | |

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 [99], [122], 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) \times (# degrés de liberté du treillis) |
| problème en horizon n | exécution de n pas |
| limite du problème en horizon fini | invariant optimal (borne) |
| itération sur les valeurs | itération de Kleene |

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 [177]). 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 [101], 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 [123], [94], [95],[7]. Une version assez générale de cet algorithme, adaptée au domaine des templates, est décrite dans [122] 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 [100], 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 [99], [122], 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 [177]). 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 [101], 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 [123], [94], [95],[7]. A rather general version of this policy iteration algorithm, adapted to the domain of templates, is described in [122], 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 [98].

English version

Max-plus algebra arises naturally in the computation of similarity scores, in biological sequence comparison. See for instance [98].

MC2 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 ([90]) and P. Vigneaux ([93]). 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 [73]. 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 [56]. 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 [70].

4.3. Cancer modeling

- Specific models : investigation of particular cancers : gliomas (brain tumors), meningioma, colorectal cancers lung and liver metastasis, breast cancer. This is one part of the PhD works of P. Berment, J. Jouganous, G. Lefebvre and post-doc of J. Joie.
- Modelling of electrochemotherapy
- Parameter estimations with the help of low order models : see the PhD of J. Jouganous
- Patient-specific simulations
- Theoretical biology of the metastatic process: dynamics of a population of tumors in mutual interactions, dormancy, pre-metastatic and metastatic niche, quantification of metastatic potential and differential effects of anti-angiogenic therapies on primary tumor and metastases.
- Mathematical models for preclinical cancer research: description and prediction of tumor growth and metastatic development, effect of anti-cancerous therapies

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 [75].
- Simulation of compressible flows on cartesian grids : see the thesis of Gabriele Ottino's Thesis [86], 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 [95]. 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 [68]
- Modeling and numerical simulation of the flow around a real wind turbine. Phd thesis of Xin Jin. This includes reduced order model to design more efficient blades.

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 [74] and Yong-Liang Xiang [97] and CH Bruneau and Iraj Mortazavi [60]. See also project [65] founded by the European Community.
- active control : see the three PhD thesis: M. Buffoni, J. Weller [94], E. Lombardi and FFAST project funded by EU and led by the University of Bristol and AIRBUS UK.
- shape optimization for turbo-machines : See [91].
- 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 [94] and E. Lombardi.
- passive control of flows with porous media : see [62], [59], [58], [85], [63].
- inverse problems in imagery : see [67].

MCTAO Project-Team

4. Application Domains

4.1. Space engineering, satellites, low thrust control

Space engineering is very demanding in terms of safe and high-performance control laws (for instance optimal in terms of fuel consumption, because only a finite amount of fuel is onboard a satellite for all its “life”). 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): orbit transfer rather than station keeping; also we do not discuss attitude control.

In the geocentric case, a space vehicle is subject to

- gravitational forces, from one or more central bodies (the corresponding acceleration is denoted by F_{grav} 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
- other “perturbating” forces (the corresponding acceleration is denoted by F_2 below).

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

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

In the case of a single attracting central body (the earth) and in a geocentric frame, $F_{\text{grav.}}$ does not depend on time, or consists of a main term that does not depend on time and smaller terms reflecting the action of the moon or the sun, that depend on time. The second term is often neglected in the design of the control at first sight; it contains terms like atmospheric drag or solar pressure. G could also bear an explicit dependence on time (here we omit the variation of the mass, that decreases proportionally to $\|u\|$).

4.1.1. Low thrust

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 in general not 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.

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” structure⁰.

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 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

These applications started by a collaboration between B. Bonnard and 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

⁰However, 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.

of the finite dimensional Schrödinger equation, known as Lindblad equation. It is a 3-dimensional system depending upon 3 parameters, yielding a very complicated optimal control problem that we have solved for prescribed boundary conditions. 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 appropriate numerical methods (shooting and numerical continuation) [29], [28].

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 is the “contrast”. This research is conducted with Prof. S. Glaser (TU-München), whose group is performing both in vivo and in vitro experiments; experiments using our techniques have successfully measured the improvement in contrast between materials chemical species that have an importance in medicine, like oxygenated and de-oxygenated blood, see [27]; this is however still to be investigated and improved. 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.

Note that a reference book by B. Bonnard and D. Sugny has been published on the topic [30].

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 [54].
- 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 [56] in the GEOMETRICA project-team. The relevant distance between measures is again the one coming from optimal transportation.
- A collaboration between Ludovic Rifford and Robert McCann from the University of Toronto aims at applications of optimal transportation to the modeling of markets in economy; it was to subject of Alice Erlinger’s PhD, unfortunately interrupted.

Applications *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.

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 [44]; also, closed geodesics for perturbed Riemannian metrics were constructed in [47], [48] using control techniques.

Inside McTAO, a work like [39], [38] is definitely in this line, applying techniques from control to construct some perturbations under constraints of Hamiltonian systems to solve longstanding open questions in the field of dynamical systems. Also, in [61], geometric control is applied successfully to obtain genericity properties for Hamiltonian systems.

MEPHYSTO Team

4. Application Domains

4.1. Mechanics of heterogeneous media

The mechanics of heterogeneous materials aims at characterizing the macroscopic properties of heterogeneous materials using the properties of their constituents.

The homogenization theory is a natural tool for this task. In particular, for linear problems (linear conductivity or linear elasticity), the macroscopic properties are encoded into a single (conductivity or elasticity) homogenized tensor. The numerical approximation of this homogenized tensor is a typical objective of quantitative homogenization.

For nonlinear problems, such as rubber elasticity, the macroscopic properties are no longer characterized by a single tensor, but rather by a nonlinear energy density. Our aim is to relate qualitatively and quantitatively the (precise but unpractical) statistical physics picture to explicit macroscopic constitutive laws that can be used for practical purposes. This endeavor is relevant both in science and technology. The rigorous derivation of rubber elasticity from polymer-physics was indeed emphasized by John Ball as an important open problem of nonlinear elasticity in his survey [40] on the field. Its solution could shed light on some aspects of polymer-physics. The associated *ab initio* derivation of constitutive laws (as an alternative to phenomenological laws) would also be of interest to computational mechanics and rubber industry.

For this application domain, we work in close collaboration with physicists (François Lequeux, ESPCI) and researchers from mechanics and computational mechanics (Patrick Le Tallec, Ecole polytechnique).

4.2. Numerical simulation in heterogeneous media

Solving numerically PDEs in highly heterogeneous media is a problem encountered in many situations, such as the transport of pollutants or the design of oil extraction strategies in geological undergrounds. When such problems are discretized by standard numerical methods the number of degrees of freedom may become prohibitive in practice, whence the need for other strategies.

Numerical solution methods inspired by asymptotic analysis are among the very few feasible alternatives, and started fifteen years ago with the contributions of Hou and Wu [49], Arbogast [37] etc. We refer to [45], [57],[3] for a recent state of the art. Numerical homogenization methods usually amount to looking for the solution of the problem (1) in the form $u_\varepsilon(x) \simeq u_0(x) + \varepsilon \nabla u_0(x) \cdot \Phi(x, \frac{x}{\varepsilon})$, where $\Phi(x, \cdot)$ is a proxy for the corrector field computed locally at point $x \in D$ (in particular, one does not use explicitly that the problem is periodic so that the method can be used for more general coefficients) and u_0 is a function which does not oscillate at scale ε .

Relying on our quantitative insight in stochastic homogenization, a first task consists in addressing the three following prototypical academic examples: periodic, quasi-periodic, and stationary ergodic coefficients with short range dependence. The more ambitious challenge is to address more complex coefficients (of interest to practitioners), and design adaptive and efficient algorithms for diffusion in heterogeneous media.

4.3. Laser physics

Our contribution to the analysis of models in laser physics is motivated by the LabEx CEMPI (Centre Européen pour les Mathématiques, la Physique et leurs Interactions, a large eight-year research and training project approved by the French government in February 2012 as a "Laboratoire d'Excellence" and an initiative of mathematicians and physicists of the Université Lille 1). For this application domain, we work in close collaboration with physicists, which ensures our direct impact on these scientific issues. We focus on two applications: optical fibers and cold atoms.

In collaboration with physicists from the PhLAM laboratory in Lille, we aim at developing new techniques for the numerical integration of a family of 1D Schrödinger-like equations modelling the propagation of laser pulses in optical fibers. The questions arising are challenging since physicists would like to have fairly fast and cheap methods for their problems, with correct qualitative and quantitative behaviors. Another point is that they are interested in methods and codes that are able to handle different physical situations, hence different terms in the NLS equation. To meet these requirements, we will have to use numerical time-integration techniques such as splitting methods or exponential Runge-Kutta methods, space discretization techniques such as finite differences and fast Fourier transforms, and absorbent boundary conditions. Our goal, together with the physicists is to be able to reproduce numerically the results of the experiments they make in actual optical fibers, and then to be able to tune parameters numerically to get more insight into the appearance of rogue waves beyond the dispersive blowup phenomenon.

Recall that the Schrödinger equation also describes Bose-Einstein condensates. A second experimental team at PhLAM projects to study questions related to Anderson localization in such condensates. In fact, they will realize the “kicked rotor” (see [43]), which provides a paradigm for Anderson localization, in a Bose-Einstein condensate. We plan to collaborate with them on the theoretical underpinnings of their findings, which pose many challenging questions.

MISTIS Project-Team

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Aina Frau Pascual, Thomas Vincent, Florence Forbes, Stéphane Girard, Flor Vasseur, Alessandro Chiancone, Farida Enikeeva, Thomas Perret, Pablo Mesejo Santiago.

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 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. We worked on hyperspectral images. In the context of the "pole de competitivite" project I-VP, we worked of images of PC Boards.

4.2. Biology, Environment and Medicine

Participants: Thomas Vincent, Aina Frau Pascual, Florence Forbes, Stéphane Girard, Gildas Mazo, Angelika Studeny, Seydou-Nourou Sylla, Marie-José Martinez, Jean-Baptiste Durand.

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 neurosciences are also considered. Finally, in the context of the ANR VMC project Medup, we studied the uncertainties on the forecasting and climate projection for Mediterranean high-impact weather events.

MODAL Project-Team

4. Application Domains

4.1. Application domains

Potential application areas of statistical modeling 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.

MOKAPLAN Team

4. Application Domains

4.1. Continuous models in economics

- As already mentioned the CFD formulation is a limit case of simple variational Mean-Field Games (MFG) [65]. MFG is a new branch of game theory recently developed by J-M. Lasry and P-L. Lions. MFG models aim at describing the limiting behavior of stochastic differential games when the number of players tends to infinity. They are specifically designed to model economic problems where a large number of similar interacting agents try to maximize/minimize a utility/cost function which takes into account global but partial information on the game. The players in these models are individually insignificant but they collectively have a significant impact on the cost of the other players. Dynamic MFG models often lead to a system of PDEs which consists of a backward Hamilton-Jacobi Bellman equation for a value function coupled with a forward Fokker-Planck equation describing the space-time evolution of the density of agents.
- In microeconomics, the *principal-agent problem* [83] with adverse selection plays a distinguished role in the literature on asymmetric information and contract theory (with important contributions from several Nobel prizes such as Mirrlees, Myerson, Spence or Tirole) and it has many important applications in optimal taxation, insurance, nonlinear pricing. The problem can be reduced to the maximization of an integral functional subject to a convexity constraint This is an unusual calculus of variations problem and the optimal price can only be computed numerically. Recently, following a reformulation of Carlier [12], convexity/well-posedness results of McCann, Figalli and Kim [52], connected to optimal transport theory, showed that there is some hope to numerically solve the problem for general utility functions.
- In [9] a class of games are considered with a continuum of players for which Cournot-Nash equilibria can be obtained by the minimisation of some cost, related to optimal transport. This cost is not convex in the usual sense in general but it turns out to have hidden strict convexity properties in many relevant cases. This enables us to obtain new uniqueness results and a characterisation of equilibria in terms of some partial differential equations, a simple numerical scheme in dimension one as well as an analysis of the inefficiency of equilibria. The mathematical problem has the structure of one step of the JKO gradient flow method.
- Many relevant markets are markets of indivisible goods characterized by a certain quality: houses, jobs, marriages... On the theoretical side, recent papers by Ekeland, McCann, Chiappori [45] showed that finding equilibria in such markets is equivalent to solving a certain optimal transport problem (where the cost function depends on the sellers and buyers preferences). On the empirical side, this allows for trying to recover information on the preferences from observed matching; this is an inverse problem as in a recent work of Galichon and Salanié [57] [58] Interestingly, these problems naturally lead to numerically challenging variants of the Monge-Kantorovich problem: the multi-marginal OT problem and the entropic approximation of the Monge-Kantorovich problem (which is actually due to Schrödinger in the early 30's).

4.2. Finance

The Skorohod embedding problem (SEP) consists in finding a martingale interpolation between two probability measures. When a particular stochastic ordering between the two measures is given, Galichon et al [56] have shown that a very natural variational formulation could be given to a class of problems that includes the SEP. This formulation is related to the CFD formulation of the OT problem [2] and has applications to *model-free bounds of derivative prices in Finance*. It can also be interpreted as a multi marginal Optimal Mass Transportation with infinitely many marginals [78].

4.3. Congested Crowd motion

The volume preserving property appears naturally in this context where motion is constrained by the density of player.

- Optimal Mass Transportation and MFG theories can be an extremely powerful tool to attack some of these problems arising from spatial economics or to design new ones. For instance, various urban/traffic planning models have been proposed by Buttazzo, Santambrogio, Carlier ([10] [40] [32]) in recent years.
- Many models from PDEs and fluid mechanics have been used to give a description of *people or vehicles moving in a congested environment*. These models have to be classified according to the dimension (1D model are mostly used for cars on traffic networks, while 2D models are most suitable for pedestrians), to the congestion effects (“soft” congestion standing for the phenomenon where high densities slow down the movement, “hard” congestion for the sudden effects when contacts occur, or a certain threshold is attained), and to the possible rationality of the agents Maury et al [69] recently developed a theory for 2D hard congestion models without rationality, first in a discrete and then in a continuous framework. This model produces a PDE that is difficult to attack with usual PDE methods, but has been successfully studied via Optimal Mass Transportation techniques again related to the JKO gradient flow paradigm.

4.4. Astrophysics

In [54] and [37], the authors show that the deterministic past history of the Universe can be uniquely reconstructed from the knowledge of the present mass density field, the latter being inferred from the 3D distribution of luminous matter, assumed to be tracing the distribution of dark matter up to a known bias. Reconstruction ceases to be unique below those scales – a few Mpc – where multi-streaming becomes significant. Above 6 Mpc/h we propose and implement an effective Monge-Ampere-Kantorovich method of unique reconstruction. At such scales the Zel’dovich approximation is well satisfied and reconstruction becomes an instance of optimal mass transportation. After discretization into N point masses one obtains an assignment problem that can be handled by effective algorithms with not more than cubic time complexity in N and reasonable CPU time requirements. Testing against N -body cosmological simulations gives over 60% of exactly reconstructed points.

4.5. Image Processing and inverse problems

The Wasserstein distance between densities is the value function of the Optimal Mass Transportation problem. This distance may be considered to have “orthogonal” properties to the widely used least square distance. It is for instance quadratic with respect to dilations and translation. On the other hand it is not very sensitive to rigid transformations, [75] is an attempt at generalizing the CFD formulation in this context. The Wasserstein distance is an interesting tool for applications where distances between signals and in particular oscillatory signals need to be computed, this is assuming one understands how to transform the information into positive densities.

- Tannenbaum and co-authors have designed several variants of the CFD numerical method and applied it to warping, morphing and registration (using the Optimal Mass Transportation map) problems in medical imaging. [86] [30]
- Gabriel Peyre and co-authors [82] have proposed an easier to compute relaxation of the Wasserstein distance (the sliced Wasserstein distance) and applied it to two image processing problems: color transfer and texture mixing.
- Froese Engquist [51] use a Monge-Ampère Solver to compute the Wasserstein distance between synthetic 2D Seismic signals (After some transformations). Applications to waveform inversion and registration are discussed and simple numerical examples are presented.

4.6. Meteorology and Fluid models

In, [34] Brenier reviews in a unified framework the connection between optimal transport theory and classical convection theory for geophysical flows. Inspired by the numerical model proposed in [30], the starting point is a generalization of the Darcy-Boussinesq equations, which is a degenerate version of the Navier-Stokes-Boussinesq (NSB) equations. In a unified framework, he relates different variants of the NSB equations (in particular what he calls the generalized hydrostatic-Boussinesq equations) to various models involving optimal transport and the related Monge-Ampère equation. This includes the 2D semi-geostrophic equations [61] [49] [48] [4] [67] and some fully nonlinear versions of the so-called high-field limit of the Vlasov-Poisson system [73] and of the Keller-Segel system for chemotaxis [63] [44] .

4.7. Mesh motion/Lagrangian methods

The necessity to preserve areas/volumes is an intrinsic feature of mesh deformations more generally Lagrangian numerical methods. Numerical method of Optimal Mass Transportation which preserve some notions of convexity and as a consequence the monotonicity of the computed transport maps can play a role in this context, see for instance [43] [46] [66].

4.8. Density Functionnal Theory (DFT)

The precise modeling of electron correlations continues to constitute the major obstacle in developing high-accuracy, low-cost methods for electronic structure computations in molecules and solids. The article [47] sheds a new light on the longstanding problem of how to accurately incorporate electron correlation into DFT, by deriving and analyzing the semiclassical limit of the exact Hohenberg-Kohn functional with the single-particle density ρ held fixed. In this limit, in the case of two electrons, the exact functional reduces to a very interesting functional that depends on an optimal transport map M associated with a given density ρ . The limit problem is known in the DFT literature with the optimal transport map being called a correlation function or a co-motion function, but it has not been rigorously derived, and it appears that it has not previously been interpreted as an optimal transport problem. The article [47] thereby links for the first time DFT, which is a large and very active research area in physics and chemistry, to optimal transportation theory with a Coulombian repulsive cost. Numerics are still widely open [38].

NACHOS Project-Team

4. Application Domains

4.1. Electromagnetic wave propagation

Electromagnetic devices are ubiquitous in present day technology. Indeed, electromagnetism has found and continues to find applications in a wide array of areas, encompassing both industrial and societal purposes. Applications of current interest include (among others) those related to communications (e.g transmission through optical fiber lines), to biomedical devices (e.g microwave imaging, micro-antenna design for telemedicine, 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. Equally notable and motivating are applications in defence which include the design of military hardware with decreased signatures, automatic target recognition (e.g bunkers, mines and buried ordnance, etc.) propagation effects on communication and radar systems, etc. Although the principles of electromagnetics are well understood, their application to practical configurations of current interest, such as those that arise in connection with the examples above, 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.).

Although many of the above-mentioned application contexts can potentially benefit from numerical modeling studies, the team currently concentrates its efforts on two physical situations.

4.1.1. Microwave interaction with biological tissues

Two main reasons motivate our commitment to consider this type of problem for the application of the numerical methodologies developed in the NACHOS project-team:

- First, from the numerical modeling point of view, the interaction between electromagnetic waves and biological tissues exhibit the three sources of complexity identified previously and are thus particularly challenging for pushing one step forward the state-of-the art of numerical methods for computational electromagnetics. The propagation media is strongly heterogeneous and the electromagnetic characteristics of the tissues are frequency dependent. Interfaces between tissues have rather complicated shapes that cannot be accurately discretized using cartesian meshes. Finally, the source of the signal often takes the form of a complicated device (e.g a mobile phone or an antenna array).
- Second, the study of the interaction between electromagnetic waves and living tissues is of interest to several applications of societal relevance such as the assessment of potential adverse effects of electromagnetic fields or the utilization of electromagnetic waves for therapeutic or diagnostic purposes. It is widely recognized nowadays that numerical modeling and computer simulation of electromagnetic wave propagation in biological tissues is a mandatory path for improving the scientific knowledge of the complex physical mechanisms that characterize these applications.

Despite the high complexity both in terms of heterogeneity and geometrical features of tissues, the great majority of numerical studies so far have been conducted using variants of the widely known FDTD (Finite Difference Time Domain) method due to Yee [55]. In this method, the whole computational domain is discretized using a structured (cartesian) grid. Due to the possible straightforward implementation of the algorithm and the availability of computational power, FDTD is currently the leading method for numerical assessment of human exposure to electromagnetic waves. However, limitations are still seen, due to the rather difficult departure from the commonly used rectilinear grid and cell size limitations regarding very detailed structures of human tissues. In this context, the general objective of the contributions of the NACHOS project-team is to demonstrate the benefits of high order unstructured mesh based Maxwell solvers for a realistic numerical modeling of the interaction of electromagnetic waves and biological tissues with emphasis on

applications related to numerical dosimetry. Since the creation of the team, our works on this topic have mainly been focussed on the study of the exposure of humans to radiations from mobile phones or wireless communication systems (see Fig. 1). This activity has been conducted in close collaboration with the team of Joe Wiart at Orange Labs/Whist Laboratory <http://whist.institut-telecom.fr/en/index.html> (formerly, France Telecom Research & Development) in Issy-les-Moulineaux [18].

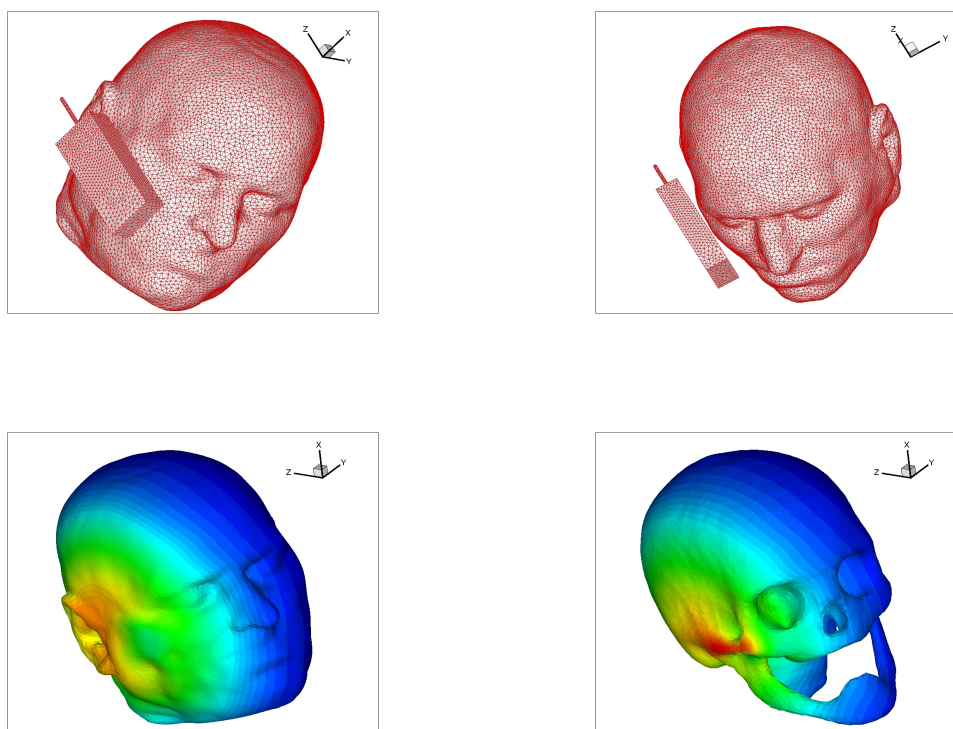


Figure 1. Exposure of head tissues to an electromagnetic wave emitted by a localized source. Top figures: surface triangulations of the skin and the skull. Bottom figures: contour lines of the amplitude of the electric field.

4.1.2. Light/matter interaction on the nanoscale

Nanostructuring of materials has opened up a number of new possibilities for manipulating and enhancing light-matter interactions, thereby improving fundamental device properties. Low-dimensional semiconductors, like quantum dots, enable one to catch the electrons and control the electronic properties of a material, while photonic crystal structures allow to synthesize the electromagnetic properties. These technologies may, e.g., be employed to make smaller and better lasers, sources that generate only one photon at a time, for applications in quantum information technology, or miniature sensors with high sensitivity. The incorporation of metallic structures into the medium add further possibilities for manipulating the propagation of electromagnetic waves. In particular, this allows subwavelength localisation of the electromagnetic field and, by subwavelength structuring of the material, novel effects like negative refraction, e.g. enabling super lenses, may be realized. Nanophotonics is the recently emerged, but already well defined, field of science and technology aimed at establishing and using the peculiar properties of light and light-matter interaction in various nanostructures. Nanophotonics 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. Because of its numerous scientific and technological applications (e.g. in relation to telecommunication, energy production and biomedicine),

nanophotonics represents an active field of research increasingly relying on numerical modeling beside experimental studies.

Plasmonics is a related field to nanophotonics. Metallic nanostructures whose optical scattering is dominated by the response of the conduction electrons are considered as plasmomeric media. If the structure presents an interface with e.g. a dielectric with a positive permittivity, collective oscillations of surface electrons create surface-plasmons-polaritons (SPPs) that propagate along the interface. SPPs are guided along metal-dielectric interfaces much in the same way light can be guided by an optical fiber, with the unique characteristic of subwavelength-scale confinement perpendicular to the interface. Nanofabricated systems that exploit SPPs offer fascinating opportunities for crafting and controlling the propagation of light in matter. In particular, SPPs can be used to channel light efficiently into nanometer-scale volumes, leading to direct modification of mode dispersion properties (substantially shrinking the wavelength of light and the speed of light pulses for example), as well as huge field enhancements suitable for enabling strong interactions with nonlinear materials. The resulting enhanced sensitivity of light to external parameters (for example, an applied electric field or the dielectric constant of an adsorbed molecular layer) shows great promise for applications in sensing and switching. In particular, very promising applications are foreseen in the medical domain [48]- [56].

Numerical modeling of electromagnetic wave propagation in interaction with metallic nanostructures at optical frequencies requires to solve the system of Maxwell equations coupled to appropriate models of physical dispersion in the metal, such the Drude and Drude-Lorentz models. Here again, the FDTD method is a widely used approach for solving the resulting system of PDEs [53]. However, for nanophotonic applications, the space and time scales, in addition to the geometrical characteristics of the considered nanostructures (or structured layouts of the latter), are particularly challenging for an accurate and efficient application of the FDTD method. Recently, unstructured mesh based methods have been developed and have demonstrated their potentialities for being considered as viable alternatives to the FDTD method [51]- [52]- [46]. Since the end of 2012, nanophotonics/plasmonics is increasingly becoming a focused application domain in the research activities of the team in close collaboration with physicists from CNRS laboratories, and also with researchers from international institutions.

4.2. Elastodynamic wave propagation

Elastic wave propagation in interaction with solids are encountered in a lot of scientific and engineering contexts. One typical example is geoseismic wave propagation, in particular in the context of earthquake dynamics or resource prospection.

4.2.1. Earthquake dynamics

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 buildings in high hazard 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 the physical understanding of the earthquake rupture processes and seismic wave propagation. Large-scale simulations of earthquake rupture dynamics and wave propagation are currently the only means to investigate these multiscale 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.

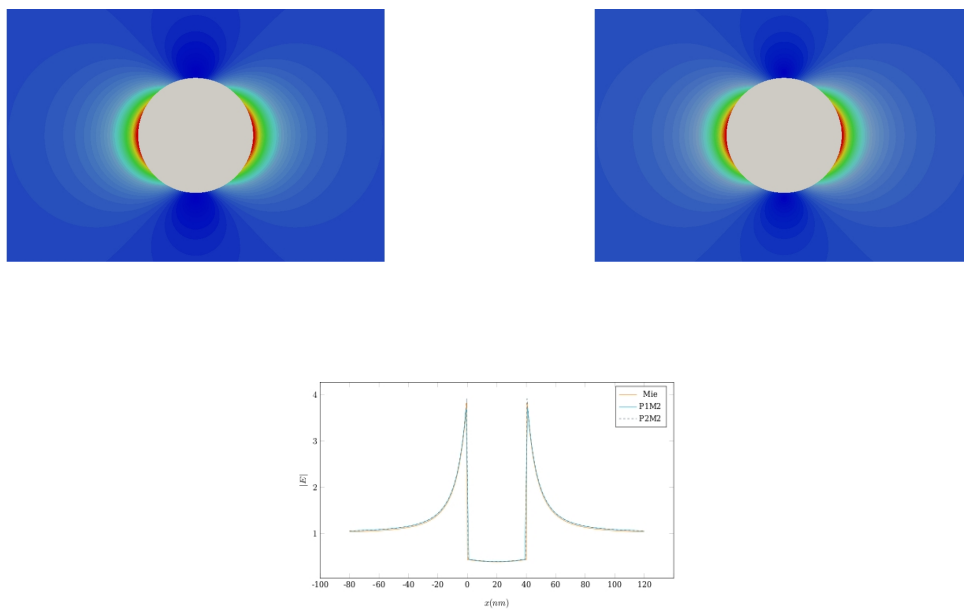


Figure 2. Scattering of a 20 nanometer radius gold nanosphere by a plane wave. The gold properties are described by a Drude dispersion model. Modulus of the electric field in the frequency domain. Top left figure: Mie solution. Top right figure: numerical solution. Bottom figure: 1d plot of the electric field modulus for various orders of approximation (PhD thesis of Jonathan Viquerat).

Numerical methods for the propagation of seismic waves have been studied for many years. Most of existing numerical software rely on finite difference type methods. Among the most popular schemes, one can cite the staggered grid finite difference scheme proposed by Virieux [54] and based on the first order velocity-stress hyperbolic system of elastic waves equations, which is an extension of the scheme derived by Yee [55] for the solution of the Maxwell equations. Many improvements of this method have been proposed, in particular, higher order schemes in space or rotated staggered-grids allowing strong fluctuations of the elastic parameters. Despite these improvements, the use of cartesian grids is a limitation for such numerical methods especially when it is necessary to incorporate surface topography or curved interface. Moreover, in presence of a non planar topography, the free surface condition needs very fine grids (about 60 points by minimal Rayleigh wavelength) to be approximated. In this context, our objective is to develop high order unstructured mesh based methods for the numerical solution of the system of elastodynamic equations for elastic media in a first step, and then to extend these methods to a more accurate treatment of the heterogeneities of the medium or to more complex propagation materials such as viscoelastic media which take into account the intrinsic attenuation. Initially, the team has considered in detail the necessary methodological developments for the large-scale simulation of earthquake dynamics [1]. More recently, the team has initiated a close collaboration with CETE Méditerranée <http://www.cete-mediterranee.fr/gb> which is a regional technical and engineering centre whose activities are concerned with seismic hazard assessment studies, and IFSTTAR <http://www.ifsttar.fr/en/welcome> which is the French institute of science and technology for transport, development and networks, conducting research studies on control over aging, risks and nuisances.

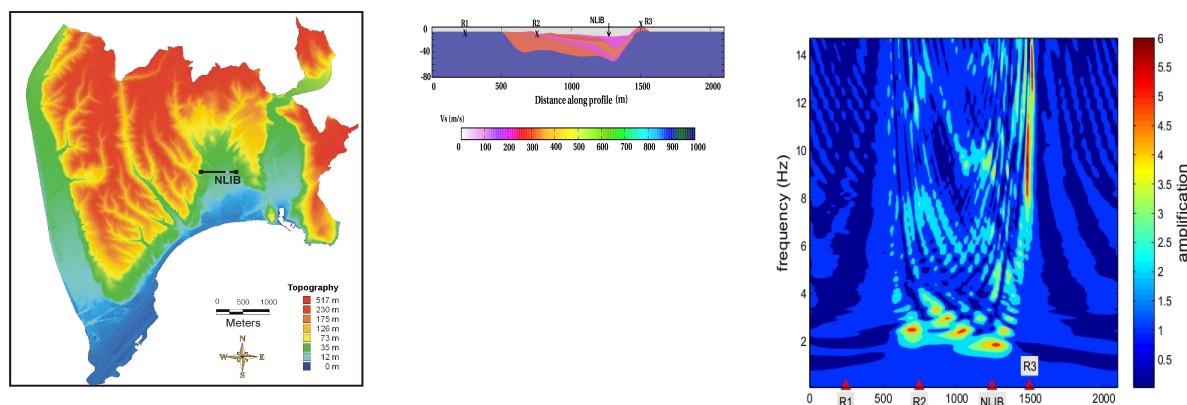


Figure 3. Propagation of a plane wave in a heterogeneous model of Nice area (provided by CETE Méditerranée).

Left figure: topography of Nice and location of the cross-section used for numerical simulations (black line).

Middle figure: S-wave velocity distribution along the cross-section in the Nice basin. Right figure: transfer functions (amplification) for a vertically incident plane wave ; receivers every 5 m at the surface. This numerical simulation was performed using a numerical method for the solution of the elastodynamics equations coupled to a Generalized Maxwell Body (GMB) model of viscoelasticity (PhD thesis of Fabien Peyrusse).

4.2.2. Seismic exploration

This application topic has been considered recently by the NACHOS project-team and this is done in close collaboration with the MAGIQUE-3D project-team at Inria Bordeaux - Sud-Ouest which is coordinating the Depth Imaging Partnership (DIP) <http://dip.inria.fr> between Inria and TOTAL. The research program of DIP includes different aspects of the modeling and numerical simulation of seismic wave propagation that must be considered to construct an efficient software suites for producing accurate images of the subsurface. Our common objective with the MAGIQUE-3D project-team is to design high order unstructured mesh based

methods for the numerical solution of the system of elastodynamic equations in the time-domain and in the frequency domain, that will be used as forward modelers in appropriate inversion procedures.

NANO-D Project-Team

3. Application Domains

3.1. 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.

3.2. 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.

3.3. Pharmaceutics 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.

3.4. Nano-engineering

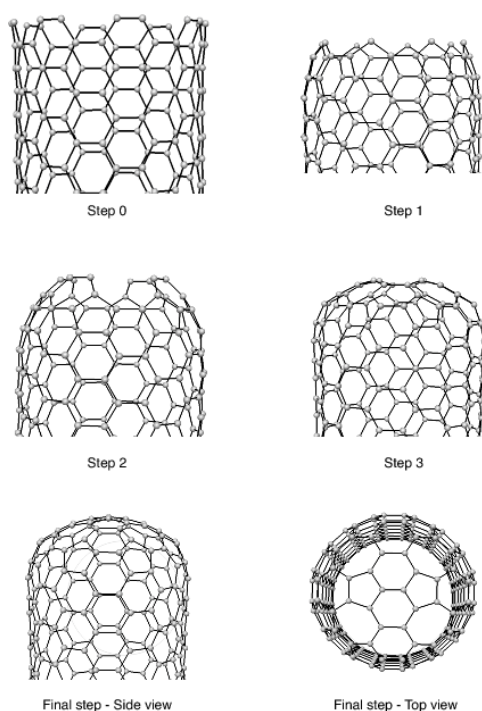


Figure 1. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.

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.

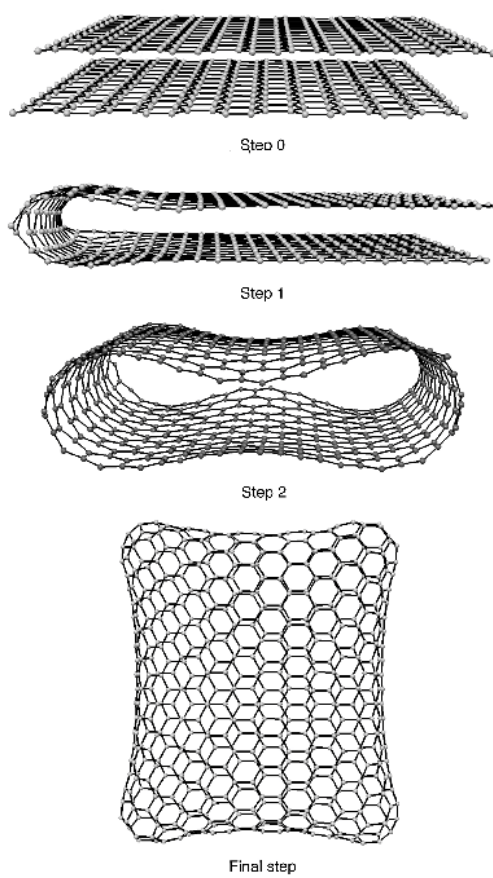


Figure 2. Different steps to prototype a “nano-pillow” with the adaptive interactive modeler.

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. 1) and the “nano-pillow” below (Fig. 2).

NECS Project-Team

4. Application Domains

4.1. A large variety of application domains

Sensor and actuator networks are ubiquitous in modern world, thanks to the advent of cheap small devices endowed with communication and computation capabilities. Potential application domains for research in networked control and in distributed estimation are extremely various, and include the following examples.

- 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.
- Smart grids: the operation of electrical networks is changing from a centralized optimization framework towards more distributed and adaptive protocols, due to the high number of small local energy producers (e.g., solar panels on house roofs) that now interact with the classic large power-plants.
- Disaster relief operations, where data collected by sensor networks can be used to guide the actions of human operators and/or to operate automated rescue equipment.
- Surveillance using swarms of Unmanned Aerial Vehicles (UAVs), where sensor information (from sensors on the ground and/or on-board) can be used to guide the UAVs to accomplish their mission.
- Environmental monitoring and exploration using self-organized fleets of Autonomous Underwater Vehicles (AUVs), collaborating in order to reach a goal such as finding a pollutant source or tracing a seabed map.
- 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 threats.

In particular, NECS team is currently focusing in the areas described in detail below.

4.2. Vehicular transportation systems

4.2.1. 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. The research activity is on two main challenges: forecasting, so as to provide accurate information to users, e.g., travel times; and control, via ramp-metering and/or variable speed limits. The Grenoble Traffic Lab (see Sect. 5.1 and <http://necs.inrialpes.fr/pages/grenoble-traffic-lab.php>) is an experimental platform, collecting traffic infrastructure information in real time from Grenoble South Ring, together with innovative software e.g. for travel-time prediction, and a show-case where to graphically illustrate results to the end-user. This activity is done in close collaboration with local traffic authorities (DIR-CE, CG38, La Metro), and with the start-up company Karrus (<http://www.karrus-its.com/>)

4.2.2. Advanced and interactive vehicle control

Car industry has been already identified as a potential homeland application for Networked Control [44], 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.), loss 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).

More recently, in the ANR project VOLHAND (2009-2013), the team has been developing a new generation of electrical power-assisted steering specifically designed for disabled and aged persons.

Currently, on-going work under a grant with IFPEN studies how to save energy and reduce pollution, by controlling a vehicle's speed in a smart urban environment, where infrastructure-to-vehicle and vehicle-to-vehicle communications happen and can be taken into account in the control.

4.3. Inertial navigation

Inertial navigation is a research area related to the determination of 3D attitude and position of a rigid body. Attitude estimation is based on data fusion from accelerometers, magnetometers and gyroscopes. Attitude is used in general to determine the linear acceleration, which needs to be integrated after to calculate the position. More recently, in the Persyval project LOCATE-ME (2014-2015), the team starts to explore Pedestrian navigation algorithms in collaboration with Tyrex team from INRIA-Rhône-Alpes Center in Montbonnot. The goal behind is to provide guidance e.g. to first responders after a disaster, or to blind people walking in unfamiliar environments. This tasks is particularly challenging indoor, where no GPS is available.

4.4. Multi-robot collaborative coordination

Due to the cost or the risks of using human operators, many tasks of exploration, or of after-disaster intervention are performed by un-manned drones. When communication becomes difficult, e.g., under water, or in spatial exploration, such robots must be autonomous. Complex tasks, such as exploration, or patrolling, or rescue, cannot be achieved by a single robot, and require a self-coordinated fleet of autonomous devices. NECS team has studied the marine research application, where a fleet of Autonomous Underwater Vehicles (AUVs) self-organize in a formation, adapting to the environment, and reaching a source, e.g., of a pollutant. This has been done in collaboration with IFREMER, within the national project ANR CONNECT and the European FP7 project FeedNetBack [11]. On-going research in the team concerns source localization, with a fleet of mobile robots, including wheeled land vehicles.

4.5. Control design of hydroelectric powerplants

We have started a collaboration with ALSTOM HYDRO, on collaborative and reconfigurable resilient control design of hydroelectric power plants. This work is within the framework of the joint laboratory Inria/ALSTOM (see <http://www.inria.fr/innovation/actualites/laboratoire-commun-inria-alstom>). A first concrete collaboration has been established with the CIFRE thesis of Simon Gerwig, who is currently studying how to improve performance of a hydro-electric power-plant outside its design operation conditions, by adaptive cancellation of oscillations that occur in such operation range.

NON-A Project-Team

4. Application Domains

4.1. Networked Robots

Both economically and scientifically, cooperation in robot swarms represents an important issue since it concerns many service applications (health, handicap, urban transports...) and can increase the potential of sensor networks⁰. It involves several challenges such as:

- Because autonomy is a key for being able to increase the network size, maximize the autonomy of the robots in their different tasks of localization, motion, communication;
- Aiming at making 1+1 be more than 2, extend the global potential of the swarm by introducing collaboration (exchanging information with other robots) and cooperation (acting with other robots);
- Include time and energy saving considerations at the design stage. The self deployment of autonomous groups of mobile robots in an unknown environment (including different kinds of static or moving obstacles) involves localization, path planning and robust control problems. Both the control and signal aspects of our researches are oriented to solve some problems coming from - or taking advantage of - such collaboration frameworks. To mention a few:
- Localization using as few as possible landmarks and exteroceptive information by means of derivative estimates;
- Image-based sensing algorithms inspired by our multidimensional estimation techniques;
- Detection and adaptation to sudden loss of communication, time-varying topology, or communication delays;
- Robust, autonomous, energy-aware controllers based on either model-free or model-based techniques.

Several algorithms have already been applied to the control of formations of mobile robots: an illustrative platform is currently developed at EuraTechnologie center within the framework of Non-A⁰. They are now being extended to medical devices (such as wheelchairs) within the European project SYSIASS (see <http://www.sysiass.eu>), in collaboration with partners from hospital settings. Another future application concerns Wireless Sensor and Robot Networks (WSRN, Fig. 2), dedicated to the surveillance of zones, to the exploration of hostile areas, or to the supervision of large scale sensor networks. The main idea here is to integrate mobile nodes (the mobile robots) within the sensor network, allowing to overcome a sensor defection, to maintain the connectivity of the network, or to extend the coverage area during a random deployment. This involves consideration about mobile actuators within a mobile network of sensors and control networks (wireless) with strong constraints on the possibilities of communication in a noisy and non-homogeneous environment. This work is made in close collaboration with the Inria project-team POPS (Lille), which brings its expertise in terms of sensor networks. It takes place in the framework of the Inria ADT SENSAS and represents our contribution to the LABEX proposal ICON.

4.2. Nano/Macro machining

Nano machining

⁰Integrating wireless sensor networks and multi-robot systems increases the potential of the sensors: robots, in comparison, are resource-rich and can be involved in taking decisions and performing appropriate actions on themselves on sensors and/or the environment.

⁰"RobotCity" was exhibited for the first time during the opening ceremony held on April 6th, 2011

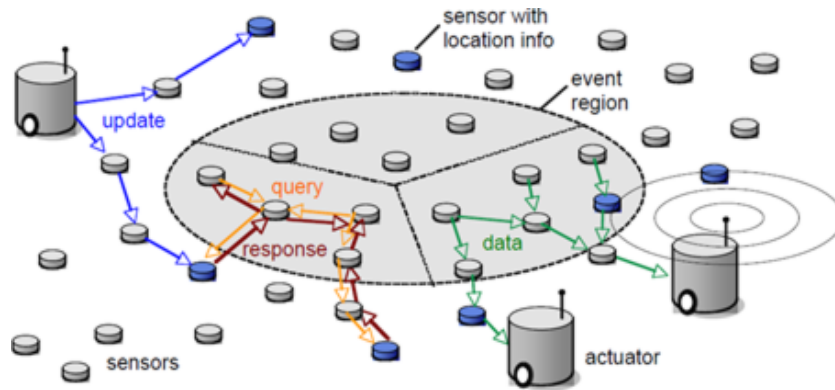


Figure 2. An illustration of collaboration in a Wireless Sensor and Robot Network.

Recent research investigations have reported the development of a number of process chains that are complementary to those used for batch manufacturing of Micro Electro Mechanical Systems (MEMS) and, at the same time, broaden the application domain of products incorporating micro and nano scale features. Such alternative process chains combine micro and nano structuring technologies for master making with replication techniques for high volume production such as injection moulding and roll-to-roll imprinting. In association with the Manufacturing Engineering Center of Cardiff, Arts et Metiers ParisTech center of Lille develops a new process chain for the fabrication of components with nano scale features. In particular, AFM probe-based nano mechanical machining is employed as an alternative master making technology to commonly used lithography-based processes (Fig.). Previous experimental studies demonstrated the potential of this approach for thermoplastic materials. Such a manufacturing route also represents an attractive prototyping solution to test the functionalities of components with nano scale features prior to their mass fabrication and, thus, to reduce the development time and cost of nano technology-enabled products. Application of our control and estimation techniques improves the trajectory tracking accuracy and the speed of the machining tools.

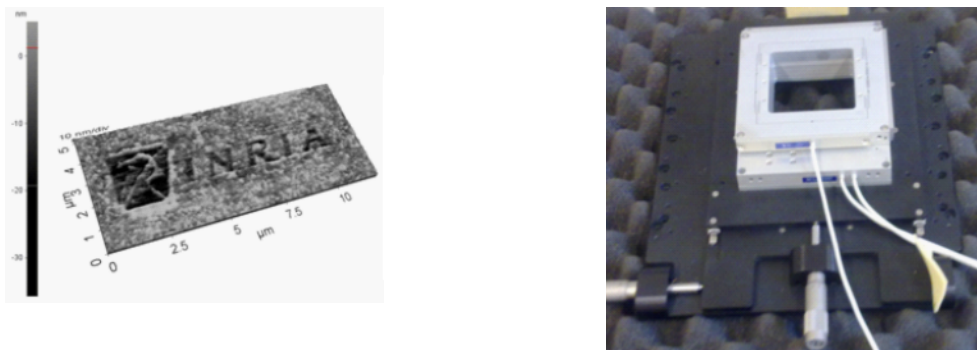


Figure 3. Left: A machined nano structure: $16 \mu\text{m} \times 8 \mu\text{m} \times \text{some nm}$. Right: Nano-positioning system available at Arts et Métiers ParisTech Lille ($75 \mu\text{m}$ range of motion).

Machining with industrial robots

Industrials are enthusiastic to replace machine-tools with industrial robots: compared to machine-tools, industrial articulated robots are very cheaper, more flexible, and exhibit more important workspaces. They can carry out machining applications like prototyping, cleaning and pre-machining of cast parts, as well as end-machining of middle tolerance parts. Such applications require high accuracy in the positioning and path tracking. Unfortunately, industrial robots have a low stiffness and are not that accurate⁰ and they deserve an increased quality of control. We deal with the modelling and the on-line identification of flexible-joint robot models. This can be used both for dynamic simulation and model-based control of industrial robots. We address the problem of real-time identification of the parameters involved in the dynamic linear model of an industrial robot axis. This is possible thanks to a special sensor developed by Arts et Métiers, subject to an EADS project within the FUI (Fonds Unique Interministériel). Control algorithms for other machining actuators such as active magnet bearings are also under study. Within the framework of LAGIS, we also consider the remote control of industrial robots (via internet or Wi-Fi links, for instance), which sets numerous problems in relation with the communication delays.

4.3. Multicell Chopper

On the basis of benchmarks developed at ECS-lab (ENSEA Cergy), we intend to work on the control and observation of serial and parallel multicell choppers, as well as more usual power converters. These power electronic systems associated with their respective loads are typical hybrid dynamical systems and many industrial and/or theoretical challenging problems occur. For example, in the industrial problem of power supply for a supercomputer, the parallel multicell chopper appears as a new solution particularly with respect to the power efficiency. Nevertheless, the observation and control of such hybrid dynamical systems is a difficult task, where non asymptotic estimation and control can be useful.

⁰Industrial robots were designed to realize repeatable tasks. The robot repeatability ranges typically from 0.03 to 0.1mm, but the accuracy is often measured to be within several millimetres. Due to their serial structure, articulated robot has lower stiffness (less than 1 N / mm) than classical machine-tools (greater than 50 N / mm). These poor accuracy and stiffness are caused by many factors, such as geometric parameter errors (manufacturing tolerances), wear of parts and components replacement, as well as flexibility of links and gear trains, gear backlashes, encoder resolution errors and thermal effects.

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. We focus more precisely on developing and validating cell dynamics models. 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. Acoustics

Two particular subjects have retained our attention recently.

Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, has been for our team a very challenging topic, which gave rise to a lot of open questions, from the modeling until the numerical approximation of existing models. Our works in this area are partially supported by EADS and Airbus. The final objective is to reduce the noise radiated by Airbus planes. Musical acoustics constitute a particularly attractive application. We are concerned by 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. We have successively considered the timpani, the guitar and the piano. This activity is continuing in the framework of the European Project BATWOMAN.

4.2. Electromagnetism

Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology and electromagnetic compatibility. These areas are still motivating research in computational sciences (large scale computation) and mathematical modeling (derivation of simplified models for multiscale problems). These topics are developed in collaboration with CEA, DGA and ONERA.

Electromagnetic propagation in non classical media opens a wide and unexplored field of research in applied mathematics. This is the case of wave propagation in photonic crystals, metamaterials or magnetized plasmas. Two ANR projects (METAMATH and CHROME) support this research.

Finally, the simulation electromagnetic (possibly complex, even fractal) networks is motivated by non-destructive testing applications. This topic is developed in partnership with CEA-LIST.

4.3. 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. A major application topic has emerged during the past years : the non destructive testing by ultra-sounds which is the main topic of our collaboration with CEA-LIST. On the other hand, we are developing efficient integral equation modelling for geophysical applications (soil-structure interaction for civil engineering, seismology).

QUANTIC Team

4. Application Domains

4.1. Quantum engineering

A new field of quantum systems engineering has emerged during the last few decades. This field englobes a wide range of applications including nano-electromechanical devices, nuclear magnetic resonance applications, quantum chemical synthesis, high resolution measurement devices and finally quantum information processing devices for implementing quantum computation and quantum communication. Recent theoretical and experimental achievements have shown that the quantum dynamics can be studied within the framework of estimation and control theory, but give rise to new models that have not been fully explored yet.

The QUANTIC team's activities are defined at the border between theoretical and experimental efforts of this emerging field with an emphasis on the applications in quantum information, computation and communication. The main objective of this interdisciplinary team is to develop quantum devices ensuring a robust processing of quantum information.

On the theory side, this is done by following a system theory approach: we develop estimation and control tools adapted to particular features of quantum systems. The most important features, requiring the development of new engineering methods, are related to the concept of measurement and feedback for composite quantum systems. The destructive and partial⁰ nature of measurements for quantum systems lead to major difficulties in extending classical control theory tools. Indeed, design of appropriate measurement protocols and, in the sequel, the corresponding quantum filters estimating the state of the system from the partial measurement record, are themselves building blocks of the quantum system theory to be developed.

On the experimental side, we develop new quantum information processing devices based on quantum superconducting circuits. Indeed, by realizing superconducting circuits at low temperatures and using microwave measurement techniques, the macroscopic and collective degrees of freedom such as the voltage and the current are forced to behave according to the laws of quantum mechanics. Our quantum devices are aimed to protect and process quantum information through these integrated circuits.

⁰Here the partiality means that no single quantum measurement is capable of providing the complete information on the state of the system.

REALOPT Project-Team

4. Application Domains

4.1. Introduction

Our group has tackled applications in logistics, transportation and routing [72], [71], [67], [69], in production planning [93] and inventory control [67], [69], in network design and traffic routing [49], [58], [65], [96], [47], [59], [79], [86], in cutting and placement problems [74], [75], [90], [91], [92], [94], and in scheduling [5], [80], [45].

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 [58]. 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 [98], [97], [96] 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 [95]. 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 [65], [66].

We studied several time dependent formulations for the unit demand vehicle routing problem [51], [50] [30]. 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 [85] 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.

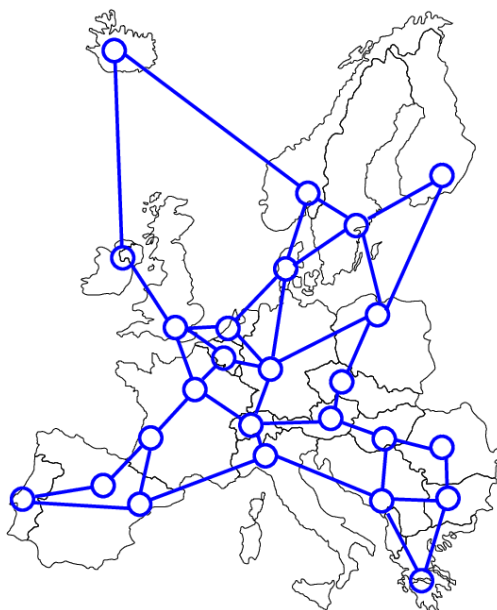


Figure 1. Design of a SDH/SONET european network where demands are multiplexed.

4.3. Packing and Covering Problems

Realopt team has a strong experience on exact methods for cutting and packing problems. These problems occur in logistics (loading trucks), industry (wood or steel cutting), computer science (parallel processor scheduling).

We developed a branch-and-price algorithm for the Bin Packing Problem with Conflicts which improves on other approaches available in the literature [84]. 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.

We also developed a branch-and-price algorithm for a variant of the bin-packing problem where the items are fragile. In [43] we studied empirically different branching schemes and different algorithms for solving the subproblems.

We studied a variant of the knapsack problem encountered in inventory routing problem [69]: we faced a multiple-class integer knapsack problem with setups [68] (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 [62], [60], [64], [63]. 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. We are currently developing exact optimization tools for glass-cutting problems in a collaboration with Saint-Gobain. This 2D-3stage-Guillotine cut problems are very hard to solve given the scale of the instance we have to deal with. Moreover one has to issue cutting patterns that avoid the defaults that are present in the glass sheet the are used as raw material. They are extra sequencing constraints regarding the production that make the problem even more complex.

Finally, let us add that we are now organizing a european challenge on packing with society Renault: see <http://challenge-esicup-2015.org/>. This challenge is about loading trucks under practical constraints. The final results will be announced in March 2015.

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 [70]. 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 [72].

Together with our partner company GAPSO from the associate team SAMBA, we worked on the equipment routing task scheduling problem [78] arising during port operations. In this problem, a set of tasks needs to be performed using equipments of different types with the objective to maximum the weighted sum of performed tasks.

We participated to the project on an airborne radar scheduling. For this problem, we developed fast heuristics [57] and exact algorithms [45]. A substantial research has been done on machine scheduling problems. A new compact MIP formulation was proposed for a large class of these problems [44]. An exact decomposition algorithm was developed for the NP-hard maximizing the weighted number of late jobs problem on a single machine [80]. A dominant class of schedules for malleable parallel jobs was discovered in the NP-hard problem to minimize the total weighted completion time [82]. We proved that a special case of the scheduling problem at cross docking terminals to minimize the storage cost is polynomially solvable [83], [81].

Another application area in which we have successfully developed MIP approaches is in the area of tactical production and supply chain planning. In [42], we proposed a simple heuristic for challenging multi-echelon problems that makes effective use of a standard MIP solver. [41] 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 [46] 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 [53], [52], 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 [48]. We considered train timetabling problems and their re-optimization after a perturbation in the network [55], [54]. 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.

Currently, we are conducting investigations on a real-world planning problem in the domain of energy production, in the context of a collaboration with EDF. The problem consists in scheduling maintenance periods of nuclear power plants as well as production levels of both nuclear and conventional power plants in order to meet a power demand, so as to minimize the total production cost. For this application, we used a Dantzig-Wolfe reformulation which allows us to solve realistic instances of the deterministic version of the problem [31]. In practice, the input data comprises a number of uncertain parameters. We deal with a scenario-based stochastic demand with help of a Benders decomposition method. We are working on Multistage Robust Optimization approaches to take into account other uncertain parameters like the duration of each maintenance period, in a dynamic optimization framework. The main challenge addressed in this work is the joint management of different reformulations and solving techniques coming from the deterministic (Dantzig-Wolfe decomposition, due to the large scale nature of the problem), stochastic (Benders decomposition, due to the number of demand scenarios) and robust (reformulations based on duality and/or column and/or row generation due to maintenance extension scenarios) components of the problem [28].

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 [48]:

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. 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. Risk modelling in finance

- A striking feature of many financial logs is that they are both irregular in the Hölder sense and display jumps. Furthermore, the local roughness as well as the size of jumps typically vary in time. This hints that multifractional multistable processes may provide well-adapted models. As a first step, we shall investigate the simple case of multistable Lévy motions and concentrate on understanding how a time-varying α function translates in terms of risk, in particular for VaR computation. This will require both a deeper understanding of the stochastic properties of these processes and a fine analysis of the microstructure of financial logs.
- In another direction, we will study whether multifractional Brownian motion (mBm) and SRP provide useful models in the frame of financial modeling. Fractional Brownian motion-based option pricing and portfolio selection has attracted a lot of interest in recent years. This process is certainly a more adequate model than pure Brownian motion, as many studies have shown. However, it is also clear that it suffers various limitations. One of the most obvious is that the local regularity of financial logs is not constant, as is apparent on any sufficiently long sample. The most direct way of generalizing fractional Brownian motion to account for this fact is to consider mBm, as we have done in [35], using the theory of stochastic calculus with respect to mBm that we have recently developed in [39], [38]. Another possibility is to use SRP. This requires to extend both the theoretical results (mainly those related to stochastic calculus) and their applications (pricing, portfolio selection) beyond the case of fractional Brownian motion. A disadvantage of mBm is that, in order to price for instance, one has to know the regularity function ahead of time, which usually requires additional assumptions, or to build a model for its evolution. This problem is not present for the SRP: no further information is required once the function relating the amplitude and the regularity has been identified. On the other hand, stochastic integration with respect to SRP (which is neither a Gaussian process nor a semi-martingale) does not seem to be within reach at present, since little is known indeed about this process. This nevertheless constitutes one of our long term goals.

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 and image analysis.

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.

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).

Moreover, a collaboration has started with Dassault Aviation on modal analysis of mechanical structures, which aims at identifying the vibration behavior of structures under dynamic excitations. From algorithmic view point, modal analysis amounts to estimation in parametric models on the basis of measured excitations and structural responses data. As it appears from literature and existing implementations, the model selection problem attached to this estimation is currently treated by a rather heavy and very heuristic procedure. The model selection via penalisation tools are intended to be tested on this model selection problem.

4.4. Dynamic contrast Enhanced imaging

Since Yves Rozenholc joins SELECT, we are involved in quantifying tumor microcirculation to monitor treatments in cancer. Dynamic Contrast Enhanced (DCE) imaging provides information on the qualities of a vascular network. It enables biostatisticians to design biomarkers that can be used for diagnosis, prognosis and treatment monitoring. To make available robust tumoral microcirculation biomarkers in DCE imaging, Yves Rozenholc is developing several tools for denoising and clustering the dynamics found in DCE imaging sequences, to realize in the blood flow model, and testing equality of the survival functions coming from two DCE imaging sequences.

4.5. Analysis of genomic data

Since many years SELECT collaborates 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. Yann Vasseur started a thesis cosupervised by Gilles Celeux and Marie-Laure Martin-Magniette on this topic which is also an interesting investigation domain for the latent block model developed by SELECT. On the other hand, SELECT is involved in ANR “jeunes chercheurs” MixStatSeq directed by Cathy Maugis (INSA Toulouse) which is concerned with Statistical analysis and clustering of RNASeq genomics data.

4.6. Pharmacovigilance

A collaboration has started with Pascale Tubert-Bitter, Ismael Ahmed and Mohamed Sedki (Pharmacoepidemiology and Infectious Diseases, PhEMI) for the analysis of pharmacovigilance data. In this framework, the objective is to detect as soon as possible potential associations between some drugs and adverse effects which appeared after the authorisation marketing of these drugs. Instead of working on aggregated data (contingency table) like it is usually the case, the developed approach aims at dealing with the individual data which perhaps give more information. Valerie Robert started a thesis cosupervised by Gilles Celeux and Christine Keribin on this topic which enables to develop a new model based-clustering inspired of the latent block model.

4.7. Environment

A study has been achieved by Jean-Michel Poggi, Benjamin Auder and Bruno Portier (INSA de Rouen), in the context of a collaboration between AirNormand, Orsay University and INSA of Rouen. It is an application of sequential prediction. To build the prediction, the question is to optimally combine before every term of forecast, the predictions of a set of experts. The study is original not only because of the specific field of application and the adaptation to the concrete context of the work of the air quality monitor in regional agency, but the main originality is that the initial set of experts contains at the same time experts coming from statistical models built by means of different methods and of different predictors and from experts coming from deterministic physico-chemical models. The interest of this kind of sequential prediction method in this specific context is under investigation and the first results on three monitoring stations are promising.

4.8. 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. In Short

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,
- 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 [47]. 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 software: in the 1990's, RL has been the basis of a very successful Backgammon program, TD-Gammon [53] 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 [47]
- 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 [51], [50] 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 Detspiter 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. 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 sometimes 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.

4.5. 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 (section vide)

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 had several related projects (Citines, a European (FP7) project; IOMCA, a ANR project), and we now work on the POST project, a ADEME BIA about investments in power systems. We have a collaboration with a company, Artelys, working on optimization in general, and in particular on energy management; this is a Inria ILAB.

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 6.2 . 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. We study various modelling errors, such as bias due to finite samples, linearization, and propose corrections.

Related Activities:

- 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. In particular, the Crystal tool is adopted by the European Community (<http://www.artelys.com/news/120/90/Energy-The-European-Commission-Chooses-Artelys-Crystal>)
- 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 since 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 (Areski 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. A new formulation of the problem, modeling the plane flows with Bayesian Networks, has been proposed in the Air Traffic Control community in 2013. In 2014, the corresponding stochastic multi-objective optimization problem has been tackled by Evolutionary Algorithms, leading to a general approach to uncertainty handling in Multi-Objective Evolutionary Algorithms [38], [59]. All details in Gaëtan’s PhD [4].

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 models with singular coefficients, with applications in Geophysics, Molecular Dynamics and Neurosciences; Lagrangian modeling in Fluid Dynamics and Meteorology; Population Dynamics, Evolution and Genetics; Neurosciences; and Financial Mathematics.

4.1.1. *Stochastic models with singular coefficients: Analysis and simulation*

Stochastic differential equations with discontinuous coefficients arise in Geophysics, Chemistry, Molecular Dynamics, Neurosciences, Oceanography, etc. In particular, they model changes of diffusion of fluids, or diffractions of particles, along interfaces.

For practitioners in these fields, Monte Carlo methods are popular as they are easy to interpret — one follows particles — and are in general easy to set up. However, dealing with discontinuities presents many numerical and theoretical challenges. Despite its important applications, ranging from brain imaging to reservoir simulation, very few teams in mathematics worldwide are currently working in this area. The Tosca project-team has tackled related problems for several years providing rigorous approach. Based on stochastic analysis as well as interacting with researchers in other fields, we developed new theoretical and numerical approaches for extreme cases such as Markov processes whose generators are of divergence form with discontinuous diffusion coefficient.

The numerical approximation of singular stochastic processes can be combined with backward stochastic differential equations (BSDEs) or branching diffusions to obtain Monte Carlo methods for quasi-linear PDEs with discontinuous coefficients. The theory of BSDEs has been extensively developed since the 1980s, but the general assumptions for their existence can be quite restrictive. Although the probabilistic interpretation of quasi-linear PDEs with branching diffusions has been known for a long time, there have been only a few works on the related numerical methods.

Another motivation to consider stochastic dynamics in a discontinuous setting came to us from time evolution of fragmentation and coagulation phenomena, with the objective to elaborate stochastic models for the avalanche formation of soils, snow, granular materials or other geomaterials. Most of the models and numerical methods for avalanches are deterministic and involve a wide variety of physical parameters such as the density of the snow, the yield, the friction coefficient, the pressure, the basal topography, etc. One of these methods consists in studying the safety factor (or limit load) problem, related to the shallow flow of a visco-plastic fluid/solid with heterogeneous thickness over complex basal topography. The resulting nonlinear partial differential equation of this last theory involves many singularities, which motivates us to develop an alternative stochastic approach based on our past works on coagulation and fragmentation. Our approach consists in studying the evolution of the size of a typical particle in a particle system which fragments in time.

4.1.2. *Stochastic Lagrangian modeling in Computational Fluid Dynamics*

Stochastic Lagrangian models were introduced in the eighties to simulate complex turbulent flows, particularly two-phase flows. In Computational Fluid Dynamics (CFD), they are intensively used in the so-called Probability Density Functions (PDF) methods in order to model and compute the reaction-phase terms in the fundamental equations of fluid motions. The PDF methods are currently developed in various laboratories by specialists in scientific computation and physicists. However, to our knowledge, we are innovating in two ways:

- our theoretical studies are the pioneering mathematical analysis of Lagrangian stochastic models in CFD;
- our work on the Stochastic Downscaling Method (SDM) for wind simulation is the first attempt to solve the fundamental equations themselves by a fully 3D stochastic particle method.

We emphasize that our numerical analysis is essential to the SDM development which takes benefits from our deep expertise on numerical schemes for McKean-Vlasov-non-linear SDEs.

4.1.3. Population Dynamics, Evolution and Genetics

The activity of the team on stochastic modeling in population dynamics and genetics mainly concerns application in adaptive dynamics, a branch of evolutionary biology studying the interplay between ecology and evolution, ecological modeling, population genetics in growing populations, and stochastic control of population dynamics, with applications to cancer growth modeling. Stochastic modeling in these areas mainly considers individual-based models, where the birth and death of each individual is described. This class of model is well-developed in Biology, but their mathematical analysis is still fragmentary. Another important topic in population dynamics is the study of populations conditioned to non-extinction, and of the corresponding stationary distributions, called quasi-stationary distributions (QSD). This domain has been the object of a lot of studies since the 1960's, but we made recently significant progresses on the questions of existence, convergence and numerical approximation of QSDs using probabilistic tools rather than the usual spectral tools.

Our activity in population dynamics also involves a fully new research project on cancer modeling at the cellular level by means of branching processes. In 2010 the International Society for Protons Dynamics in Cancer was launched in order to create a critical mass of scientists engaged in research activities on Proton Dynamics in Cancer, leading to the facilitation of international collaboration and translation of research to clinical development. Actually, a new branch of research on cancer evolution is developing intensively; it aims in particular to understand the role of proteins acting on cancerous cells' acidity, their effects on glycolysis and hypoxia, and the benefits one can expect from controlling pH regulators in view of proposing new therapies.

4.1.4. Stochastic modeling in Neuroscience

It is generally accepted that many different neural processes that take place in the brain do so in the presence of noise. Indeed, one typically observes experimentally underlying variability in the spiking times of an individual neuron in response to an unchanging stimulus, while a predictable overall picture emerges if one instead looks at the average spiking time over a whole group of neurons. Sources of noise that are of interest include ionic currents crossing the neural membrane, synaptic noise, and the global effect of the external environment (such as other parts of the brain).

It is likely that these stochastic components play an important role in the function of both the neurons and the networks they form. The characterization of the noise in the brain, its consequences at a functional level and its role at both a microscopic (individual neuron) level and macroscopic level (network of thousands of neurons) is therefore an important step towards understanding the nervous system.

To this end, a large amount of current research in the neuroscientific literature has involved the addition of noise to classical purely deterministic equations resulting in new phenomena being observed. The aim of the project is thus to rigorously study these new equations in order to be able to shed more light on the systems they describe.

4.1.5. Stochastic modeling in Financial Mathematics

4.1.5.1. Technical Analysis

In the financial industry, there are three main approaches to investment: the fundamental approach, where strategies are based on fundamental economic principles; the technical analysis approach, where strategies are based on past price behaviour; and the mathematical approach where strategies are based on mathematical models and studies. The main advantage of technical analysis is that it avoids model specification, and thus calibration problems, misspecification risks, etc. On the other hand, technical analysis techniques have limited theoretical justifications, and therefore no one can assert that they are risk-less, or even efficient.

4.1.5.2. *Financial Risks Estimation and Hedging*

Popular models in financial mathematics usually assume that markets are perfectly liquid. In particular, each trader can buy or sell the amount of assets he/she wants at the same price (the “market price”). They moreover assume that the decision taken by the trader does not affect the price of the asset (the small investor assumption). In practice, the assumption of perfect liquidity is never satisfied but the error due to liquidity is generally negligible with respect to other sources of error such as model error or calibration error, etc.

Derivatives of interest rates are singular for at least two reasons: firstly the underlying (interest rate) is not directly exchangeable, and secondly the liquidity costs usually used to hedge interest rate derivatives have large variation in times.

Due to recurrent crises, the problem of risk estimation is now a crucial issue in finance. Regulations have been enforced (Basel Committee II). Most asset management software products on the markets merely provide basic measures (VaR, Tracking error, volatility) and basic risk explanation features (e.g., “top contributors” to risk, sector analysis, etc).

4.1.5.3. *Energy and Carbon Markets*

With the rise of renewable energy generation (from wind, waves...), engineers face new challenges which heavily rely on stochastic and statistical problems.

Besides, in the context of the beginning of the second phase (the Kyoto phase) in 2008 of the European carbon market, together with the fact that French carbon tax was scheduled to come into law on Jan. 1, 2010, the year 2009 was a key year for the carbon price modeling. Our research approach adopts the point of view of the legislator and energy producers. We used both financial mathematical tools and a game theory approach. Today, with the third phase of the EU-ETS, that didn't yet start, and the report from the Cour des Comptes (October 2013) that pointed out (among many others point) the lack of mathematical modeling on such carbon market design, we continue our research in this direction.

4.1.5.4. *Optimal Stopping Problems*

The theory of optimal stopping is concerned with the problem of taking a decision at the best time, in order to maximise an expected reward (or minimise an expected cost). We work on the general problem of optimal stopping with random discounting and additional cost of observation.

4.1.5.5. *First hitting times distributions*

Diffusion hitting times are of great interest in finance (a typical example is the study of barrier options) and also in Geophysics and Neurosciences. On the one hand, analytic expressions for hitting time densities are well known and studied only in some very particular situations (essentially in Brownian contexts). On the other hand, the study of the approximation of the hitting times for stochastic differential equations is an active area of research since very few results still are available in the literature.

ABS Project-Team (section vide)

AMIB Project-Team (section vide)

ANGE Project-Team

4. Application Domains

4.1. Overview

Sustainable development and environment preservation have a growing importance and scientists have to address difficult issues such as: management of water resources, renewable energy production, biogeochemistry of oceans, resilience of society w.r.t. hazardous flows,...

As mentioned above, the main issue is to propose models of reduced complexity, suitable for scientific computing and endowed with stability properties (continuous and/or discrete). In addition, models and their numerical approximations have to be confronted with experimental data, as analytical solutions are hardly accessible for these problems/models. A. Mangeney (IPGP) and N. Goutal (EDF) may provide useful data.

4.2. Geophysical flows

Reduced models like the shallow water equations are particularly well-adapted to the modelling of geophysical flows since they are characterized by large time or/and space scales. For long time simulations, the preservation of equilibria is essential as global solutions are a perturbation around them. The analysis and the numerical preservation of non-trivial equilibria, more precisely when the velocity does not vanish, are still a challenge. In the fields of oceanography and meteorology, the numerical preservation of the so-called geostrophic quasi-steady state, which is the balance between the gravity and the Coriolis forces, can significantly improve the forecasts. In addition, the data assimilation is required to improve the simulations and correct the dissipative effect of the numerical scheme.

The sediment transport modelling is of major interest in terms of applications, in particular to estimate the sustainability of facilities with silt or scour, such as canals and bridges. Dredging or filling-up operations are costly and generally not efficient in long term. The objective is to determine a configuration almost stable with the facilities. In addition, it is also important to determine the impact of major events like emptying dam. It is aimed at evacuating the sediments in the dam reservoir and requires a large discharge. However, the downstream impact should be measured in terms of turbidity, river morphology and flood.

4.3. Hydrological disasters

It is a violent, sudden and destructive flow. Between 1996 and 2005, nearly 80% of natural disasters in the world have meteorological or hydrological origins. The main interest of their study is to predict the areas in which they may occur most probably and to prevent the damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worse damages. In addition, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last requires the use of reduced models like the shallow water equations. In urban areas, the flow can be largely impacted by the debris, in particular cars, and required fluid/structure interactions to be well understood. Moreover, underground flows, in particular in sewer, can accelerate and amplify the flow. To take them into account, the model and the numerical resolution should be able to treat the transition between free surface and underground flows.

Tsunamis are another hydrological disaster largely studied. Even if the propagation of the wave is globally well described by the shallow water model in oceans, it is not the case close to the epicenter and in the coastal zone where the bathymetry leads a vertical accretion and produce substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

Whereas the viscous effects can often be neglected in water flows, they have to be taken into account in situations such as avalanches, debris flows, pyroclastic flows, erosion processes,...*i.e.* when the fluid rheology becomes more complex. Gravity driven granular flows consist of solid particles commonly mixed with an interstitial lighter fluid (liquid or gas) that may interact with the grains and decrease the intensity of their contacts, thus reducing energy dissipation and favoring propagation. Examples include subaerial or subaqueous rock avalanches (*e.g.* landslides).

4.4. Biodiversity and culture

Nowadays, simulations of the hydrodynamic regime of a river, a lake or an estuary, are not restricted to the determination of the water depth and the fluid velocity. They have to predict the distribution and evolution of external quantities such as pollutants, biological species or sediment concentration.

The potential of micro-algae as a source of biofuel and as a technological solution for CO₂ fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds. One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps,...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel). Consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs. Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed. However, it is clear that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature and density, evaporation and external forcing.

Many problems previously mentioned also arise in larger scale systems like lakes. Hydrodynamics of lakes is mainly governed by geophysical forcing terms: wind, temperature variations,...

4.5. Sustainable energy

One of the booming lines of business is the field of renewable and decarbonated energies. In particular in the marine realm, several processes have been proposed in order to produce electricity thanks to the recovering of wave, tidal and current energies. We may mention water-turbines, buoys turning variations of the water height into electricity or turbines motioned by currents. Although these processes produce an amount of energy which is less substantial than in thermal or nuclear power plants, they have smaller dimensions and can be set up more easily.

The fluid energy has a kinetic and potential part. The buoys use the potential energy whereas the turbines (hydrolian) are activated by currents. To become economically relevant, these systems need to be optimized in order to improve their productivity. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy.

This is a complex and original issue which requires a fine model of energy exchanges and efficient numerical tools. In a second step, the optimisation of parameters that can be changed in real-life, such as bottom bathymetry and buoy shape, must be studied. Eventually, physical experiments will be necessary for the validation.

ARAMIS Project-Team

4. Application Domains

4.1. Introduction

We develop different applications of our new methodologies to brain pathologies, mainly neurodegenerative diseases, epilepsy and cerebrovascular disorders. These applications aim at:

- better understanding the pathophysiology of brain disorders;
- designing biomarkers of pathologies for diagnosis, prognosis and assessment of drug efficacy;
- developing brain computer interfaces for clinical applications;
- improving the localisation of stimulation targets in Deep Brain Stimulation protocol.

These applications are developed in close collaboration with biomedical researchers of the ICM and clinicians of the Pitié-Salpêtrière hospital.

4.2. Understanding brain disorders

The approaches that we develop allow to characterize anatomical and functional alterations, thus making it possible to study these alterations in different clinical populations. This can provide provide new insights into the mechanisms and progression of brain diseases. This typically involves the acquisition of neuroimaging data in a group of patients with a given pathology and in a group of healthy controls. Measures of anatomical and functional alterations are then extracted in each subject (for instance using segmentation of anatomical structures, shape models or graph-theoretic measures of functional connectivity). Statistical analyses are then performed to identify: i) significant differences between groups, ii) correlations between anatomical/functional alterations on the one hand, and clinical, cognitive or biological measures on the other hand, iii) progression of alterations over time.

We propose to apply our methodologies to study the pathophysiology of neurodegenerative diseases (mostly Alzheimer's disease and fronto-temporal dementia), epilepsy, cerebrovascular pathologies and neurodevelopmental disorders (Gilles de la Tourette syndrome). In neurodegenerative diseases, we aim at establishing the progression of alterations, starting from the early and even asymptomatic phases. In Gilles de la Tourette syndrome, we study the atypical anatomical patterns that may contribute to the emergence of symptoms. In epilepsy, we aim at studying the relationships between the different functional and structural components of epileptogenic networks.

4.3. Biomarkers for diagnosis, prognosis and clinical trials

Currently, the routine diagnosis of neurological disorders is mainly based on clinical examinations. This is also true for clinical trials, aiming to assess the efficacy of new treatments. However, clinical diagnoses only partially overlap with pathological processes. For instance, the sensitivity and specificity of clinical diagnosis of Alzheimer's disease (AD) based on established consensus criteria are of only about 70-80% compared to histopathological confirmation. Furthermore, the pathological processes often begin years before the clinical symptoms. Finally, clinical measures embed subjective aspects and have a limited reproducibility and are thus not ideal to track disease progression. It is thus crucial to supplement clinical examinations with biomarkers that can detect and track the progression of pathological processes in the living patient. This has potentially very important implications for the development of new treatments as it would help: i) identifying patients with a given pathology at the earliest stage of the disease, for inclusion in clinical trials; ii) providing measures to monitor the efficacy of treatments.

The derivation of biomarkers from image analysis approaches requires large-scale validation in well-characterized clinical populations. The ARAMIS team is strongly engaged in such efforts, in particular in the field of neurodegenerative disorders. To that purpose, we collaborate to several national studies (see section Partnerships) that involve multicenter and longitudinal acquisitions. Moreover, ARAMIS is strongly involved in the CATI which manages over 15 multicenter studies, including the national cohort MEMENTO (2000 patients).

4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the base of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor patients' progress and their compliance with the protocol, monitoring that they are actually imagining movements. It also follows that feedback from such a BCI can provide patients with an early reinforcement in the critical phase when there is not yet an overt sign of movement recovery. The BCI approaches that we develop are based on the characterization of the information contained in the functional connectivity patterns. We expect to significantly increase the performance of the BCI system with respect to the sole use of standard power spectra of the activity generated by single local brain areas. Such an improvement will concretely provide the user with a more precise control of the external environment in open-loop BCI tasks and a more coherent feedback in the closed-loop BCI schemes.

4.5. Deep Brain Stimulation

Deep Brain Stimulation (DBS) is a surgical technique, which consists in sending electrical impulses, through implanted electrodes, to specific parts of the brain for the treatment of movement and affective disorders. The technique has been initially developed for otherwise-treatment-resistant patients with essential tremors or Parkinson's disease. Its benefit in other affections, such as dystonia, obsessive-compulsive disorders, Tourette syndrome is currently investigated. The localisation of the stimulation target in specific nucleus in deep brain regions is key to the success of the surgery. This task is difficult since the target nucleus, or the precise sub-territory of a given nucleus is rarely visible in the Magnetic Resonance Image (MRI) of the patients. To address this issue, a possible technique is to personalize a high-resolution histological atlas of the brain to each patient. This personalization is achieved by registering the histological atlas, which consists of an image and meshes of deep brain structures, to the pre-operative MRI of each patient. The registration is currently done by optimally aligning image intensities in the atlas and patient's MRI using a block-matching algorithm. The linear nature of the transform makes the technique robust at the cost of a lack of precision, especially for elderly patients with expanded ventricles. We investigate the use of non-linear registration techniques to optimally align both image intensities and contours of visible structures surrounding the target. We expect to improve the localisation of the target for patients with large ventricles while keeping the method robust in all cases.

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 [77], [81]. 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**

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 collaborations with the La Timone hospital in Marseille.

Subtopics include:

- 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.
- Collaboration with the *Laboratory for Neurobiology of Cognition* in order to develop methods that suit their needs for sophisticated data analysis.

Brain Computer Interfaces (BCI) aim to allow direct control of external devices using brain signals such as measured through EEG. In our project, BCI can be seen as an application of EEG processing techniques, but also as an object of fundamental and applied research as they open the way for more dynamical and active brain cognitive protocols.

We are developing research collaborations with the Neurelec company in Sophia Antipolis (subsidiary of Oticon Medical) and with the leading EEG software company BESA based in Munich. We are conducting a feasibility study with the Nice University Hospital on the usage of BCI-based communication for ALS ⁰ patients.

⁰Nice University Hospital hosts a regional reference center for patients suffering from Amyotrophic Lateral Syndrome.

BAMBOO Project-Team

4. Application Domains

4.1. Domain

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

BEAGLE Project-Team (section vide)

BIGS Project-Team (section vide)

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 members of the team at the Laboratory of Environmental Biotechnology (LBE), 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 members of the team at the Oceanographic Laboratory of Villefranche-sur-Mer (LOV), 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) [81].

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 [6].

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 [105] 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:

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

4.6. Software development

4.6.1. ODIN

We are developing **ODIN**, a software platform for the supervision of bioreactors. ODIN [80] supports the smart management of bioreactors (data acquisition, fault diagnosis, automatic control algorithm,...). This C++ application (working under Windows and Linux) is structured in order to rapidly develop and deploy advanced control algorithms through the use of a Scilab interpreter. It also contains a Scilab-based process simulator (developed jointly with Inria Chile) which can be harnessed for experimentation and training purposes. ODIN is made of different modules which can be distributed along different platforms, and which interact through CORBA.

It has been implemented and validated with four different applications in four different laboratories. A licence with the start-up BioEnTech was signed for remote monitoring of anaerobic digesters.

4.6.2. In@lgae

The simulation platform In@lgae is jointly developed with the Inria Ange team. Its objective is to simulate the productivity of a microalgae production system, taking into account both the process type and its location and time of the year. A first module (Freshkiss) developed by Ange computes the hydrodynamics, and reconstructs the Lagrangian trajectories perceived by the cells. Coupled with the Han model, it results in the computation of an overall photosynthesis yield. A second module is coupled with a GIS (geographic information system) to take into account the meteorology of the considered area (any location on earth). The evolution of the temperature in the culture medium together with the solar flux is then computed. Finally, the productivity in terms of biomass, lipids, pigments together with CO₂, nutrients, water consumption, ... are assessed. The productivity map which is produced can then be coupled with a resource map describing the availability in CO₂ nutrients and land.

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 have experienced remarkable advances with Next Generation Sequencing (NGS), that allow 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. 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 noncoding RNA. Noncoding RNA plays 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 acid 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 on linkage maps and fifteen year old mathematical models. But the usage of computational tools was still limited due to the 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 years a tight collaboration with Probiogem lab on nonribosomal peptides, and has become 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 [14]. 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.

CASTOR Project-Team

4. Application Domains

4.1. Tokamaks

In the conception of the ITER tokamak, several key challenging points have been identified. One of them is the necessity to understand and control the huge thermal loads that are directed to the divertor target plates from the scrape-off layer (SOL) region since they are at the edge of or above what can be handled by today's materials. In the same spirit, the control of ELMs type instabilities that can also result in huge energy losses impacting the plasma facing components is considered as of crucial importance for the ITER program. The optimization of scenarii for designing the discharges of ITER and WEST will be addressed as well as some problems of ionospheric plasma.

CLIME Project-Team

4. Application Domains

4.1. Introduction

The first 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 assess ocean surface circulation, by assimilating fronts and vortices displayed on image acquisitions. Concerning meteorology, the focus is on correcting the 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 improving 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. The ensemble methods allow 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. The two main objectives are: management of combinations of sensor types and deployment modes and 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 that represents 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 of these structures are developed and take into account the underlying physical processes. Image acquisitions are assimilated into these 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.

COFFEE Project-Team

4. Application Domains

4.1. Porous Media

Clearly, the analysis and simulation of flows in porous media is a major theme in our team. It is strongly motivated by industrial partnerships, with Total, GdF-Suez, ANDRA, BRGM, etc. with direct applications in geothermy, geological storages, and oil and gas recovery.

Our research has first dealt with the discretization and convergence analysis of multiphase Darcy flows on general polyhedral meshes and for heterogeneous anisotropic media. We have investigated both the Vertex Approximate Gradient (VAG) scheme using both cell and vertex unknowns and the Hybrid Finite Volume (HFV) scheme using both cell and face unknowns. It is remarkable that the VAG scheme is much more accurate than existing nodal approaches (such as CVFE) for heterogeneous test cases: since it avoids the mixing of different rocktypes inside the control volumes, while preserving the low cost of nodal discretizations thanks to the elimination of cell unknowns without any fill-in. The convergence of the numerical discretizations has been studied for the problem of contaminant transport with adsorption in the case of HFV scheme and for two phase Darcy flows in global pressure formulation using particular VAG or HFV schemes, as well as the more general framework of gradient schemes. To reduce the Grid Orientation Effect, a general methodology is proposed in on general meshes. It is based on the recombination of given conservative fluxes to define new conservative fluxes on a richer stencil. On the same token, we have considered the transport of radionuclides by water in porous media. The question is naturally motivated by security studies of nuclear waste storage. We have dealt with the non linear Peaceman system, set on a heterogeneous domain, typically a layered geological medium. The system couples anisotropic diffusion equation and a diffusion-dispersion equation for the pollutant concentration. We have developed and analyzed a specific DDFV scheme to investigate such flows

4.2. Particulate and mixture flows

We investigate fluid mechanics models referred to as “multi-fluids” flows. A large part of our activity is more specifically concerned with the case 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 modelling of fluidized beds, the dynamic of sprays and in particular biosprays with medical applications, engine fine particles emission... There are many possible modelings 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. It turns out that the leading effect in such models is the drag force. However, the role of other terms, of more or less phenomenological nature, deserves to be discussed (close packing terms, lift term, Basset force...). Of course the fluid/kinetic model is interesting in itself and needs further analysis and dedicated numerical schemes. In particular, in collaboration with the Atomic Energy Commission (CEA), we have proposed a semi-Lagrangian scheme for the simulation of particulate flows, extending the framework established in plasma physics to such flows.

We also think it is worthwhile to identify hydrodynamic regimes: it leads to discuss hierarchies of coupled hydrodynamic systems, the nature of which could be quite intriguing and original, while they share some common features of the porous media problems. We are particularly interested in revisiting the modeling of mixture flows through the viewpoint of kinetic models and hydrodynamic regimes. We propose to revisit the derivation of new mixture models, generalizing Kazhikov-Smagulov equations, through hydrodynamic asymptotics. The model is of “hybrid” type in the sense that the constraint reduces to the standard incompressibility condition when the disperse phase is absent, while it involves derivatives of the particle volume fraction when the disperse phase is present.

4.3. Biological degradation, biofilms formation and algae proliferation

Members of the team have started an original research program devoted to biofilms formation and algae proliferation. We started working on this subject through a collaboration with Roberto Natalini and a group of experts in Firenze interested in preventing damages on historical monuments. It is also motivated by *Ostreopsis* proliferation in the Mediterranean Sea. The multidisciplinary character of this research relies on discussions with researchers of the Oceanography Laboratory in Villefranche-sur-Mer, a leading marine research unit, and the Inria team BIOCORE, led by J-L Gouzé. This research is supported by a ANR-project, led by M. Ribot, and it is the main topic of the PhD thesis of B. Polizzi.

DEMAR Project-Team (section vide)

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 [42] 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. Blood cell functions

(i) *O₂ 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 3) in normal and stress conditions.

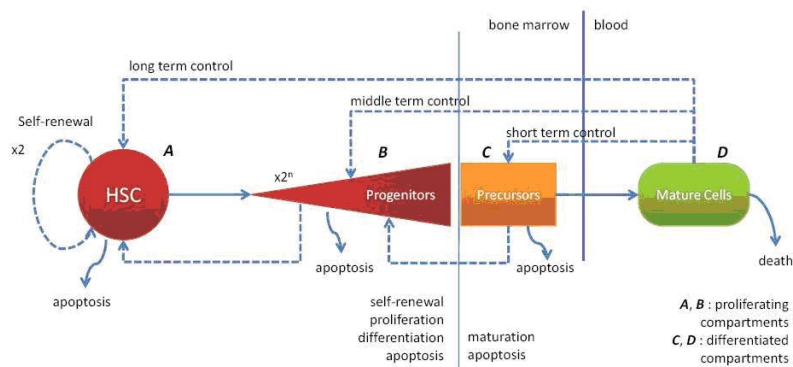


Figure 3. Scheme of Erythropoiesis Modelling ([33]). Without considering explicitly growth factor mediated regulation, all controls (proliferation, self-renewal, differentiation, apoptosis) are mediated by cell populations (dashed arrows). Mature cells can either regulate immature (HSC, progenitors) 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 [46]. A first version of this model is shown in Figure 2 .

(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 ([41], [38]), chronic myelogenous leukemia [43]).

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.

(iii) Coagulation: platelet lineage

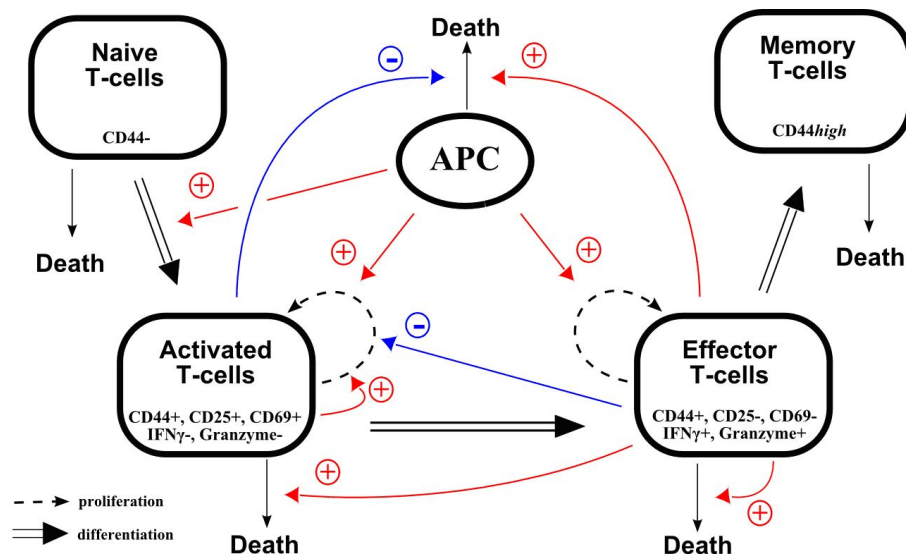


Figure 4. Model of the immune response resulting in the generation of CD8 memory T cells. The response starts with a viral infection resulting in the presentation of viral antigens through antigen presenting cells (APC) to naive T-cells. These latter, once activated, differentiate into activated cells 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.

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. [36]) 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 [44]. 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 Project-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.

Biological data integration. The first axis of the project (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). See Sec. 3.1 .

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. This is the second axis of the project. 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 Sec. 3.2 and Fig. 2). This specific approach is motivated by the quite restricted spectrum of available physiological observations over the asymptotic dynamics of the biological system.

Biological sequence annotation In order to check the accuracy of in-silico predictions, a third research axis of the team 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 Fig. 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 Fig. 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. See Sec. 3.3 for details.

4.2. Application fields

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco*, *memap* and *memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [22], [19]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [28]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [2]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [47]. We are currently initiating two new research programs in this domain: the team will participate to a collaboration program with the Biocore and Ange Inria teams, focused on the understanding on green microalgae; and we will be involved in the deciphering of phytoplankton variability at the system biology level in collaboration with the Station Biologique de Roscoff.

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *bioquali*, *ingranalysis*, *shogun*, *lombarde* were designed in this context [5]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [55]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [9] [27].

Similarly, in **animal biology**, our goal is to propose methods to identify regulators of very complex phenotypes related to nutritional issues. In collaboration with researchers from Inra/Pegase and Inra/Igeep laboratories, we develop methods to distinguish the response of cows, chicken or porks to different diaries or treatments and characterize upstream transcriptional regulators for such a response. The system biology tool *nutritional analyzer* was designed in this framework [14]. The pattern matching tool *logol* also allows for a fine identification of transcription factor motifs [51] [27]. Constraints-based programming also allows us to decipher regulators of reproduction for pea aphids [30], [13]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [20].

We are less involved in **bio-medical applications** as the models and data studied in this application field are well informed and rather data-driven. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [46]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [23] and N. Theret group at Inserm/Irset (Rennes) [15]. The dynamical system tools *caspo* and *cadbiom* were designed within these collaborations. Future studies will focus on the understanding of the metabolism of xenobiotics, still in collaboration with Inserm/Irset (Rennes).

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.

We detail hereafter these two application domains.

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 techniques 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 close 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 many 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 Project-Team

4. Application Domains

4.1. Brain Tumors and Neuro-degenerative diseases

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. Furthermore, in collaboration with the Neurospin center of CEA and the Brookhaven National Laboratory at StonyBrook University we investigate the use of machine learning methods towards automatic interpretation of functional magnetic resonance imaging between cocaine addicted and normal subjects. Last, but not least in collaboration with the Georges Pompidou European Hospital an effort toward understanding tumor perfusion process through comportemental models is carried out with emphasis given on elastic organs.

4.2. Image-driven Radiotherapy Treatment & Surgery Guidance

The use of CT and MR imaging for cancer guidance treatment in collaboration with the Gustave Roussy Institute of Oncology. The aim is to provide tools for automatic dose estimation as well as off-line and online positioning guidance through deformable fusion between imaging data prior to each session and the ones used for scheduling/planning and dose estimation. 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.

4.3. Fundus Image Analysis

Retinal images—also known as fundus images or retinographies—are projective color images of the inner surface of the human eye. In collaboration with Pladema Institute, UNCPBA, Argentina, we are developing a suite of software tools for automatic analysis of retinal images driven by statistical learning approaches.

GENSCALE Project-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)

KALIFFE Project-Team

4. Application Domains

4.1. Plasma Physics and fusion energy

Taking into account spatial effects and time evolution of hot plasmas therefore leads to severe numerical challenges first on modeling aspects and second on purely numerical issues like oscillations and multiscale phenomena. Time scales are very different, ranging from pico-seconds to seconds. This requires special treatment to avoid excessive computing time, as for instance slow/fast manifold decomposition or time averaging. These two difficulties lead to the study and development of new robust numerical schemes and algorithms in order to preserve accuracy and stability within reasonable computing time. To speed up model development the use of refined numerical schemes must be as automatized as possible. Slow/fast manifold should be constructed automatically from the model, and spatial discretization should be as transparent as possible in order to apply former works (semi-Lagrangian or particle methods for transport, spectral or finite difference methods for collisions) on this topic.

Our program can be split into three different tasks :

- derivation of new collision operators based both on Coulombian interactions and strong external forces, approximation of collisional operators for multi-species by developing fast algorithms ;
- numerical treatment of multi-scale problems applying AP schemes to problems taking into account electromagnetic effects and collisions ;
- hybrid methods to take advantage of different regimes and reduce the computational cost.

Approximation of collision operators in plasma physics & conception of softwares for collisional plasmas.

An important step in the understanding of high temperature and dense plasmas is to investigate the effect of weakly Coulombian interactions, namely the Landau or Landau-Fokker-Planck collision operator. Due to the high temperature, collisions between particles have been neglected most of the time, but for the long time simulations, it seems that collisions may contribute and induce some nonlinear effects stabilizing the plasma. Furthermore, for inertial and magnetic confinement fusion, classical collision operators are no longer valid since their derivation, based on microscopic interactions only take into account self-interactions but no external forces, which are not negligible in our applications. There are only few works of physicists on this topic in 80's.

Then, our objective is to derive such operators to describe collisional plasmas and to simulate the transport of classical, as well as relativistic electrons, within a multi-species plasma, containing mobile electrons and ions. Issues to be addressed on this topic involve the derivation of multi-scale models due to different scales of effective constants, spatial heterogeneity and strength of boundary conditions.

Moreover, because of the quadratic aspect of the kernel and the multiple integrations in its analytical formulation, the Landau-Fokker-Planck equation is complicated to compute even if fast algorithm are available $O(N \log N)$, where N is the number of degree of freedom. Henceforth, different simpler models have been introduced, especially the BGK model which is mainly a relaxation towards a Maxwellian equilibrium state, or the linear Fokker-Planck which is a diffusive operator or a nonlinear Fokker-Planck operator taking into account Coulombian interactions. Although, these operators describes correctly the hydrodynamical limit, they usually do not give the correct transport coefficient in the Chapman-Enskog expansion. Our interest here is to compare the different operators in the description of binary collisions between ions-electrons or electrons-electrons and to select the one which is adapted with respect to the physical situation.

Collisional plasma and fluid models One characteristic of plasma physics problems is that they involve many different phenomena (instabilities, saturation phenomena due to nonlinear effects which couple different modes), many different time (plasma frequency, gyrokinetic frequency, etc) and space scales. Splitting a model in sub-models and studying their interactions is a central point, leading to new questions: how to define sub-models? How to simplify or complexify them?

For instance, the interaction of intense lasers with solid matter generates a hot plasma state that is well described by the Vlasov-Maxwell equation at the ignition point, whereas collective effects (electromagnetic fields) and collisions have to be taken into account around the impact and fluid models are sufficient in the capsule (see Figure 1). Accurate and efficient modeling of the physics in these scenarios is highly pertinent, because it relates to experimental campaigns to produce energy by inertial confinement fusion on facilities such as the Laser Méga-Joules in Bordeaux. Calculations involving the Vlasov-Fokker-Planck equation are computationally intensive, but are crucial to proper understanding of a wide variety of physical effects and instabilities in inertial fusion plasmas.

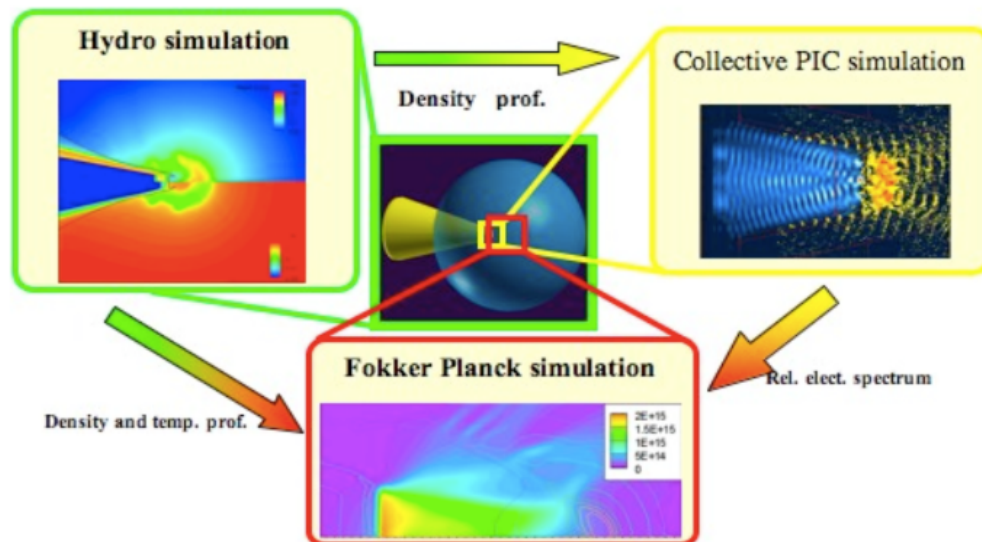


Figure 1. Multiscale modeling at higher laser intensities corresponding to the fast ignition approach for Inertial Confinement Fusion : a relativistic treatment should be considered and collision operators with a large energy exchange are required.

One of the main challenges from the numerical point of view is to propose a general methodology to design macroscopic fluid models that take into account localized kinetic up-scaling effects (which represents the meso-scale). One approach will consist in considering fluid models, which are solved in the whole domain together with a localized kinetic upscaling that corrects the fluid model wherever it is necessary (non-equilibrium events occurring in the flow). This upscaling is obtained by solving a kinetic equation on the non-equilibrium part of the distribution function. This equation is solved only locally and is related to the fluid equation through a downscaling effect. We want to demonstrate that this approach applies to problems that have a hydrodynamic time scale as well as to problems with diffusion time scale.

The project will therefore combine physical modeling and mathematical analysis in order to achieve an understanding and propose a model of the plasma behavior over the various scales involved. The milestones involved in this project are therefore:

- set up a new phenomenology allowing to describe non-local effects in any geometry, and based on a state-of-the-art of mathematical modeling of hot plasmas;
- set up a multiscale model of the physical mechanisms at play at the different scales (micro, meso and macro scales), with a crucial emphasis on the connections between the scales ;
- propose a mathematical analysis and numerical development of the models, and provide systematic derivations of the connections between the scales.

This program therefore organizes naturally over the various scales at play in the problem and their connections: macroscale phenomenology ; kinetic at mesoscales ; statistical behavior at microscales.

LEMON Team

4. Application Domains

4.1. Coastal Oceanography

Participants: Arnaud Duran, Fabien Marche, Antoine Rousseau.

Saint-Venant and Boussinesq equations have been widely applied until recently to model and simulate the propagation and transformations of waves in the nearshore area, over rapidly varying topography. However, the first equations do not include dispersive effects, and consequently have a domain of validity limited to the surf zone. The second set of equations overcome the limitations of the SV equations but relies on a “small amplitude assumption” and is therefore unable to model the whole range of waves transformations. This is the reason why they are usually called “weakly nonlinear Boussinesq equations”. A better suited set of equations is known as the Green-Naghdi equations, but until recently, they have received far less attention, both from the theoretical and numerical point of view. In particular, there is no available numerical method of arbitrary order for 2d simulations on unstructured meshes. Additionally, the construction of rigorous positive preserving schemes is a paramount for the study of waves run-up.

4.2. Urban Floods

Participant: Vincent Guinot.

Floods have been identified by the National Accounting Authority (Cour des Comptes) to represent up to 1% of the GNP in terms of damage cost. For crisis management purposes, modeling urban floods at the scale of the conurbation is highly desirable. This however cannot be achieved in the current state of technology because of the meshing and computational cost (5569 up to one billion cells being needed to mesh an entire urban area). This can be overcome by upscaling the shallow water equations so as to obtain large scale models that can operate three orders of magnitude faster than refined 2D models. Various upscaled versions of the upscaled 2D Shallow Water Equations have been proposed in the literature, some of which by members of the Lemon team. Further developments are being carried out, including the subgrid-scale description of topography variations and a better representation of energy dissipation terms. Laboratory experiments are also needed to discriminate between the various existing models.

4.3. River Hydraulics

Participants: Vincent Guinot, Antoine Rousseau.

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 required accuracy, 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.

LIFEWARE Team

4. Application Domains

4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. We work mainly on eukaryotic cells. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Moreover, one important application of our research is the development of a **modeling platform** for systems biology.

4.2. Modeling platform for systems biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (**BIOCHAM**), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

4.3. Couplings between the cell cycle and circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characteristics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR Hyclock project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio **C5Sys** and FP6 **TEMPO** projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints⁰. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

4.4. Biosensor design and implementation in non-living vesicles

In collaboration with Franck Molina (CNRS, Sysdiag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan), we ambition to apply our techniques to the design and implementation of biosensors in non-living vesicles for medical applications. Our approach is based on protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sysdiag, which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis is also particularly appealing for security considerations in synthetic biology and compliance to forthcoming EU regulation.

⁰Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. *Theoretical Computer Science*, 412(21):2108–2127, 2011.

M3DISIM Team

4. Application Domains

4.1. Clinical applications

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [9]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [1].

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.

4.2. Modeling of Multiperforated plates in turboreactors

In the turbo-engine, the temperature can reach 2000 K inside the combustion chamber. To protect its boundary, "fresh" air at 800 K is injected through thousands of perforations. The geometry of the network of perforations is chosen in order to optimize the cooling and the mechanical properties of the chamber. It has been experimentally observed that these perforations have a negative impact on the stability of the combustion. This is due to the interaction with an acoustic wave generated by the combustion. Due to the large number of holes (2000) and their small sizes (0.5 mm) with respect to the size of the combustion chamber (50 cm), it is not conceivable to rely on numerical computations (even with supercomputers) to predict the influence of these perforations.

In collaboration with ONERA, we develop new models which allow to take into account these multiperforated plates at the macroscopic scale.

4.3. Helioseismology

This collaboration with the Max Planck Institute for solar system, which started in 2014, aims at designing efficient numerical methods for the wave propagation problems that arise in helioseismology in the context of inverse problems. The final goal is to retrieve information about the structure of the sun i.e. inner properties such as density or pressure via the inversion of a wave propagation problem. Acoustic waves propagate inside the sun which, in a first approximation and regarding the time scales of physical phenomena, can be considered as a moving fluid medium with constant velocity of motion. Some other simplifications lead to computational saving, such as supposing a radial or axisymmetric geometry of the sun. Aeroacoustic equations must be

adapted and efficiently solved in this context, this has been done in the finite elements code Montjoie 5.2 . In other situations, a full 3D simulation is required and demands large computational resources. Ultimately, we aim at modeling the coupling with gravity potential and electromagnetic waves (MHD equations) in order to be able to recover sun spots.

MAGNOME Project-Team

4. Application Domains

4.1. Function and history of 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. 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 <http://www.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. 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.

Comparative genomics is a powerful tool for strain selection even when genetic engineering must be excluded. Large-scale comparison of the genomes of experimentally characterized strains can be used to identify quantitative trait loci, which can be used as markers in selective breeding strategies. Identifying individual SNPs and predicting their effect can lead to better understanding of the function of genes implicated in improved strain performance, particularly when those genes are naturally mutated or are the result of the transfer of genetic material from other strains. And understanding the combined effect of groups of genes or alleles can lead to insight in the phenomenon of heterosis.

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.

MAMBA Team

4. Application Domains

4.1. Cancer modelling

Evolution of healthy or cancer cell populations under environmental pressure; drug resistance. Considering cancer as an *evolutionary disease* – evolution meaning here Darwinian evolution of populations structured according to relevant phenotypes – in collaboration with our biologist partners within the Institut Universitaire de Cancérologie (IUC) of UPMC, we tackle the problem of understanding and limiting a) evolution from pre-malignancy to malignancy in cell populations, and b) in established cancer cell populations, evolution towards (drug-induced) drug resistance. Environmental pressure guiding evolution is of various natures, including signalling molecules induced by the peritumoral stroma (e.g., between a breast tumour and its adipocytic stroma), and anticancer drugs and their effects on both the tumour and its stromal environment. The models we use [59], [58] [40] are akin to models used in ecology for adaptive dynamics.

Multi-scale modelling of EMT. The major step from a benign tumour that can be eradicated by surgery and an invasive cancer is the development step at which cells detach from the tumour mass and invade individually the surrounding tissue⁰. The invasion is preceded by a transition (called EMT - epithelial mesenchymal transition) of the cancer phenotype from an epithelial type to a mesenchymal type cell. We so far worked on multi-scale modelling of EMT⁰, and the step by which invading cancer cells enter blood vessels, called intravasation⁰. We now perform in-vitro simulations of cancer cell invasion for Non Small Cell Lung Cancer (NSCLC) having a 5-year survival fraction of about 20%, and for breast cancer. Under development (in collaboration with our biologist partners within the IUC for the experimental part) is also a phenotype-structured PDE model of the interactions between colonies of MCF7 breast cancer and adipocyte stromal support populations.

4.2. Cancer therapies and their optimisation

Drugs. We focus on multi-drug multi-targeted anticancer therapies aiming at finding combinations of drugs that theoretically minimise cancer cell population growth with the constraint of limiting unwanted toxic side effects under an absolute threshold (i.e., this is not L^2 nor L^1 , but L^∞ optimisation) in healthy cell populations and avoiding the emergence of resistant cell clones in cancer cell populations [49], [58] [4], [40]. Prior to using optimisation methods, we design models of the targeted cell populations (healthy and tumour, including molecular or functional drug targets [48]) by PDEs or agent-based models [3], and molecular pharmacological (pharmacokinetic-pharmacodynamic, PK-PD) models of the fate and effects of the drugs used, usually by ODE models. A particular aspect of such modelling is the representation of multi-cellular spatio-temporal patterns emerging from therapies.

Radiotherapy. Radiation is still a major treatment in cancer. We have recently published new results on this topic, please refer to the *New Results* section.

⁰Weinberg, The biology of cancer, Garland, 2007

⁰Ramis-Conde, Drasdo, Anderson, Chaplain, Biophys. J., 2008

⁰Ramis-Conde, Chaplain, Anderson, Drasdo, Phys. Biol. 2009

4.3. Cell motion

Several processes are employed by cells to communicate, regulate and control their movements and generate collective motion. Among them, chemotaxis is the phenomenon by which cells direct their active motion in response to an external chemical (or physical) agent. In chemotaxis, cells not only respond but can also produce the chemical agent, leading to a feedback loop. Understanding this phenomenon is a major challenge for describing the collective behaviour of cells. Many mathematical models have been proposed at different scales, yielding a good description of cell aggregation. In collaboration with biophysicists at Institut Curie in Paris, we develop and study ⁰ mathematical models based on kinetic equations for bacterial travelling waves in a microchannel. These models have shown a remarkable quantitative agreement with experimental observations.

Cell motion arises also in the growth of solid tumours, which can be described through cell population models or multiphase flows ⁰. This is a very active subject because several bio-chemico-physical mechanisms are at work; for instance motion can arise from pressure forces resulting from cell divisions and from active cell motility. At the smaller scale stochastic agent-based models of tumour cells invading the tumour environment or blood vessels are considered ⁰, and allow to include detailed behaviours and interactions. At a larger scale, free boundary problems are widely used, e.g. for image-based prediction because of the reduced number of parameters ⁰. Asymptotic analysis makes a link between these different mechanistic models [63].

One other setting where we will study cell motion is epithelial gap closure, a form of collective cell migration that is a very widespread phenomenon both during development and adult life - it is essential for both the formation and for the maintenance of epithelial layers. Due to their importance, wound healing in vivo and morphogenetic movements involving closure of holes in epithelia have been the object of many studies (including some involving members of this project like [47]). Several theoretical models have also been proposed recently for the advancement of tissue covering unoccupied areas (see, for instance, [46]). It is particularly interesting to study epithelial gap closure in vivo. However, the complexity of the process and the difficulty to measure relevant quantities directly and to control the parameters in vivo, lead people to seek alternative systems where epithelial gap closure can be studied under better-defined and better-controlled conditions.

4.4. Contraction of acto-myosin structures in morphogenesis and tissue repair

In 2014, L. Almeida, I. Cheddadi, C. Emako-Kazianou, P. Bagnerini ⁰, A. Jacinto ⁰, P. Patricio ⁰, B. Ladoux ⁰ and N. Gov ⁰ have continued to investigate the dependence of physical and biological mechanisms of actomyosin cable formation and wound closure depending on the geometry of the wound and adhesion to the substrate, with particular emphasis on the effect of the wound edge curvature. We extended our work from in vivo studies to in vitro situations taking advantage of a collaboration with the group of Benoît Ladoux who did experiments on cell monolayers of human keratinocytes and of MDCK cells. We could single out some similar geometry dependence of the wound closure strategies between these two settings indicating the existence of conserved mechanisms that should be very general across living beings.

In our model under development, we consider viscous behaviour and friction in the tissue plus boundary terms associated to cable and lamellipodial forces. The numerical simulations obtained using this model are in good agreement with the experimental results. This work is attracting considerable attention from the community.

⁰N. Bournaveas, V. Calvez, S. Gutiérrez and B. Perthame, Global existence for a kinetic model of chemotaxis via dispersion and Strichartz estimates, *Comm. PDE*, 2008

⁰J. Ranft et al, Fluidization of tissues by cell division and apoptosis, *PNAS*, 2010 and L. Preziosi and A. Tosin, Multiphase modelling of tumour growth and extracellular matrix interaction: mathematical tools and applications, *J. Math. Biol.*, 2009.

⁰I. Ramis-Conde et al., *J. Phys. Biol.*, 2009

⁰Works by O. Saut, T. Colin, A. Iollo, N. Ayache, J. Lowengrub

⁰Univ. Genova

⁰CEDOC, Lisbon

⁰CEDOC, Lisbon

⁰Inst. J. Monod

⁰Weizmann Institute

4.5. Protein polymerisation

Protein polymerisation is the key feature of amyloid diseases, among which we can quote Alzheimer's, Prion (in particular variant Creutzfeldt-Jakob disease, epidemically linked to bovine spongiform encephalopathy, or so-called "mad cow", disease), Parkinson's, Huntington's diseases. However, the mechanisms of polymerisation are far from being quantitatively understood by biologists. They can be modelled with the help of coagulation-fragmentation equations, a field of expertise of MAMBA [51], [50]. Hence the European starting grant SKIPPER^{AD}, which follows the ANR project TOPPAZ, came up very naturally from the encounter with Human Rezaei, a biologist expert in amyloid diseases at INRA Jouy-en-Josas. Moreover, this field of applications brings new questions to us, which is both a stimulation for our mathematical research and a very promising tool for the biologists.

4.6. Physics of tissue organisation

Many new insights in the last years indicate that migration, growth and division of cells are largely impacted by cell and tissue mechanics (0, 0, 0). Centre-based growth models already permit to account for many of the observed phenomena (e.g. 0, 0). They furthermore permit calculation of the stress tensor in the tissue. ABMs resolving cells at higher resolution 0 permit to calculate cell deformation as a response of stress emerging in the tissue, hence the stress tensor cannot only be resolved at the position of the cell centre, as in the case of centre-based models, but in this case at any point on the cell surface or inside the cell. This permits relating stress and strain in tissues and the deformation and stress a cell feels at subcellular scale. We extended a deformable cell model towards cell-division which enables us to calculate precise stress - strain relationships for cells, that later can be used to calibrate forces in center-based models. This is fundamental to understand the impact of mechanical stress on cell cycle progression or other cell decisions. Moreover, we established a model to explain the proliferation pattern of cells growing in closed capsules.

4.7. Liver modelling

Liver is the main detoxifying organ of the human body and can regenerate up to about 70% of its mass. It pursues its task due to a complex tissue architecture, with hepatocytes aligning along micro-capillaries and forming a dense network. Incidence rate of liver diseases are steadily increasing, liver cancer ranks 6th among all cancers. About one person in 12 suffer from viral hepatitis, which makes 500 million people worldwide. Hepatitis B and C as well as misuse of drugs or alcohol are major reasons to develop liver cancer. Notwithstanding the importance of this public health problem, disease pathogenesis and regeneration in liver are still not well understood.

So far systems biology approaches addressing the tissue scale are rare. Most of those which do so base on compartment models (e.g. 0); only recently are being developed approaches addressing the tissue scale ([57], 0, 0, 0, 0). We are developing a multi-scale model of liver regeneration representing the tissue architecture, the different cell types, the flow systems, hepatocyte metabolism and signal transduction controlling cell cycle entrance in the regeneration processes, taking into account extrahepatic compartments when relevant. Applications are regeneration after drug-induced damage and after partial hepatectomy, drug pharmacodynamics and pharmacokinetics in liver and liver cancer, and model-based prediction of in-vivo drug toxicity from in-vitro measurements 0. The research work is performed within the EU project NOTOX, the BMBF project VIRTUAL LIVER NETWORK and the ANR project IFLOW.

⁰Ingber, Proc. Natl. Acad. Sci (USA), 2005

⁰Trepast et. al., Nat. Phys. 2009

⁰Alessandri et. al., Proc. Natl. Acad. Sci. (USA) 2013

⁰Drasdo and Hoehme, Phys. Biol. 2005

⁰Drasdo and Hoehme, New Journal of Physics 2012

⁰Odenthal, Smeets, van Liedekerke, et. al., PloS Comput Biol. 2013

⁰Diaz-Ochoa et. al. Frontiers in Pharmacology, 2013

⁰Ricken, Dahmen, Dirsch, Biomech. Model. Mechanobiol. 2010

⁰Debbaut et. al., J. Biomech. Eng. 2014

⁰Siggers, Leungchavphongse, Ho, Repetto, Biomech. Model. Mechanobiol. 2014

⁰Schwen et. al., PloS Comput. Biol. 2014

⁰Godoy et al., Arch Toxicol. 2013 Aug;87(8):1315-1530

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. During the last decade, 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.

MNEMOSYNE Project-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. Wastewater treatment systems

The water resources of our planet are limited, and today the quality of drinking water is considered to be responsible of more human deaths than malnutrition. Pollution and over-exploitation of water resources affect almost all the water reservoirs on Earth. Preserving the quality of water has thus become a worldwide problem. The industry of decontamination is thus a necessity, but waste-water treatment is costly and requires large plants. It relies on the use of micro-organisms that concentrate toxic soluble substances into sludge (that can be used as a fertilizer in agriculture). Today, a water decontamination plant costs about 1000 to 5000 euros per inhabitant. 30 to 40% of its running costs are devoted to the energy necessary for pool ventilation.

The waste-water treatment industry use software to optimize the plant design (number, size, interconnections of tanks), but design and improvements of bio-processes remain costly. This is why modelling allows numerical simulations of *virtual* bio-processes that can save substantial amount of money, avoiding tests at a real scale.

There is presently a growing need to conceive treatment systems in a more global framework, including the valorization of the “outputs” such as:

- the biogas production,
- the reuse of treated water for agriculture or dam refill in case of drought.

This requires to re-think the use of the models or to couple them with other models with new outputs and novel criteria to be optimized.

This is our most important domain of transfer and dissemination.

4.2. Environmental microbiology

Chemostat-like models (see Section 3.1.1) are also quite popular in theoretical marine ecology or in soil bio-chemistry, because micro-organisms play again a crucial role in the bio-geo-chemical cycles on Earth. Questioning are here a bit different than the ones depicted in Section 4.1 , because it is much more oriented towards comprehension and prediction than decision making (at the present time). Grasping the role of the microbial biodiversity appears to be an everlasting and common important question among scientists of various domains.

Nevertheless, mathematical models are quite similar but with some specificity (much more resources are available in marine microbiology; the spatial heterogeneity play a crucial role in underground processes).

A recent trend of considering natural microbial ecosystems on Earth to be able to delivering new 'eco-systemic services' has emerged, especially in terms of bio-remediation. Modelling and simulating tools are much relevant as in site experiments are quite costly and time-consuming.

4.3. Bioprocesses industry

Several industries use micro-organisms or yeasts to product substances of commercial interest (in pharmaceuticals, green biotechnology, food making...). Novel investigation techniques in microbiology (such as multistage continuous bioreactors) brings new insights on the metabolic functioning of the various strains. This conducts to revisit old models such as Monod's one, and to look for new estimation and piloting strategies. Those questions are quite closed from the ones studied in 4.1 and 4.2 , although the ecological dimension is less present (most of the culture are pure ones). The team is naturally solicited to contribute together with the specialists about problems related to modelling, simulation and control of these bio-processes.

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, Laurent Debreu, Jérémie Demange, Marc Honorat, Christine Kazantsev, Eugène Kazantsev, François-Xavier Le Dimet, Florian Lemarié, Bénédicte Lemieux-Dudon, Maëlle Nodet, Antoine Rousseau, Arthur Vidard.

Keywords: 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, Maëlle Nodet.

Keywords: 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 loosing 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, Mehdi-Pierre Daou.

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 tackled these issues, in the past in the framework of a partnership with the French electricity company EDF, and now thanks to another contract with ARTELIA Group.

MORPHEME Project-Team (section vide)

MYCENAE Project-Team

4. Application Domains

4.1. Introduction

MYCENAE addresses rather “upstream” questions in neuroendocrinology and neuroscience. Nevertheless, MYCENAE’s expected results can contribute to more applied issues in these fields, mainly by helping understand the mechanisms underlying physiological and pathological processes and also by designing new concepts for biomedical data analysis. MYCENAE thematics are related to societal issues concerning endocrine disruptors, reproductive biotechnologies, and neurological diseases, especially in case of pathological synchronizations encountered in epilepsy and Parkinson’s disease.

4.2. Neuroendocrinology and Neuroscience

We are interested in the complex dynamical processes arising within neuroendocrine axes, with a special focus on the reproductive (hypothalamo-pituitary-gonadal) axis. This axis can be considered as the paragon of neuroendocrine axes, since it both concentrates all remarkable dynamics that can be exhibited by these axes and owns its unique specificities, as gonads are the only organs that host germ cells. Since, in neuroendocrine axes, neural systems are embedded within endocrine feedback loops and interact with peripheral organs, one also needs to get interested in the peripheral dynamics to be able to “close the loop” and account for the effect of peripheral inputs on neural dynamics. In the case of the HPG axis, these dynamics are especially complex, because they involve developmental processes that occur even in adult organisms and combine the glandular function of the gonads with their gametogenic function.

Neuroendocrinology is thus a scientific field at the interface between Neuroscience, Endocrinology and Physiology (and even of Developmental Biology in the case of the HPG axis). On a neuroscience ground, mathematical neuroendocrinology is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood stream. Neuroendocrine networks are characterized by the emergence of very slow rhythms (on the order of an hour), finite size effects due to their relative small number of neurons (on the order of a few thousands for the Gonadotropin-Releasing-Hormone network) and neuroanatomical particularities, that impact the way they can synchronize and desynchronize. On a physiological ground, gonadal cell biology raises specific cell biology issues on more than one account. First, the gonads are the only organs sheltering the germ cell lines (corresponding to oogenesis in ovaries and spermatogenesis in testes). Hence, the two modes of cell division, mitosis and meiosis are encountered in these tissues. Second, there are intricate interactions between the gonadal somatic cells (granulosa cells in the ovaries, sertoli cells in the testes) and the germ cells. Third, the control of gonadal cell populations is exerted within endocrine feedback loops involving both the hypothalamus and pituitary, which results naturally in multiscale population dynamics coupled with hormonally-controlled cell kinetics.

MYCENAE’s research topics in mathematical neuroscience deal with complex oscillations, synchronization and plasticity. We study (i) the emergence of network-level behaviors from individual dynamics of excitable cells (mainly neurons, but not exclusively, as the pituitary cells belong to the family of excitable cells): complete synchronization or synchronization of specific events, effect of the recruitment rate in the synchronization process, dependence on the neuro-anatomical and functional coupling properties; (ii) the control of the different possible configurations of the network depending on external (e.g. daylength) and/or internal inputs (e.g. metabolic status), at the source of plasticity processes in cognitive (vision learning) or neuroendocrine systems (differential sensitivity to gonadal steroids and peptides across the different steps of the reproductive life); (iii) the encoding of neuro-hormonal signals as complex oscillations, on the electrical, ionic (calcium dynamics) and secretory levels; and (iv) the decoding of these signals by their target neuronal or non-neuronal cells.

The biological relevance of our modeling and model-based signal analysis approaches is grounded on our network of collaborations with teams of experimentalist biologists. In particular, we have long standing collaborations with the UMR 6175 (INRA-CNRS-Université François Rabelais-Haras Nationaux) “Physiologie de la Reproduction et des Comportements” that covers most our research topics in reproductive neuroendocrinology. We have especially closed links with the Bingo (Integrative Biology of the ovary) and Bios (Biology and Bioinformatics of Signaling Systems) teams, which were partners of the REGATE LSIA. We have been jointly investigating issues relative to terminal or basal follicular development [5], [6], analysis of neurosecretory patterns [13] and modeling of GPCR (G-Protein Coupled Receptors) signaling networks [8]. We also have special links with the Center for Interdisciplinary Research in Biology (CIRB, Collège de France), headed by Alain Prochiantz, that help us get a better understanding of how the brain connectivity develops and how it is functionally organized. An instance of a recent collaborative work is the study of the organization of spatial frequencies in the primary visual cortex [42].

NEUROMATHCOMP Project-Team (section vide)

NEUROSYS Team

4. Application Domains

4.1. General remarks

The research directions of the team are motivated by general anaesthesia (GA) that has attracted our attention in the last years. The following paragraphs explain in some detail the motivation of our work on the four major phenomena of GA: loss of consciousness, immobility, amnesia and analgesia.

During general anaesthesia, the electroencephalogram (EEG) on the scalp changes characteristically: increasing the anaesthetic drug concentration the amplitudes of oscillations in the α -band ($\sim 8 - 12$ Hz) and in the δ -band ($2 - 8$ Hz) increase amplitudes in frontal electrodes at low drug concentrations whereas the spectral power decreases in the γ -band ($\sim 20 - 60$ Hz). This characteristic change in the power is the basis of today's EEG-monitors that assist the anaesthetist in the control of the anaesthesia depths of patients during surgery. However, the conventional monitors exhibit a large variability between the patients detected anaesthetic depth and their real depth. Moreover, a certain number of patients re-gain consciousness during surgery (about 1 - 2 out of 1000) and a large percentage of patients suffer from diverse after-effects, such as nausea or long-lasting cognitive impairments such as partial amnesia (from days to weeks). Since surgery under general anaesthesia is part of a hospital's everyday practice, a large number of patients suffer from these events everyday. One reason for the lacking control of such disadvantageous effects is the dramatic lack of knowledge on what is going on in the brain during general anaesthesia and a weak EEG-online monitoring system during anaesthesia. Consequently, to improve the situation of patients during and after surgery and to develop improved anaesthetic procedures or even drugs, research is necessary to learn more about the neural processes in the brain and develop new monitoring machines.

4.2. Level of consciousness

The EEG originates from coherent neural activity of populations in the cortex. Hence to understand better the characteristic power changes in EEG during anaesthesia, it is necessary to study neural population dynamics subject to the concentration of anaesthetic drugs and their action on receptors on the single neuron level. We study mathematical models which will be constrained by the signal features extracted from experimental data, such as EEG (data provided by Jamie Sleight, University of Auckland and Christoph Destrieux, University of Tours), Local Field Potentials (data provided by Flavio Frohlich, University of North Carolina - Chapel Hill) and behavior. The combination of model and analysis of experimental data provides the optimal framework to reveal new knowledge on the neural origin of behavioral features, such as the loss of consciousness or the un-controlled gain of consciousness during surgery. For instance, modelling studies show that the characteristic changes of spectral power (second-order statistics) are not sufficient to deduce all underlying neural mechanisms. Consequently, additional higher-order statistical measures may provide additional insight into underlying neural mechanisms and may provide a novel marker for the loss of consciousness.

Moreover, the constant supervision of anaesthetized patients in intensive care is a demanding task for the personnel in hospital practice. It is almost not possible to take care of a patient constantly and hence the today's medicine demands monitoring devices that control automatically the level of anaesthetic drugs based on the patients' neural activity (e.g., EEG). Brain-Computer-Interfaces (BCI) have already demonstrated their potential for the detection of consciousness in non-responsive patients. We will apply the data analysis techniques known in BCI to extract new markers for the depth of anaesthesia. More specifically, for deeper anaesthesia, auditory-evoked and Event-Related Desynchronization/Event-Related Synchronization (ERD/ERS) BCI could be used to better identify the state of consciousness in patients under anaesthesia. In this context, we have established a first contact to the University of Wuerzburg. Another research direction will link intracranial EEG and scalp EEG by characterising micro-awake episodes during sleep.

4.3. Immobility

A research direction will be to take benefit of the relationship between the motor activity and anesthesia. Indeed, even if no movement is visually perceptible, a study by electroencephalographic recordings of brain activity in motor areas, quantifying the characteristics of amplitude and phase synchronization observed in the alpha and beta frequency bands, may reveal an intention movement. This feature is important because it demonstrates that the patient is aware. Thus, we will develop an experimental protocol in collaboration with an anesthesiologist of the regional hospital on stimulating the median nerve at forearm level to track the evolution of the shape of the beta rebound in the motor cortex for various doses of the anesthetic agent.

4.4. Amnesia

Patients sometimes develop post-traumatic disorders associated with the surgery they underwent because they either woke up during the surgery or because the amnesiant effect of the general anaesthesia was only partial, declarative memory being maintained in some unexplained cases. It is still unknown how memory can be maintained under general anaesthesia and it needs to be investigated to improve the recovery from anaesthesia and to avoid as much as possible post-traumatic disorders. To learn more about memory under anaesthesia, we will focus our theoretical studies on the oscillation regimes observed in the hippocampus, mainly in the theta and gamma ranges, which are correlated with memory formation and retrieval.

4.5. Analgesia

One of the most important aspect in general anaesthesia is the loss of pain. During surgery, it is very difficult to find out whether the anesthetized patient feels pain and hence will develop cognitive impairment after surgery. Today, the anesthesiologist knows and detects physiological signs of pain, such as sweat, colour of skin or spontaneous involuntary movements. However, more objective criteria based on EEG may assist the pain detection and hence improves the patients' situation. To this end, we analyze large sets of patient EEG-data observed during surgery and aim to extract EEG signal features of pain.

NUMED Project-Team (section vide)

PARIETAL Project-Team

4. Application Domains

4.1. Human neuroimaging data and their use

Human neuroimaging consists in acquiring non-invasively image data from normal and diseased human populations. Magnetic Resonance Imaging (MRI) can be used to acquire information on brain structure and function at high spatial resolution.

- T1-weighted MRI is used to obtain a segmentation of the brain into different different tissues, such as gray matter, white matter, deep nuclei, cerebro-spinal fluid, at the millimeter or sub-millimeter resolution. This can then be used to derive geometric and anatomical information on the brain, e.g. cortical thickness.
- Diffusion-weighted MRI measures the local diffusion of water molecules in the brain at the resolution of 1 to 2mm, in a set of directions (60 typically). Local anisotropy, observed in white matter, yields a local model of fiber orientation that can be integrated into a geometric model of fiber tracts along which water diffusion occurs, and thus provides information on the connectivity structure of the brain.
- Functional MRI measures the blood-oxygen-level-dependent (BOLD) contrast that reflects neural activity in the brain, at a spatial resolution of 1.5 to 3mm, and a temporal resolution of about 2s. This yields a spatially resolved image of brain functional networks that can be modulated either by specific cognitive tasks or exhibit spontaneous co-activations.
- Electro- and Magneto-encephalography (MEEG) are two additional modalities that complement functional MRI, as they directly measure the electric and magnetic signals elicited by neural activity, at the millisecond scale. These modalities rely on surface measurements and do not localize brain activity very accurately in the spatial domain.

4.2. High-field MRI

High field MRI as performed at NeuroSpin (7T on humans, 11.7T in 2017, 17.6T on rats) brings an improvement over traditional MRI acquisitions at 1.5T or 3T, related to a higher signal-to-noise ratio in the data. Depending on the data and applicative context, this gain in SNR can be traded against spatial resolution improvements, thus helping in getting more detailed views of brain structure and function. This comes at the risk of higher susceptibility distortions of the MRI scans and signal inhomogeneities, that need to be corrected for. Improvements at the acquisition level may come from the use of new coils (such as the 32 channels coil on the 7T at NeuroSpin), as well as the use of multi-band sequences [44].

4.3. Technical challenges for the analysis of neuroimaging data

The first limitation of Neuroimaging-based brain analysis is the limited Signal-to-Noise Ratio of the data. A particularly striking case is functional MRI, where only a fraction of the data is actually understood, and from which it is impossible to observe by eye the effect of neural activation on the raw data. Moreover, far from traditional i.i.d. Gaussian models, the noise in MRI typically exhibits local and long-distance correlations (e.g. motion-related signal) and has potentially large amplitude, which can make it hard to distinguish from true signal on a purely statistical basis. A related difficulty is the *lack of salient structure* in the data: it is hard to infer meaningful patterns (either through segmentation or factorization procedures) based on the data only. A typical case is the inference of brain networks from resting-state functional connectivity data.

Regarding statistical methodology, neuroimaging problems also suffer from the relative paucity of the data, i.e. the relatively small number of images available to learn brain features or models, e.g. with respect to the size of the images or the number of potential structures of interest. This leads to several kinds of difficulties, known either as *multiple comparison problems* or *curse of dimensionality*. One possibility to overcome this challenge is to increase the amount of data by using images from multiple acquisition centers, at the risk of introducing scanner-related variability, thus challenging the homogeneity of the data. This becomes an important concern with the advent of cross-modal neuroimaging-genetics studies.

POMDAPI Project-Team

3. Application Domains

3.1. Environmental sciences

Applications are in hydrogeology and water resources.

3.2. Energy sciences

Applications are in oil reservoir and sedimentary basin simulations, and in optimization of the power flow in an electricity transportation network.

POPIX Team

4. Application Domains

4.1. Pharmacometrics

Participants: Marc Lavielle, Kevin Bleakley, Célia Barthélémy.

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
- protocol optimization tools
- diagnostic tools
- model selection tools
- data exploration tools
- estimation techniques for complex models (eg, stochastic differential equations, partial differential equations)

4.2. Gene expression

Participant: Marc Lavielle.

Mixed effects models can also be successfully used in quantitative biology for modeling the dynamics of biological networks in cell populations. Indeed, the population approach is relevant for building predictive computational models of intracellular processes. POPIX was interested with the experiments performed by the CONTRAINTES Inria team looking at the high-osmolarity glycerol (HOG) pathway in budding yeast. Yeast cells are exposed to osmotic shocks, i.e., sudden changes in the solute concentration of their surroundings. Signal transduction pathways, most notably the HOG pathway, provide information to the cell about the osmolarity of its environment and activate responses to deal with these stress conditions. In particular, a large set of genes is turned on and corresponding stress-responsive proteins are produced. This protein production process can be quantified by replacing one target protein, for example STL1, by a fluorescent protein such as yECitrine. This can be done by genetically modifying the yeast genome.

Thanks to time-lapse microscopy and cell tracking algorithms, single cell responses can be measured over time. Significant inter-cell variability is often observed.

The related Hog1-induced gene expression model is given by a parametric reaction network. MONOLIX can then be used to estimate the model parameters.

A collaboration with LIFEWARE (formerly CONTRAINTES) is starting on this subject.

4.3. Oncology

Participants: Marc Lavielle, Célia Barthélémy.

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 project Lab 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 ongoing, with the aim of extending the statistical methods developed by POPIX to partial differential equation-based models. NUMED works on models of tumor growth and has previously implemented an extension of MONOLIX to KPP-type reaction-diffusion models.

4.4. Respiratory system

Participants: Bertrand Maury, Astrid Decoene.

Comprehensive models to simulate the whole pulmonary system, i.e., the mechanical behavior of the lung and gas exchanges within the pulmonary system, are built upon ODE and PDE approaches. For instance, the mechanical behavior of a lung is often described by single or multi-compartment ODE models, whereas air flow may be determined by the coupling of a 3D PDE system in the proximal part of the bronchial tree with a 0D ODE system in the distal part of the bronchial tree. Gas exchange has so far been investigated using 0D or 1D models in which heterogeneity of gas exchange along the path length may be investigated.

In a mathematical representation of such physiological systems, model parameters can be associated with specific quantities in the real system, such as the resistance and compliance of the pulmonary system. These quantities are time-dependent and nonlinear and are measured by pneumologists in order to characterize chronic obstructive pulmonary diseases (COPD) such as asthma and emphysema. These parameters may be useful in assessing lung conditions.

Although most physiological studies have used averaged deterministic models of the tracheobronchial tree geometry, morphometric studies show that inter-subject and intra-subject variability in the structural components of the human lung is significant. In particular, the resistance of the respiratory tract may be significantly affected as it is directly related to the inner diameter of the bronchi. Feedback from such variability to resistance and, as a consequence efficiency of the gas exchange process, within the framework of a fully coupled model, is unclear. In this situation, the statistical and numerical approaches being developed by POPIX are clearly promising estimation methods for respiratory system analysis.

4.5. Blood flow modeling

Participants: Bertrand Maury, Astrid Decoene.

Modeling and numerical simulation of blood flow in arteries and veins may become an important tool for medical applications, as for instance in the prediction of cardiovascular disease. Analyzing the pressure waves and estimating the wall compliance of arteries is fundamental, as these exhibit strong inter- and intra-subject variability. Currently, non-invasive pressure measurements involve excessive errors; intensive direct estimation is thus not applicable in practice. Physiologists therefore hope to be able to predict the time and space evolution of the pressure in the arterial network from a small amount of flow data measured at a few points.

Several numerical models have been developed in order to simulate blood flow in arteries and veins. They mainly consist of one to three-dimensional systems of partial differential equations, depending on the level of complexity one desires to achieve. Coupling the various models is also an issue. These numerical models allow us to compute the transversal section area, as well as the velocity or flow at different points in space, leading to a rather complete description of the arterial flow (velocity, pressure, section). But for these models to be adapted to each patient, certain numerical and physical parameters must be fitted, such as the compliance of walls and the viscosity of the blood. These parameters are difficult to estimate experimentally and may be related to measurements which involve a non-negligible error. Furthermore, their optimal value is linked to the particular modeling framework and therefore can differ from the value given by their physical definition.

Mixed models appear to be an appropriate framework for taking into account the specific nature of each patient and quantifying uncertainty in the numerical model. Flow data are available as it is possible to non-invasively measure the mean velocity in and diameter of an artery.

We aim to introduce statistical mixed models to the framework for the classical one-dimensional blood flow model.

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.

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.* [62], [66], [65] 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). One of the goal is to use our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macs/M3disym project-team, we are interested in the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology, in particular the simulation of experimental measurement of the field potential of cardiac stem cells in multi-electrode arrays.

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 at OSUR, Pprime at University of Poitiers and CDCSP at University of Lyon. It is also done in the context of the group Momas and previous Andra grants.

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 Project-Team

4. Application Domains

4.1. Biological pilot models: Birbeck granule and Melanosome biogenesis

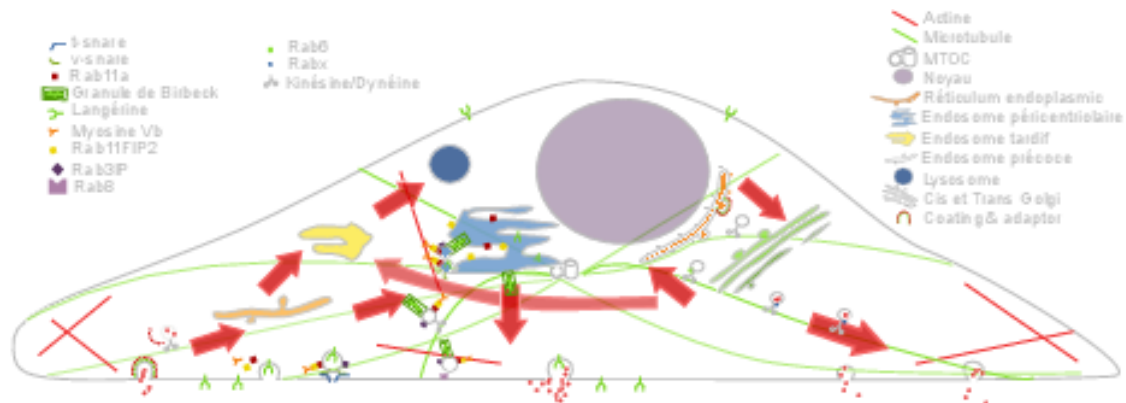


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried at UMR 144 CNRS-Institut Curie (“Space Time imaging of Endomembranes and organelles Dynamics”) (STED) team) 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: i) multiple sorting and structural events involved in the biogenesis of these organelles; ii) complexity of the endo-melanosomal network of these highly specialized cells; iii) complex molecular architecture organizing and coordinating their dynamics; iv) 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 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 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.

SHACRA Project-Team (section vide)

SISTM Team

4. Application Domains

4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [47], "Systems vaccinology" [44], "Systems medicine" [35]. From the statistician point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

4.2. The case of HIV immunology

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [39]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [34]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

SISYPHE Project-Team

3. Application Domains

3.1. Medical applications of modeling, signal analysis and control

- *3D cardiac modeling for personalized medicine.* Our main contribution to Inria's collective efforts in this field (project-teams Asclepios, MACS, REO, Sisyphe) is the so-called "Bestel-Clément-Sorine" model of contraction of cardiac muscle [39], at the origin of the 3D electromechanical direct and inverse modeling of the heart at Inria. This model is based on ideas originating from the kinetic equation theory, used to model, on the molecular scale, the controlled collective behavior of actin-myosin nanomotors at the root of muscle contraction. The classical Huxley model was recovered on the sarcomere scale by using moment equations and a controlled constitutive law on the tissue scale was obtained using the same type of scaling techniques. The model, now embedded in heart simulators, is used in various studies [51], [3], [54], [53].

- *Semiclassical analysis of cardiovascular signals.* This work began with the article [41] and the PhD of M. Laleg-Kirati [47], [46], [49]. The theory and a validation of a new method of blood pressure analysis are now published [17], [48].

The main idea is to consider a signal $x \rightarrow y(x)$ to analyze as the multiplication operator $\phi \rightarrow y\phi$ on some function space, and to analyze it as a potential. The signal is represented by the spectrum of an associated Schrödinger operator, combined with a semi-classical quantification: $-h^2 \frac{d^2}{dx^2} - y(x)$ with $h > 0$ small. For signals looking as "superpositions of bumps" (e.g. the systolic pulse, the dichrotic notch for the arterial pulse pressure), this leads to some kind of nonlinear Fourier analysis [17]. The spectral parameters associated with the arterial pressure can be useful cardiovascular indices, e.g. for noninvasive blood flow estimation [48]. In the arterial pressure case, this is equivalent to approximate the traveling pressure pulse by a N -soliton solution of a Korteweg-de Vries (KdV) equation [41] and using ideas similar to the Lax pair representation of N -solitons and proof technique for the weak dispersion limit of KdV. A striking result is that an N -soliton is a very good representation of the arterial pressure waveform for values of N as small as $N = 3$. The representation of pulse-shaped signals is parsimonious, having only $2N$ parameters [55].

- *Multiscale signal analysis of cardiovascular signals:* collaboration with Julien Barral (former member of Sisyphe) and partners of the ANR project DMASC. The starting point was the common idea that "A Healthy Heart Is a Fractal Heart". We have developed a method to test the existence of scale laws in signals and applied it to RR signals: the heart rate is not always fractal or even multifractal in an Healthy Heart [19].

- *Modeling and control of CARMAT Total Artificial Heart.* This TAH has been implanted for the first time in a patient in Dec 2013. We have contributed to this industrial project since 2008 on modeling and control questions during the post-doc of Karima Djabella (now engineer at CARMAT), Frédéric Vallais and the two-year contract for supervising Julien Bernard (CARMAT control engineer). It was an opportunity to exploit some results on the baroreflex control [42] or heart rate variability during exercise [40].

- *Glycemic control in Intensive Care Units (ICUs):* Blood glucose is a key biological parameter in ICU since the study of van den Berghe et al [60] who demonstrated decreased mortality in surgical intensive care patients in association with tight glycemic control (TGC), based on intensive insulin therapy. In their work, however, there was only one ICU and the protocol was not formalized. Trying to decrease mortality in standard ICUs by using computer aided glycemic control is still a challenge. Previous studies have failed because of high rates of severe hypoglycemia. The last one was NICE-SUGAR [57] with a 2% increase in mortality (death ratio from any cause within 90 days after randomization compared between control and TGC patients). In cooperation with Pierre Kalfon (Intensive Care, Hospital of Chartres) and in the framework of a CIFRE contract with a small medtech company LK2 (Tours, France), we have studied the origins of these failures and proposed more robust control algorithms tuned using a database of representative "virtual patients". See [44] and the PhD of A. Guerrini, [43]. A first version of the controller has been tested in a large clinical study CGAO-REA [14].

3.2. Engineering applications of modeling, signal analysis and control

Identification of nonlinear systems: from algorithms to a popular matlab toolbox:

Block-oriented nonlinear system identification with Jiandong Wang (Associate Professor, Beijing University, China) [58]; Development of the Matlab System Identification ToolBox (SITB).

Identification of transmission line characteristics: from algorithms to electronic experiments. Collaboration with CEA LIST (Lab of applied research on software-intensive technologies) and LGEP (Laboratoire de génie électrique de Paris) with Florent Loete [50] (ANR projects SEEDS, 0-DEFECT, INSCAN, SODDA).

We have extended to some networks the seminal work of Jaulent [45] for the real line: all the information contained in a measured reflection coefficient can be obtained by solving an inverse scattering problem for a system of Schrödinger or Zakharov-Shabat equations on the graph of the network, which allows one to recover the geometry of the network and some electrical characteristics for nonuniform lossless electrical star-shaped networks [26]. An efficient method to solve the associated Guelfand-Levitan-Marchenko equations has been studied and is used in the software ISTL that has been developed in Sisyphé [59], [56]. This development will continue in the project-team I4S. An engineering methodology based on this approach has been described [29] and some first experimental results obtained [36], [50].

Monitoring and control of automotive depollution systems: with RENAULT (Karim Bencherif, Damiano Di Penta and PhD students): [52], [20], [38].

3.3. Modeling and optimizing patient pathways in hospital

This research theme was inspired by the observation that the practice of modern industrialized medical care proceeds by queueing and forwarding patients from one step of care to the next, with each step involving specialized personnel and machinery. Whereas the human competence and machine performance available in hospital are both highly evolved, the deployment of these resources for the patient's benefit is problematic.

If we change our point of view and, rather than look in the traditional way at the hospital as a collection of "vertical" silos (emergency room, cardiac ward, respiratory ward, operating rooms, imagery services, blood services, logistics, etc.), we instead consider the patient's "horizontal" trajectory crossing through many different silos, we can see significant problems. In between each step, whose value added part typically lasts at most a few minutes, there are long stretches of time (on the order of hours) during which the patient simply waits for the hospital to arrange the next step of care — with serious consequences:

- delayed treatment is directly correlated with increased rates of hospital acquired nosocomial infection for the immuno-compromised and to loss of autonomy for the elderly;
- the cost in terms of scarce resources (for example, bed-hours in hospital) is significant, but the root causes of these costs due to problems at the interfaces between silos are hard to observe since they do not fit neatly into the hospital's traditional hierarchy;
- the slow and error prone hand-offs of information between silos are dangerous for the patient who is vulnerable to medical errors. In the US, for example, it is estimated that there are 100,000 deaths per year attributable to hospital error (triple the number due to road traffic accidents by way of comparison).

STEER Team

4. Application Domains

4.1. Introduction

In the context described in the previous sections, we can distinguish two connected and complementary strategies for analyzing environmental pressures: a sectorial approach and a spatial one. The first one is more directly connected to ecological accounting, the second one has more direct relations to urban economy and land cover modelling. Let us start by describing the former.

4.2. Ecological accounting for sectorial pressure assessment

One of the major issues in the assessment of the long-term sustainability of urban areas is related to the concept of “imported sustainability”. Cities bring in from the outside most of their material and energy resources, and reject to the outside the waste produced by their 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. Such an assessment brings to light the most relevant direct and indirect lines of action on these issues. In this respect, it is useful to introduce the alternative concepts of consumer versus producer responsibility (or point of view).

The producer point of view is the most useful to pinpoint relevant direct lines of actions on environmental pressures due to production. In other respects, any territory imports and exports goods and services from and to the rest of the world. The consumer point of view provides information on the indirect pressures associated with these exchanges, as production responds to a final demand. Tracking the various supply chains through the analysis of the structure of the local economy and its relations and dependencies to the external world allows us to identify critically important contributions to environmental pressures; this also enables us to define fair environmental indicators in order not to attribute environmental pressures to producers only (whose responsibility is the easier to quantify of the two). In this approach, the producer responsibility follows directly from the measurement of its energy and material uses, while the consumer responsibility is established indirectly through an allocation of the impacts of production to the final consumers, but this second mode of allocation is to some extent virtual and partly subjective.

Four methods stand out:

- Material Flow Analysis (MFA)
- Input-Output Analysis (IOA)
- Life-Cycle Analysis (LCA)
- Ecological Footprint (EF)

Each of these is based on a well-defined structuring element: mass conservation for MFA, measure of industrial inter-dependencies for IOA, identification of all the steps from cradle to grave for LCA, measure of biocapacity demand for EF. The different methods have preferred areas of application. For example, EF is more relevant for analyzing primary production such as agricultural staples, wood, etc. IOA is more focused on whole industrial sectors, while LCA is geared towards end-user products, taken as functional units; finally, primary materials (such as metals), waste and emissions are more easily characterized through MFA. Methodological choices are driven by the type of question one needs to address, data availability and collection method and the spatial scales under consideration. Indeed, data can be used in two different ways: bottom-up or top-down. The bottom-up data is more precise, but in general precludes comprehensiveness; on the contrary, the top-down data is by nature more comprehensive, but is not suited for a detailed, fine-scale analysis of the results.

STEPP has already initiated its research program on this theme with three major goals: 1) Creating a comprehensive database enabling pressure analyses; 2) Developing methodologies and models resolving scaling issues, and developing algorithms allowing us to rigorously and automatically obtain adequate assessments; 3) Providing a synthetic analysis of environmental pressures associated to the major material flows, at various geographic levels (employment catchment area, *département* and *région*, for France), with the explicit aim of incorporating this type of information in the public decision process on environmental issues, via specifically designed decision-help procedures.

4.3. Urban economy and land use/land cover changes: assessment of spatial distributions of the pressures

The preceding section was focused on territorial metabolism, in particular on the analysis of supply chains. Here territories are examined with a more prominent emphasis on their spatial dimension, with attention to: the spatial distribution of local pressures previously identified (from a land use point of view), and the modeling of future land use and activity location (from an economic point of view). These two questions correspond to very different modeling strategies: the first one is more statistical in nature, extrapolating future land use from past evolution combined with global territory scenarios; the other one has a more fundamental flavor and focuses on an understanding of the processes driving urbanization. For this, we focus more precisely on the question of household and businesses choices of localization, as well as on spatial fluxes within the territory (transportation of goods and persons). The critical point here is to understand and manage urban sprawl and its environmental effects (GHG emission, loss of arable land, ecosystem fragmentation, and so on).

Land Use/land Cover Change models (LUCC)

LUCC models are mostly used in environmental sciences, e.g. to evaluate the impact of climate change on agriculture, but they can also be used to analyze urban sprawl. There is a variety of models, static or dynamic, grid- or agent- based, local or global, etc., and with varying degrees of sophistication concerning spatio-temporal analysis or decision structures incorporated in the model.

The models of interest here are statistical in nature but spatially explicit. Following decades of development, they are robust, versatile and mature. In principle, agent-models have a larger potential for representing decision processes, but in practice this advantage results in a loss of universality of the models. Among the most well-known and most mature models, one can mention the CLUE family of models, DINAMIC, or LCM (Land Change Modeler). These models are well described in the literature, and will only be briefly presented here.

These models analyze change in land use in a statistical way; they are structured around three different modules:

- The first module determines the probability of change of pixels of the territory (pixels are typically tens to hundreds of meters in size).
- The second module defines the global changes between the various land uses of interest per time step (usually, a few years), based on global scenarios of evolution of the territory under study. These first two modules are independent of one another.
- The last module distributes changes of land use in an explicit manner, pixel per pixel, at each time step, on the basis of the information provided by the first two modules.

Probabilities of change are calibrated on past evolution, from the differences between two past maps of land use in the more favorable cases, or from a single map otherwise (under the assumption that the logic of occupation changes is the same as the logic of land use at this single date). Such changes are then characterized in a statistical way with the help of modeling variables identified by the modeler as having potential explaining or structuring power (typically, a few to a dozen variables are used for one type of land use change). For example, in the case of urban sprawl, typical explaining factors are the distance to existing urbanized zones or distances to roads and other means of transportation, elements of real estate costs, etc. Global scenarios are quantified in terms of global changes in land use over the whole studied area (e.g., how many hectares are

transformed from agricultural to urban uses in a given number of years, how does this evolve over time...); this is done either from academic expert knowledge, or from information provided by local planning agencies. Whenever feasible, models are validated by comparing the model predictions with actual evolution at a later date. Therefore, such models need from one to three land use maps at different dates for calibration and validation purposes (the larger the number of maps, the more robust and accurate the model). A large array of statistical tools is available in the literature to perform the calibration and validation of the model.

The horizon of projections of such models is limited in time, typically 20-30 years, due to the inherent uncertainty in such models, although they are occasionally used on longer time-scales. Climate change constraints are included, when needed, through scenarios, as it is not in the scope of such models to incorporate ecological processes that may translate climate change constraints into land cover change dynamics. Note that on such short time-scales, climate change is not dominated by the mean climate evolution but by decade variations which average out on longer time-scales and are not modeled in the global climate models used e.g. for IPCC projections for the end of the century; as a consequence, the various IPCC climate scenarios cannot be distinguished on such a short time horizon.

With regard to LUCC, the STEEP team is involved in the ESNET project, bearing on the characterization of local Ecosystem Services NETWORKS; the project is coordinated by LECA (*Laboratoire d'Ecologie Alpine*), in collaboration with a number of other research laboratories (most notably, IRSTEA Grenoble, besides our team), and in close interaction with 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, at the 2030/2040 horizon, under constraints of climate change. 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 project modelling effort, with the help of an ESNET funded post-doctoral researcher. In the process, areas of conceptual and methodological improvements of statistical LUCC models have been identified; implementing these improvements may be useful for the LUCC community at large, independently of the ESNET project needs.

Models for Land-Use and Transportation Interactions (LUTI)

Urban transport systems are intricately linked to urban structure and activities, i.e., to land use. Urbanization generally implies an increased travel demand. Cities have traditionally met this additional demand by extending 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.

Transportation models allow us to address questions generally limited to the impacts of new infrastructures, tolls and other legislation on traffic regulation⁰, on user behavior⁰, or on the environment⁰. LUTI models (Land-Use and Transport Integrated models) can answer a much broader spectrum of issues. For example, they allow us to understand how the localization of households and of economic activities (which generate transportation demand) adapt to changes of transportation supply. They also allow us to assess the impacts of such changes on the increase in real estate value, or more generally on their effects on the economic development of a specific sector or neighborhood. An economic vision interprets all these interactions in terms of equilibrium between demand and supply. Modelling the localization of households and employments (companies) relies on capturing the way stakeholders arbitrate between accessibility, real estate prices, and attractiveness of different areas.

State of the art and operability of LUTI models.

The first model that proved able to analyze the interactions between transport and urbanization was developed by Lowry. Since then theories and models have become increasingly complex over time. They can be classified according to different criteria. A first classification retraces the historic path of these theories and models. They

⁰ Congestion, cost and time spent for the transport, etc.

⁰ Changes in modality choice

⁰ CO2 emissions, air pollution, noise nuisance, etc.

can be associated with one or several of the approaches underlying all present theories: economic base theory and gravity models, Input/Output models and theory of urban rent, and micro-simulations. A second possibility consists in classifying the models according to their aims and means.

Significant scientific progress has been made over the last thirty years. Nevertheless, modelling tools remain largely restricted to the academic world. Today, only seven models have at least had one recent application outside academia or are commercialized or potentially marketable, in spite of the important needs expressed by the urban planning agencies: Cube Land, DELTA, MARS, OPUS/UrbanSim, PECAS, TRANUS and Pirandello.

To guide their choice of a modelling framework, users can rely on various criteria such as the strength of the theoretical framework, the quality and the diversity of the available documentation, the accessibility of the models (is the model freely available? is the code open source? is the software regularly updated and compatible with the recent operating systems?), the functionality and friendliness of user interfaces (existence of graphic user interface, possibility of interfacing with Geographic Information Systems), existence of technical assistance, volume and availability of the data required to implement the model, etc. For example, among the seven models mentioned above, only two are open source and mature enough to meet professional standards: TRANUS and UrbanSim⁰. These two models are very different but particularly representative of the main current philosophies and trends in this scientific domain. Their comparison is informative.

STEPP implication in LUTI modelling.

As yet, very few local planning authorities make use of these strategic models, mostly because they are difficult to calibrate and validate. Systematic improvement on these two critical steps would clearly increase the level of confidence in their results; these limitations hinder their dissemination in local agencies. One of the major goals of STEEP is therefore to meet the need for better calibration and validation strategies and algorithms. This research agenda lies at the core of our projects CITiES (“ANR *Modèles Numériques*”) and TRACER (Ecos Nord Venezuela). As for LUTI modeling, we have been using the TRANUS model since the creation of our team. We have also been working on UrbanSim from the beginning of the CITiES project. In this framework we work in close collaboration with AURG⁰, the local urban planning agency of Grenoble (*Agence d’Urbanisme de la Région Grenobloise*) in order to better understand and to improve the relevance of these tools for such territorial agencies.

⁰<http://www.urbansim.org>

⁰<http://www.aurg.org/>

TONUS Team

4. Application Domains

4.1. Controlled fusion and ITER

The search for alternative energy sources is a major issue for the future. Among others, controlled thermonuclear fusion in a hot hydrogen plasma is a promising possibility. The principle is to confine the plasma in a toroidal chamber, called a tokamak, and to attain the necessary temperatures to sustain nuclear fusion reactions. The International Thermonuclear Experimental Reactor (ITER) is a tokamak being constructed in Cadarache, France. This was the result of a joint decision by an international consortium made of the European Union, Canada, USA, Japan, Russia, South Korea, India and China. ITER is a huge project. As of today, the budget is estimated at 20 billion euros. The first plasma shot is planned for 2020 and the first deuterium-tritium operation for 2027.

Many technical and conceptual difficulties have to be overcome before the actual exploitation of fusion energy. Consequently, much research has been carried out around magnetically confined fusion. Among these studies, it is important to carry out computer simulations of the burning plasma. Thus, mathematicians and computer scientists are also needed in the design of ITER. The reliability and the precision of numerical simulations allow a better understanding of the physical phenomena and thus would lead to better designs. TONUS's main involvement is in such research.

The required temperatures to attain fusion are very high, of the order of a hundred million degrees. Thus it is imperative to prevent the plasma from touching the tokamak inner walls. This confinement is obtained thanks to intense magnetic fields. The magnetic field is created by poloidal coils, which generate the toroidal component of the field. The toroidal plasma current also induces a poloidal component of the magnetic field that twists the magnetic field lines (see Figure 2). The twisting is very important for the stability of the plasma. The idea goes back to research by Tamm and Sakharov, two Russian physicists, in the 50's.

Other devices are essential for the proper operation of the tokamak: divertor for collecting the escaping particles, microwave heating for reaching higher temperatures, fuel injector for sustaining the fusion reactions, toroidal coils for controlling instabilities, *etc.*

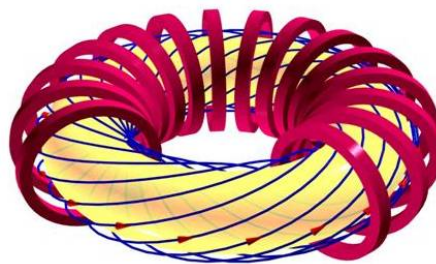


Figure 2. Poloidal coils and magnetic field lines geometry inside a tokamak

4.2. Other applications

The software and numerical methods that we develop can also be applied to other fields of physics or of engineering.

- For instance, we have a collaboration with the company AxesSim in Strasbourg for the development of efficient Discontinuous Galerkin (DG) solvers on hybrid computers. The applications is electromagnetic simulations for the conception of antenna, electronic devices or aircraft electromagnetic compatibility.
- The acoustic conception of large rooms requires huge numerical simulations. It is not always possible to solve the full wave equation and many reduced acoustic models have been developed. A popular model consists in considering "acoustic" particles moving at the speed of sound. The resulting Partial Differential Equation (PDE) is very similar to the Vlasov equation. The same modeling is used in radiation theory. We have started to work on the reduction of the acoustic particles model and realized that our reduction approach perfectly applies to this situation. We plan to supervise a new PhD with CEREMA (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) in Strasbourg. The objective is to investigate the model reduction and to implement the resulting acoustic model in our DG solver.

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, modeling 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. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then 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 labeling 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. Modeling 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, 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.

ALGORILLE Project-Team

4. Application Domains

4.1. Promoting parallelism in applications

In addition to direct contributions within our own scientific domain, numerous collaborations have permitted us to test our algorithmic ideas in connection with academics of different application domains and through our association with SUPELEC with some industrial partners: physics, geology, biology, medicine, machine learning or finance.

4.2. Experimental methodologies for the evaluation of distributed systems

Our experimental research axis has a *meta* positioning, targeting all large-scale distributed systems. This versatility allows us to factorize the efforts and maximize our efficiency. The resulting findings are typically used by researchers and developers of systems in the following domains:

- High Performance Computing systems (in particular MPI applications on high-end platforms)
- Cloud environments (in particular virtualized environments)
- Grids (in particular high throughput computing systems)
- Peer-to-peer systems

ALPINES Project-Team

4. Application Domains

4.1. Compositional multiphase Darcy flow in heterogeneous porous media

We study the simulation of compositional multiphase flow in porous media with different types of applications, and we focus in particular on reservoir/bassin modeling, and geological CO₂ underground storage. All these simulations are linearized using Newton approach, and at each time step and each Newton step, a linear system needs to be solved, which is the most expensive part of the simulation. This application leads to some of the difficult problems to be solved by iterative methods. This is because the linear systems arising in multiphase porous media flow simulations cumulate many difficulties. These systems are non-symmetric, involve several unknowns of different nature per grid cell, display strong or very strong heterogeneities and anisotropies, and change during the simulation. Many researchers focus on these simulations, and many innovative techniques for solving linear systems have been introduced while studying these simulations, as for example the nested factorization [Appleyard and Cheshire, 1983, SPE Symposium on Reservoir Simulation].

4.2. Inverse problems

The research of F. Nataf on inverse problems is rather new since this activity was started from scratch in 2007. Since then, several papers were published in international journals and conference proceedings. All our numerical simulations were performed in FreeFem++.

We focus on methods related to time reversal techniques. Since the seminal paper by [M. Fink et al., Imaging through inhomogeneous media using time reversal mirrors. *Ultrasonic Imaging*, 13(2):199, 1991.], time reversal is a subject of very active research. The main idea is to take advantage of the reversibility of wave propagation phenomena such as it occurs in acoustics, elasticity or electromagnetism in a non-dissipative unknown medium to back-propagate signals to the sources that emitted them. Number of industrial applications have already been developed: touchscreen, medical imaging, non-destructive testing and underwater communications. The principle is to back-propagate signals to the sources that emitted them. The initial experiment, was to refocus, very precisely, a recorded signal after passing through a barrier consisting of randomly distributed metal rods. In [de Rosny and Fink. Overcoming the diffraction limit in wave physics using a time-reversal mirror and a novel acoustic sink. *Phys. Rev. Lett.*, 89 (12), 2002], the source that created the signal is time reversed in order to have a perfect time reversal experiment. Since then, numerous applications of this physical principle have been designed, see [Fink, Renversement du temps, ondes et innovation. Ed. Fayard, 2009] or for numerical experiments [Larmat et al., Time-reversal imaging of seismic sources and application to the great sumatra earthquake. *Geophys. Res. Lett.*, 33, 2006] and references therein.

4.3. Numerical methods for wave propagation in multi-scale media

We are interested in the development of fast numerical methods for the simulation of electromagnetic waves in multi-scale situations where the geometry of the medium of propagation may be described through characteristic lengths that are, in some places, much smaller than the average wavelength. In this context, we propose to develop numerical algorithms that rely on simplified models obtained by means of asymptotic analysis applied to the problem under consideration.

Here we focus on situations involving boundary layers and *localized* singular perturbation problems where wave propagation takes place in media whose geometry or material characteristics are submitted to a small scale perturbation localized around a point, or a surface, or a line, but not distributed over a volumic sub-region of the propagation medium. Although a huge literature is already available for the study of localized singular perturbations and boundary layer phenomena, very few works have proposed efficient numerical methods that rely on asymptotic modeling. This is due to their natural functional framework that naturally involves singular functions, which are difficult to handle numerically. The aim of this part of our research is to develop and analyze numerical methods for singular perturbation methods that are prone to high order numerical approximation, and robust with respect to the small parameter characterizing the singular perturbation.

4.4. Data analysis in astrophysics

We focus on computationally intensive numerical algorithms arising in the data analysis of current and forthcoming Cosmic Microwave Background (CMB) experiments in astrophysics. This application is studied in collaboration with researchers from University Paris Diderot, and the objective is to make available the algorithms to the astrophysics community, so that they can be used in large experiments.

In CMB data analysis, astrophysicists produce and analyze multi-frequency 2D images of the universe when it was 5% of its current age. The new generation of the CMB experiments observes the sky with thousands of detectors over many years, producing overwhelmingly large and complex data sets, which nearly double every year therefore following the Moore's Law. Planck (<http://planck.esa.int/>) is a keystone satellite mission which has been developed under auspices of the European Space Agency (ESA). Planck has been surveying the sky since 2010, produces terabytes of data and requires 100 Petaflops per image analysis of the universe. It is predicted that future experiments will collect half petabyte of data, and will require 100 Exaflops per analysis as early as in 2020. This shows that data analysis in this area, as many other applications, will keep pushing the limit of available supercomputing power for the years to come.

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 2014, we have published results on constructive mechanisms for security and accountability properties in service-based system as well as results on service provisioning problems, in particular, service interoperability and mediation, see Sec. 6.3 . Furthermore, we take part in the European project A4Cloud on accountability challenges, that is, the responsible stewardship of third-party data and computations, see Sec. 8.3 .

4.2. Capacity Planning in Cluster, Grid and Cloud Computing

Cluster, Grid and more recently Cloud computing platforms aim at delivering large capacities of computing power. These capacities can be used to improve performance (for scientific applications) or availability (e.g., for Internet services hosted by datacenters). These distributed infrastructures consist of a group of coupled computers that work together and may 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 call for automation of the corresponding adaptation processes. 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 2014, we have proposed a mechanism to take into account locality aspects in the DVMS proposal, a fully distributed VM scheduler. Concretely, our mechanism leverages Vivaldi coordinates in order to favor live migration of virtual machines between servers belonging to the same site before performing inter-site live migrations. By such a means, we have improved the reactivity of DVMS, establishing it as one of the most scalable and reactive scheduler of virtual machines for large-scale cloud computing infrastructures. Finally, we have also provided several results on the energy efficient management of Cloud applications and infrastructures, see Sec. 6.4 .

In the energy field, we have designed a set of techniques, named OptiPlace, for cloud management with flexible power models through constraint programming. OptiPlace supports external models, named views. Specifically, we have developed a power view, based on generic server models, to define and reduce the power consumption of a datacenter's physical servers. We have shown that OptiPlace behaves at least as good as our previous system, Entropy, requiring as low as half the time to find a solution for the constrained-based placement of tasks for large datacenters.

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 2014, we have extended the language EScala, which integrates reactive programming through events with aspect-oriented and object-oriented mechanisms, see Sec. 6.3 .

ATLANMOD Project-Team

4. Application Domains

4.1. Application domains

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 [47] and to system engineering [38]. There are a lot of problems in these application domains that may be addressed by means of modeling and model transformation techniques.

As a result, AtlanMod has collaborated with a great variety of different companies ranging from the Automotive to the Insurances domains and from SMEs to large enterprises through the projects described later on in this same report. AtlanMod hopes to continue this trend in the future.

AVALON Project-Team

4. Application Domains

4.1. Overview

The Avalon team targets applications with large computing and/or data storage needs, which are still difficult to program, maintain, and deploy. Those applications can be parallel and/or distributed applications, such as large scale simulation applications or code coupling applications. Applications can also be workflow-based as commonly found in distributed systems such as grids or clouds.

The team aims at not being restricted to a particular application field, thus avoiding any spotlight. The team targets different HPC and distributed application fields, which bring use cases with different issues. This will be eased by our various collaborations: the team participates to the INRIA-Illinois Joint Laboratory for Petascale Computing, the Physics, Radiobiology, Medical Imaging, and Simulation French laboratory of excellence, the E-Biothon project, the INRIA large scale initiative Computer and Computational Sciences at Exascale (C2S@Exa), and to BioSyL, a federative research structure about Systems Biology of the University of Lyon. Moreover, the team members have a long tradition of cooperation with application developers such as CERFACS and EDF R&D. Last but not least, the team has a privileged connection with CC IN2P3 that opens up collaborations, in particular in the astrophysics field.

In the following, some examples of representative applications we are targeting are presented. In addition to highlighting some application needs, they also constitute some of the use cases we will use to validate our theoretical results.

4.2. Climatology

The 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. Simulations can be of different types: HPC applications (*e.g.*, the NEMO framework [38] for ocean modelization), code-coupling applications (*e.g.*, the OASIS coupler [44] for global climate modeling), or workflows (long term global climate modeling).

As for most applications the team is targeting, the challenge is to thoroughly analyze climate-forecasting applications to model their needs in terms of programming model, execution model, energy consumption, data access pattern, and computing needs. Once a proper model of an application has been set up, appropriate scheduling heuristics could be designed, tested, and compared. The team has a long tradition of working with CERFACS on this topic, for example in the LEGO (2006-09) and SPADES (2009-12) French ANR projects.

4.3. Astrophysics

Astrophysics is a major field to produce large volume of data. For instance, the Large Synoptic Survey Telescope (<http://www.lsst.org/lsst/>) will produce 15 TB of data every night, with the goals of discovering thousands of exoplanets and of uncovering the nature of dark matter and dark energy in the universe. The Square Kilometer Array (<http://www.skatelescope.org/>) produces 9 Tbits/s of raw data. One of the scientific projects related to this instrument called Evolutionary Map of the Universe is working on more than 100 TB of images. The Euclid Imaging Consortium will generate 1 PB data per year.

Avalon collaborates with the *Institut de Physique Nucléaire de Lyon* (IPNL) laboratory on large scale numerical simulations in astronomy and astrophysics. Contributions of the Avalon members have been related to algorithmic skeletons to demonstrate large scale connectivity, the development of procedures for the generation of realistic mock catalogs, and the development of a web interface to launch large cosmological simulations on GRID'5000.

This collaboration, that continues around the topics addressed by the CLUES project (<http://www.clues-project.org>), has been extended thanks to the tight links with the CC-IN2P3. Major astrophysics projects execute part of their computing, and store part of their data on the resources provided by the CC-IN2P3. Among them, we can mention SNFactory, Euclid, or LSST. These applications constitute typical use cases for the research developed in the Avalon team: they are generally structured as workflows and a huge amount of data (from TB to PB) is involved.

4.4. Bioinformatics

Large-scale data management is certainly one of the most important applications of distributed systems in the future. Bioinformatics is a field producing such kinds of applications. For example, DNA sequencing applications make use of MapReduce skeletons.

The Avalon team is a member of BioSyL (<http://www.biosyl.org>), a Federative Research Structure attached to University of Lyon. It gathers about 50 local research teams working on systems biology. Moreover, the team cooperates with the French Institute of Biology and Chemistry of Proteins (IBCP <http://www.ibcp.fr>) in particular through the ANR MapReduce project where the team focuses on a bio-chemistry application dealing with protein structure analysis. These collaborations bring scientific applications that are both dynamic and data-intensive.

CIDRE Project-Team

4. Application Domains

4.1. Domain

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, in which security (and safety) is a major concern can 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 or cloud services, social networks, location-based services, etc.) can also benefit from results obtained by CIDRE, in particular to solve some of the privacy issues raised by these systems that manipulate huge amount of personal data. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. Cloud computing brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

COAST Team (section vide)

COATI Project-Team

4. Application Domains

4.1. Telecommunication networks

COATI is mostly interested in telecommunications networks. 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 SME like 3-Roam.

We focus on the design and management of heterogeneous networks. The project has kept working on the design of backbone networks (optical networks, radio networks, IP networks). We also study routing algorithms such as dynamic and compact routing schemes in the context of the FP7 EULER led by Alcatel-Lucent Bell-Labs (Belgium), and the evolution of the routing in case of any kind of topological modifications (maintenance operations, failures, capacity variations, etc.).

4.2. Other domains

Our combinatorial tools may be well applied to solve many other problems in various areas (transport, biology, resource allocation, chemistry, smart-grids, speleology, etc.) and we intend to collaborate with teams of these other domains.

For instance, we have recently started a collaboration in Structural Biology with EPI ABS (Algorithms Biology Structure) from Sophia Antipolis (described in Section 6.2). Furthermore, we are also working on robot moving problems coming from Artificial Intelligence/Robotic with Xavier Defago (Associate Professor at Japan Advanced Institute of Science and Technology, Japan). We have also started a collaboration with Amadeus on complex journey planning.

CTRL-A Exploratory Action

4. Application Domains

4.1. Distributed systems and High-Performance Computing

Distributed systems have grown to levels of scale and complexity where it is difficult to master their administration and resources management, in dynamic and open environments. One of the growing concerns is that the energy consumption has reached levels where it can not be considered negligible anymore, ecologically or economically. Data centers or high performance computing grids need to be controlled in order to combine minimized power needs with sustained performance and quality of service. As mentioned above, this motivates the automation of their management, and is the major topic of, amongst others, our ANR project Ctrl-Green (see 8.2.1).

Another challenge in distributed systems is in the fast growing amounts of data to process and store. Currently one of the most common ways of dealing with these challenges is the parallel programming paradigm MapReduce which is slowly becoming the de facto tool for Big Data analytics. While its use is already widespread in the industry, ensuring performance constraints while also minimizing costs provides considerable challenges. Current approaches to ensure performance in cloud systems can be separated into three categories: static, reactive, predictive and hybrid approaches. In the industry, static deployments are the standard and usually tuned based on the application peak demand and are generally over-provisioned. Reactive approaches are usually based on reacting to an input metric such as the current CPU utilisation, request rate, response time by adding and removing servers as necessary. Some public cloud providers offer reactive techniques such as the Amazon Auto Scaler. They provide the basic mechanisms for reactive controllers, but it is up to the user to define the static scaling thresholds which is difficult and not optimal. To deal with this issue, we propose a control theoretical approach, based on techniques that have already proved their usefulness for the control community.

In the domain of parallel systems and High Performance Computing, systems are traditionally less open and more controlled by administrators, but this trend is changing, as they are facing the same challenges in energy consumption, needs for adaptivity in reaction to changing workloads, and security issues in computation outsourcing. Topics of interest for us in this domain concern problem in dynamical management of memory and communications features, which we are exploring in the HPES project of the Labex Persybal-lab (see 8.1).

4.2. Reconfigurable architectures in embedded systems

Dynamically reconfigurable hardware has been identified as a promising solution for the design of energy efficient embedded systems. A common argument in favor of this kind of architecture is the specialization of processing elements, that can be adapted to application functions in order to minimize the delay, the control cost and to improve data locality. Another key benefit is the hardware reuse to minimise the area, and therefore the static power and cost. Further advantages such as hardware updates in long-life products and self-healing capabilities are also often mentioned. In presence of context changes (e.g. environment or application functionality), self-adaptive technique can be applied as a solution to fully benefit from the runtime reconfigurability of a system.

Dynamic Partial Reconfiguration (DPR) of FPGA is another accessible solution to implement and experiment reconfigurable hardware. It has been widely explored and detailed in literature. However, it appears that such solutions are not extensively exploited in practice for two main reasons: i) the design effort is extremely high and strongly depends on the available chip and tool versions; and ii) the simulation process, which is already complex for non-reconfigurable systems, is prohibitively large for reconfigurable architectures. As a result, new adequate methods are required to fully exploit the potential of dynamically reconfigurable and

self-adaptive architectures. We are working in this topic, especially on the reconfiguration control aspect, in cooperation with teams specialized in reconfigurable architectures such as the former DaRT team at Inria Lille, and LabSticc in Lorient, as in the recently ended ANR project Famous.

4.3. Smart environments and Internet of Things

Another application domain for autonomic systems design and control is the Internet of Things, and especially the design of smart environments, at the level of homes, buildings, or cities. These domains are often considered at the level of sensors networks, with a strong emphasis on the acquisition of data in massive scales. The infrastructures are sometimes also equipped with actuators, with a wide range of applications, for example concerning lighting or heating, or access and security aspects. We are interested in closing the control loop in such environments, which is less often studied. In particular, rule-based languages are often used to define the automated systems, and we want to contribute to the safe design of such controllers with guarantees on their behaviors. We are working in this topic in cooperation with teams specialized in infrastructures for smart environments at CEA LETI/DACLE and Orange labs.

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 make 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/iBird 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 science of complex networks has been assigned an increasingly relevant role in defining a conceptual framework for the analysis of complex systems. Network science is concerned with graphs that map entities and their interactions to nodes and links. For a long time, this mathematical abstraction has contributed to the understanding of real-world systems in physics, computer science, biology, chemistry, social sciences, and economics. Recently, however, enormous amounts of detailed data, electronically collected and meticulously catalogued, have finally become available for scientific analysis and study. This has led to the discovery that most networks describing real world systems show the presence of complex properties and heterogeneities, which cannot be neglected in their topological and dynamical description. This has called forth a major effort in developing the methodology to characterize the topology and temporal behavior of complex networks, to describe the observed structural and temporal heterogeneities, to detect and measure emerging community structure, to see how the functionality of networks determines their evolving structure, and to determine what kinds of correlations play a role in their dynamics. All these efforts have brought us to a point where the science of complex networks has become advanced enough to help us to disclose the deeper roles of complexity and gain understanding about the behavior of very complicated systems.

In this endeavor the DANTE project targets the study of dynamically evolving networks, concentrating on questions about the evolving structure and dynamical processes taking place on them. During the last year we developed several projects along these lines concerning three major datasets:

- Mobile telephony data: In projects with academic partners and Grandata we performed projects based on two large independent datasets collecting the telephone call and SMS event records for million of anonymized individuals. The datasets record the time and duration of mobile phone interactions and some coarse grained location and demographic data for some users. In addition one of the dataset is coupled with anonymised bank credit information allowing us to study directly the socioeconomic structure of a society and how it determines the communication dynamics and structure of individuals.
- Skype data: Together with Skype Labs/STACC and other academic groups we were leading projects in the subject of social spreading phenomena. These projects were based on observations taken from a temporally detailed description of the evolving social network of (anonymized) Skype users registered between 2003 and 2011. This data contains dates of registration and link creation together with gradual information about their location and service usage dynamics.

- Twitter data: In collaboration with ICAR-ENS Lyon we collected a large dataset about the microblogs and communications of millions of Twitter users in the French Twitter space. This data allows us to follow the spreading of fads/opinions/hashtags/ideas and more importantly linguistic features in online communities. The aim of this collaboration is to set the ground for a quantitative framework studying the evolution of linguistic features and dialects in an social-communication space mediated by online social interactions.

DIANA Team (section vide)

DIONYSOS Project-Team (section vide)

DIVERSE Project-Team

4. Application Domains

4.1. From Embedded Systems to Service Oriented Architectures

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.

DYOGENE Project-Team

4. Application Domains

4.1. Embedded Networks

Critical real-time embedded systems (cars, aircrafts, spacecrafts) are nowadays made up of multiple computers communicating with each other. The real-time constraints typically associated with operating systems now extend to the networks of communication between sensors/actuators and computers, and between the computers themselves. Once a media is shared, the time between sending and receiving a message depends not only on technological constraints, but also, and mainly from the interactions between the different streams of data sharing the media. It is therefore necessary to have techniques to guarantee maximum network delays, in addition to local scheduling constraints, to ensure a correct global real-time behaviour to distributed applications/functions.

Moreover, pessimistic estimate may lead to an overdimensioning of the network, which involves extra weight and power consumption. In addition, these techniques must be scalable. In a modern aircraft, thousands of data streams share the network backbone. Therefore algorithm complexity should be at most polynomial.

4.2. Wireless Networks

Wireless networks can be efficiently modelled as dynamic stochastic geometric networks. Their analysis requires taking into account, in addition to their geometric structure, the specific nature of radio channels and their statistical properties which are often unknown a priori, as well as the interaction through interference of the various individual point-to-point links.

4.3. Peer-to-Peer Systems

The amount of multimedia traffic accessed via the Internet, already of the order of exabytes (10^{18} bytes) per month, is expected to grow steadily in the coming years. A peer-to-peer (P2P) architecture, where peers contribute resources to support service of such traffic, holds the promise to support its growth more cheaply than by scaling up the size of data centers. More precisely, a large scale P2P system based on resources of individual users can absorb part of the load that would otherwise need to be served by data centers. In video-on-demand applications, the critical resources at the peers are storage space and uplink bandwidth. Our objective is to ensure that the largest fraction of traffic is supported by the P2P system.

4.4. Social and Economic Networks

Networks are ubiquitous with the presence of different kinds of social, economic and information networks around us. The Internet is one of the most prominent examples of a geometric network. We also examine geometric networks from the perspective of sociologist and economist [55]. Network analysis is also attracting fundamental research by computer scientists [49]. Diffusion of information, social influence, trust, communication and cooperation between agents are heavily researched topics in e-commerce and multi-agent systems. Our probabilistic techniques are very appropriate in this case and have been largely neglected so far. While the first works on geometric networks emanated from theoretical physicists, they stay more focused on static properties of such networks and do not consider game theoretical or statistical learning (like community detection) aspects of such networks. This leaves open a range of new problems to which we will contribute.

4.5. Routing Protocols

Routing protocols enables to maintain paths for transmitting messages over a network. Those protocols, such as OSPF, are based on the transmission of periodic messages between neighbors. Nowadays, faulty behaviors result in the raising of alarms, but are mostly detected when a breakdown or a major misbehavior occurs. Indeed, alarms are so numerous that they cannot be analyzed efficiently. We aim at developing methods to detect misbehaviours of a router before a major fault occurs, and techniques to study the influence of the protocol parameters on the behavior of the network.

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.

FUN Project-Team (section vide)

GANG Project-Team

4. Application Domains

4.1. 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 of such studies.

HIEPACS Project-Team

4. Application Domains

4.1. Material physics

Participants: Pierre Blanchard, Olivier Coulaud, Arnaud Etcheverry, Matthias Messner.

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 are two ways to decrease the computational cost. In the first approach, we improve algorithms and their parallelization and in the second way, we will consider a multiscale approach.

A 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 modeling 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) modeling. 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 currently under definition **OPTIDIS** 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.

4.2. Co-design for scalable numerical algorithms in scientific applications

Participants: Pierre Brenner, Jean-Marie Couteyen, Mathieu Faverge, Luc Giraud, Xavier Lacoste, Guillaume Latu, Salli Moustapha, Pierre Ramet, Fabien Rozar, Jean Roman, Pablo Salas, Xavier Vasseur.

The research activities concerning the ITER challenge are involved in the Inria Project Lab (IPL) **C2S@EXA**.

4.2.1. MHD instabilities edge localized modes

The numerical simulations tools designed for ITER challenges aim at making a significant progress in understanding active control methods of plasma edge MHD instabilities Edge Localized Modes (ELMs) which represent particular danger with respect to heat and particle loads for Plasma Facing Components (PFC) in ITER. Project is focused in particular on the numerical modeling study of such ELM control methods as Resonant Magnetic Perturbations (RMPs) and pellet ELM pacing both foreseen in ITER. The goals of the project are to improve understanding the related physics and propose possible new strategies to improve effectiveness of ELM control techniques. The tool for the nonlinear MHD modeling (code **JOREK**) will be largely developed within the present project to include corresponding new physical models in conjunction with new developments in mathematics and computer science strategy in order to progress in urgently needed solutions for ITER.

The fully implicit time evolution scheme in the **JOREK** code leads to large sparse linear systems that have to be solved at every time step. The MHD model leads to very badly conditioned matrices. In principle the **PaStiX** library can solve these large sparse problems using a direct method. However, for large 3D problems the CPU time for the direct solver becomes too large. Iterative solution methods require a preconditioner adapted to the problem. Many of the commonly used preconditioners have been tested but no satisfactory solution has been found. The research activities presented in Section 3.3 will contribute to design new solution techniques best suited for this context.

4.2.2. Turbulence of plasma particules inside a tokamak

In the context of the ITER challenge, the **GYSELA** project aims at simulating the turbulence of plasma particules inside a tokamak. Thanks to a better comprehension of this phenomenon, it would be possible to design a new kind of source of energy based of nuclear fusion. Currently, **GYSELA** is parallalized in a MPI/OpenMP way and can exploit the power of the current greatest supercomputers (e.g., Juqueen). To simulate faithfully the plasma physic, **GYSELA** handles a huge amount of data. In fact, the memory consumption is a bottleneck on large simulations (449 K cores). In the meantime all the reports on the future Exascale machines expect a decrease of the memory per core. In this context, mastering the memory consumption of the code becomes critical to consolidate its scalability and to enable the implementation of new features to fully benefit from the extreme scale architectures.

In addition to activities for designing advanced generic tools for managing the memory optimisation, further algorithmic research will be conduced to better predict and limit the memory peak in order to reduce the memory footprint of **GYSELA**.

4.2.3. SN Cartesian solver for nuclear core simulation

As part of its activity, EDF R&D is developing a new nuclear core simulation code named COCAGNE that relies on a Simplified PN (SPN) method to compute the neutron flux inside the core for eigenvalue calculations. In order to assess the accuracy of SPN results, a 3D Cartesian model of PWR nuclear cores has been designed and a reference neutron flux inside this core has been computed with a Monte Carlo transport code from Oak Ridge National Lab. This kind of 3D whole core probabilistic evaluation of the flux is computationally very demanding. An efficient deterministic approach is therefore required to reduce the computation effort dedicated to reference simulations.

In this collaboration, we work on the parallelization (for shared and distributed memories) of the DOMINO code, a parallel 3D Cartesian SN solver specialized for PWR core reactivity computations which is fully integrated in the COCAGNE system.

4.2.4. 3D aerodynamics for unsteady problems with moving bodies

ASTRIUM has developed for 20 years the FLUSEPA code which focuses on unsteady phenomenon with changing topology like stage separation or rocket launch. The code is based on a finite volume formulation with temporal adaptive time integration and supports bodies in relative motion. The temporal adaptive integration

classifies cells in several temporal levels, zero being the level with the slowest cells and each level being twice as fast as the previous one. This repartition can evolve during the computation, leading to load-balancing issues in a parallel computation context. Bodies in relative motion are managed through a CHIMERA-like technique which allows building a composite mesh by merging multiple meshes. The meshes with the highest priorities recover the least ones, and at the boundaries of the covered mesh, an intersection is computed. Unlike classical CHIMERA technique, no interpolation is performed, allowing a conservative flow integration. The main objective of this research is to design a scalable version of FLUSEPA in order to run efficiently on modern parallel architectures very large 3D simulations.

HIPERCOM2 Team

4. Application Domains

4.1. Introduction

The HIPERCOM2 team addresses the following application domains:

- military, emergency or rescue applications,
- industrial applications,
- vehicular networks,
- smart cities,
- Internet of Things.

These application domains use the four types of wireless networks:

- wireless mesh and mobile ad hoc networks,
- wireless sensor networks,
- vehicular networks,
- cognitive radio networks.

4.2. Wireless mesh and mobile ad hoc networks

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 historically the first application of mobile ad hoc networks.

The rescue context is halfway between military and civilian applications. In emergency applications, heterogeneous wireless networks have to cooperate in order to save human lives or bring the situation back to normal as soon as possible. Wireless networks that can be quickly deployed are very useful to assess damages and take the first decisions appropriate to the disaster of natural or human origin. The primary goal is to maintain connectivity with the humans or the robots (in case of hostile environment) in charge of network deployment. This deployment should ensure the coverage of an interest area or of only some interest points. The wireless network has to cope with pedestrian mobility and robots/vehicles mobility. The environment, initially unknown, is progressively discovered and usually has many obstacles. These obstacles should be avoided. The nodes of the wireless network are usually battery-equipped. Since they are dropped by a robot or a human, their weight is very limited. The protocols supported by these nodes should be energy efficient to increase network lifetime. Furthermore, in case of aggressive environment, sensor nodes should be replaced before failing. Hence, in such conditions, it is required to predict the failure time of nodes to favor a predictive maintenance.

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. Vehicular Networks and Smart Cities

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.

Smart cities share with the military tactical networks the constraint on pedestrian and vehicular mobility. Furthermore, the coexistence of many networks operating in the same radio spectrum may cause interferences that should be avoided. Cognitive radio takes advantage of the channels temporarily left available by the primary users to assign them to secondary users. Such an opportunistic behavior can also be applied in wireless sensor networks deployed in the cities. Smart cities raise the problem of transmitting, gathering, processing and storing big data. Another issue is to provide the right information at the right place: where it is needed.

4.4. Wireless sensor networks in industrial applications and Internet of Things

Concerning wireless sensor networks, WSNs, we tackle the three following issues:

- Energy efficiency is a key property in wireless sensor networks. Various techniques contribute to save energy of battery-equipped sensor nodes. To name a few, they are: energy efficient routing protocols, node activity scheduling, adjustment of transmission power, reduction of protocols overhead, reduction of data generated and transmitted. 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. We have observed the great benefit obtained with node activity scheduling. In networks with low activity, opportunistic strategies are used to address low duty cycles.
- Large scale WSNs constitute another challenge. 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.
- Multichannel WSNs provide an opportunity:
 - to increase the parallelism between transmissions. Hence, it reduces the data gathering delays and improves the time consistency of gathered data.
 - to increase the robustness against interferences and perturbations possibly caused by the coexistence of other wireless networks.

4.5. Cognitive Radio Networks

Usually in cognitive radio, the secondary users are in charge of monitoring the channel to determine whether or not the primary users are active in the area. If they are not, the secondary users are allowed to use the spectrum left unused by the primary users. We are interested in two issues:

- Design and modeling of a new access scheme based on a generalized Carrier Sense Multiple Access scheme using active signaling. This scheme allows the primary users to capture the bandwidth even if the secondary users are transmitting in the area.
- Design of a time slot and channel assignment to minimize the data gathering performed by secondary users. This assignment should work with different detection schemes of primary user presence.

INDES Project-Team

4. Application Domains

4.1. Web programming

Along with games, multimedia applications, electronic commerce, and email, the web has popularized computers for daily 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. Robotics

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 robot as a web entity, and the use of remote web services to manage, monitor or extend the features of the robot. Among the direct benefits of relying on a web framework for robotics are the ability to use any web enabled device such as a smartphone or tablet to drive the robot.

INFINE Team

4. Application Domains

4.1. Panorama

The research in INFINE spans a wide range of application areas ranging from Internet-based, wireless sensor-based, mobile wireless-based, and OSN-based applications. These applications are related to the three main research axes described in the previous sections.

4.2. Mobile wireless network

Smart portable devices such as smartphones, PDAs or tablet PCs are being considered as pervasive mobile sensing platforms due to their increasing proliferation and their wide range of embedded heterogeneous capabilities (in terms of type of communication and data gathering possibilities - e.g., 3G, WiFi, GPS, video, camera, etc). Such devices are changing the way people are communicating, generating, and exchanging data: They allow the free sensing/gathering of data of the surrounding environment anytime and anywhere. On the other hand, the projected increase of mobile data traffic demand pushes towards additional complementary offloading methods. Novel mechanisms are thus needed, which must fit both the new context that Internet users experience now, and their forecasted demands.

In these contexts, the application domains that we are targeting are related to traffic offloading in large-scale mobile wireless networks. Among the numerous offloading solutions fitting in this application domain, we are specially interested in the ones related to: infrastructure deployment, traffic modeling, opportunistic communication, or still task delegation. A core principle of such solutions is the understanding and modeling of users behavior in terms of their context (i.e., imposed by mobility) and their content demands.

4.3. Online Social Networks

Our high-level goal here is to help increase the relevance of content accessed by users, through the elaboration of contact and content recommendation mechanisms, as well as incentive mechanisms. The scientific context in which we phrase this goal is that of:

- modeling information propagation in OSN;
- statistical inference problems raised by the search for improved information propagation. In particular these include community detection for contact and content recommendation, and bandit-like algorithms for active learning of given content type at limited “spamming” cost;
- the mechanism design branch of economic theory, which can be leveraged to conceive reward mechanisms meant to incentivize efficient collaborative content filtering by OSN users.

4.4. Spontaneous Wireless Networks applications

The advances in hardware development have made possible the miniaturization of micro-electro-mechanical systems and consequently, the development of sensor networks. The combination of inexpensive, autonomous, low-power sensing, and compact devices has established the viability of deploying large and dense wireless sensor networks (WSNs) able to sense the physical world. By essence, such networks require fully decentralized solutions in which the load is evenly balanced in the system, merely because participating entities have limited in power, storage and communication capabilities. Thus one of the applications of Spontaneous Wireless Networks has been traditionally such wireless sensor networks, where some typical applications are to continuously monitor data (real-time data collection to a sink), and to be able to do manage network after deployment (for instance reflashing nodes with firmware over the air). The challenge is to operate this with standards (such as IP), constrained devices (battery, memory, power, ...), which requires sophisticated protocols, with reliable and tested implementations.

The applications of the more recent “Internet of Object” are much broader, since they literally consists of any application running on any object (in the industrial factories, in living spaces, ...). While some of the constraints in wireless sensor networks are a still present in IoT in general, what characterizes IoT is the heterogeneity of the platforms.

KerData Project-Team

4. Application Domains

4.1. 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. It has remained poorly understood so far. Both neuroimaging- and genetic-domain observations include a huge amount of variables (of the order of millions). Performing rigorous statistical 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 was carried out within the Microsoft Research-Inria Joint Research Center. It was co-led by the KerData (Rennes) and Parietal (Saclay) Inria teams. They jointly address this computational problem using cloud related techniques on the 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. This project is a typical multi-disciplinary Data Science project which serves as background for several on-going research activities.

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.

4.2. 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 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. A 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, using a cloud computing facility unlocks 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 update all derivative predictive models periodically with new experimental data. The applications in this complete chemo-proteomic vision address the identification of new druggable protein targets, and thereby the generation of new drug candidates.

4.3. 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 record and visualize data during the simulation efficiently without impacting the performance of the computation generating that data. Conventional practice consists in storing data on disk, moving them off-site, reading them into a workflow, and analyzing them. This approach becomes increasingly harder to use because of the large data volumes generated at fast rates, in contrast to limited back-end performance. 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.

In this context, the KerData project-team started to explore ways to remove the limitations mentioned above through 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 approach, which typically consists in periodically writing a very large number of small files. This causes bursts 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 these challenges through an approach based on dedicated I/O cores.

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 services 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. Main Application Domains

MAESTRO's main application area is networking, to which we apply modeling, performance evaluation, optimization and control. Our primary focus is on protocols and network architectures, and recent evolutions include the study of the Web and social networks, as well as models for Green IT.

- 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.

MESCAL Project-Team

4. Application Domains

4.1. Cloud, Grid, Multi-core and Desktop Computing

Participants: Arnaud Legrand, Olivier Richard, Jean-Marc Vincent.

Software tools were developed to carry experiments on clouds and grids (Kameleon and Expo). Other tools (Pajé, Viva, Framesoc and Ocelotl) have been designed to monitor, trace and analyse applications running on multi-core and grid computers. Such traces have also been used in SIMGRID to simulate volunteer computing systems at unprecedented scale.

4.2. Wireless Networks

Participants: Bruno Gaujal, 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. We have also been working on optimal spectrum management of MIMO systems, routing in ad-hoc works and power allocation in future 5G networks.

4.3. 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.

4.4. Energy and Transportation

Participant: Nicolas Gast.

This work is mainly done within the Quanticol European project.

Smart urban transport systems and smart grids are two examples of collective adaptive systems. They consist of a large number of heterogeneous entities with decentralised control and varying degrees of complex autonomous behaviour. Within the QUANTICOL project, we develop an analysis tools to help to reason about such systems. Our work relies on tools from fluid and mean-field approximation to build decentralized algorithms that solve complex optimization problems. We focus on two problems: decentralized control of electric grids and capacity planning in vehicle-sharing systems to improve load balancing.

MIMOVE Team

4. Application Domains

4.1. Mobile urban systems for smarter cities

With the massive scale adoption of mobile devices and further expected significant growth in relation with the Internet of Things, mobile computing is impacting most -if not all- the ICT application domains. However, given the importance of conducting empirical studies to assess and nurture our research, we focus on one application area that is the one of "*smart cities*". The smart city vision anticipates that the whole urban space, including buildings, power lines, gas lines, roadways, transport networks, and cell phones, can all be wired together and monitored. Detailed information about the functioning of the city then becomes available to both city dwellers and businesses, thus enabling better understanding and consequently management of the city's infrastructure and resources. This raises the prospect that cities will become more sustainable environments, ultimately enhancing the citizens' well being. There is the further promise of enabling radically new ways of living in, regulating, operating and managing cities, through the increasing active involvement of citizens by ways of crowd-sourcing/sensing and social networking.

Still, the vision of what smart cities should be about is evolving at a fast pace in close concert with the latest technology trends. It is notably worth highlighting how mobile and social network use have reignited citizen engagement, thereby opening new perspectives for smart cities beyond data analytics that have been initially one of the core foci for smart cities technologies. Similarly, open data programs foster the engagement of citizens in the city operation and overall contribute to make our cities more sustainable. The unprecedented democratization of urban data fueled by open data channels, social networks and crowd sourcing enables not only the monitoring of the activities of the city but also the assessment of their nuisances based on their impact on the citizens, thereby prompting social and political actions. However, the comprehensive integration of urban data sources for the sake of sustainability remains largely unexplored. This is an application domain that we intend to focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In a first step, we concentrate on the following specialized applications, which we investigate in close collaboration with other researchers, in particular as part of the dedicated Inria Project Lab *CityLab@Inria* under creation:

- **Democratization of urban data for healthy cities.** The objective here is to integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances from raw pollution sensing (e.g., sensing noise) to the sensing of its impact on citizens (e.g., how people react to urban noise and how this affects their health).
- **Socially-aware urban mobility.** Mobility within mega-cities is known as one of the major challenges to face urgently due to the fact that today's mobility patterns do not scale and to the negative effect on the environment and health. It is our belief that mobile social and physical sensing may significantly help in promoting the use of public transport, which we have started to investigate through empirical study based on the development and release of dedicated apps.
- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. This is especially studied with the Social Apps Lab at UC Berkeley. Our objective is to study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development and development of such "*civic apps*".

Acknowledging the need for collaborative research in the application domain of smart cities, MiMove is heavily involved and actually leading the effort of creating CityLab@Inria⁰. CityLab is focused on the study of ICT solutions promoting social sustainability in smart cities, and involves the following Inria project-teams in addition to MiMove: CLIME, DICE, FUN, MYRIADS, OAK, SMIS, URBANET and WILLOW. CityLab further involves strong collaboration with Californian universities affiliated with CITRIS (Center for Information Technology Research in the Interest of Society) and especially UC Berkeley, in relation with the *Inria@SiliconValley* program. We note that Valérie Issarny acts as scientific manager of Inria@SiliconValley and is currently on leave at UC Berkeley. In this context, MiMove researchers are working closely with colleagues of UC Berkeley, including researchers from various disciplines interested in smart cities (most notably sociologists).

⁰<https://citylab.inria.fr>

MOAIS Project-Team (section vide)

MUSE Team

4. Application Domains

4.1. Home Network Diagnosis

With the availability of cheap broadband connectivity, Internet access from the home has become a ubiquity. Modern households host a multitude of networked devices, ranging from personal devices such as laptops and smartphones to printers and media centers. These devices connect among themselves and to the Internet via a local-area network—a *home network*—that has become an important part of the “Internet experience”. In fact, ample anecdotal evidence suggests that the home network can cause a wide array of connectivity impediments, but their nature, prevalence, and significance remain largely unstudied.

Our long-term goal is to assist users with concrete indicators of the causes of potential problems and—ideally—ways to fix them. We intend to develop a set of easy-to-use home network diagnosis tools that can reliably identify performance and functionality shortcomings rooted in the home. The development of home network diagnosis tools brings a number of challenges. First, home networks are heterogenous. The set of devices, configurations, and applications in home networks vary significantly from one home to another. We must develop sophisticated techniques that can learn and adapt to any home network as well as to the level of expertise of the user. Second, there are numerous ways in which applications can fail or experience poor performance in home networks. Often there are a number of explanations for a given symptom. We must devise techniques that can identify the most likely cause(s) for a given problem from a set of possible causes. Third, even if we can identify the cause of the problem, we must then be able to identify a solution. It is important that the output of the diagnosis tools we build is “actionable”. Users should understand the output and know what to do.

We are conceiving methods for two application scenarios: (i) when the end user in the home deploys our diagnostic tools either on the home gateway (the gateway often combines a DSL/cable modem and an access point; it connects the home network to the ISP) or on devices connected to the home network and (ii) when ISPs collect measurements from homes of subscribers and then correlate these measurements to help identify problems.

Assisting end users. We are developing algorithms to determine whether network performance problems lie inside or outside the home network. Given that the home gateway connects the home with the rest of the Internet, we are designing an algorithm (called *WTF*) that analyzes traffic that traverses the gateway to distinguish access link and home network bottlenecks. A measurement vantage point on the gateway is key for determining if the performance bottleneck lies within the home network or the access ISP, but we also need to deploy diagnosis tools in end-devices. First, some users may not want (or not know how) to deploy a new home gateway in their homes. Second, some problems will be hard to diagnose with only the vantage point of the gateway (for example, when a device cannot send traffic or when the wireless is poor in certain locations of a home). We can obtain more complete visibility by leveraging *multiple* measurement nodes around the home, potentially including the home gateway, all participating jointly in the measurement task. We have an ongoing project to realize a home network analyzer as a web-based measurement application built on top of our team’s recently developed browser-based measurement platform, *Fathom*. To integrate the home gateway in the analyzer, we plan to engage the BISmark Project. BISmark already provides a web server as well as extensive configurability, allowing us to experiment freely with both passive as well as active measurements. We must develop a home network analyzer that can first discover the set of devices connected to the home network that can collaborate on the diagnosis task. We will then develop tomography algorithms to infer where performance problems lie given measurements taken from the set of available vantage points.

Assisting Internet Service Providers (ISPs). Our discussions with several large access ISPs reveal that service calls are costly, ranging from \$9–25 per call, and as many as 75% of service calls from customers are usually caused by problems that have nothing to do with the ISP. Therefore, ISPs are eager to deploy techniques to assist in home network diagnosis. In many countries ISPs control the home gateway and set-top-boxes in the home. We plan to develop more efficient mechanisms for home users to report trouble to their home ISP and consequently reduce the cost of service calls. This project is in collaboration with Technicolor and Portugal Telecom. Technicolor is a large manufacturer of home gateways and set-top-boxes. Portugal Telecom is the largest broadband access provider in Portugal. Technicolor already collects data from 200 homes in Portugal. We are working with the data collected in this deployment together with controlled experiments to develop methods to diagnose problems in the home wireless.

4.2. Quality of Experience

Understanding how users react to different levels of network performance presents two main challenges:

1. User perception is subjective and contextual. Different users may have different tolerance levels to network performance and the same user may have different expectations under different circumstances. Take for example the round-trip time (RTT), a typical network performance metric. If RTTs are larger than usual, a user who is doing remote login may feel that the connection is unusable, whereas another who is watching YouTube may notice no problem (because YouTube has a playout buffer to mask some network delay). Take another example of a user downloading her email. This user may tolerate some delay when she is leisurely checking her email at home, but she may become extremely frustrated with the same delay if she is in an airplane and needs to download her email just before takeoff.
2. It is challenging to “measure” users. We must develop methods to measure the user perception of network performance as users perform their routine online tasks. It is hence important that these methods are not too intrusive. Otherwise, users are unlikely to participate in the experiment. In addition, we must capture user perception at different levels of performance and in a variety of scenarios.

We will develop tools that run on end systems to collect network performance data annotated with the user perception. These tools will adopt a hybrid measurement methodology that combines network measurement techniques to infer application performance with techniques from HCI to measure user perception. We will later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user’s traffic observed in the network. Models of user perception of network performance can be used to detect when performance is poor to trigger diagnosis or to adapt network/application performance to better serve users.

4.3. Crowd-sourced content recommendation

The Internet today serves as a large content distribution platform (online content varies from traditional news, TV series, and movies to specialized blogs and family pictures shared over social networks) as well as a platform for users to exchange opinions about practically everything (from movies to services and restaurants). The amount of information available online today overwhelms most users and selecting which content to watch or what do has become a challenge. We are applying passive measurement methods and content summarisation techniques to help users to identify relevant content in two scenarios. First, we are developing a system called WeBrowse that passively observes network traffic to extract user clicks (i.e., the URLs users visit). A user click is a good measure of interest, as users often have an idea of the type of content they are about to access (e.g., because they saw a preview or because a friend recommended it). Intuitively, the more users click on a URL, the higher the interest in the content on the corresponding page. WeBrowse then promotes “hottest” and most popular content to users of a network. We have a deployment of WeBrowse in a campus network. Second, we are working on techniques to summarise user feedback (for example, movie or restaurant reviews) with semi-structured feedback. Today reviews are either free-form text or star rating. Star rating is too coarse to capture the nuances of why a user likes or dislikes something, whereas free text is hard for users to parse and

extract a clear opinion. We are instead working with semi-structured reviewing where users enter *tags* (a short sequence of words describing the user experience). We are working with Technicolor on the summarisation of movie reviews and on building a mobile app (called TagIt) where users can review movies directly with tags.

MYRIADS Project-Team (section vide)

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 cognitive assistance.

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. Smart Cities

The Internet of Things (IoT) has become a reality with the emergence of Smart Cities, populated with large amounts of smart objects which are used to deliver a range of citizen services (e.g., security, well being, etc.) The IoT paradigm relies on the pervasive presence of smart objects or “things”, which raises a number of new challenges in the software engineering domain.

4.3.1. The Object's World project

There are an abundance of research and industry initiatives that have been undertaken with the aim of promoting the emergence of IoT [1]. In line with this goal, the Object's World project brings together stakeholders from different domains to build and support the emergence of an IoT sector in France and beyond. The project is lead by SIGFOX, the world's first cellular network operator dedicated to low-bandwidth wireless objects. The cooperation between industry and research partners (e.g., sensor manufacturers, computer science and electrical engineering research labs) is of uttermost importance in overcoming technological barriers. This issue is currently hindering the development of an IoT sector. The main objectives of this project are the development of:

- expertise in the low-bandwidth network sector,
- low-cost transmitter/receiver chips,
- low-energy autonomous sensors, and
- software frameworks which cover the entire lifecycle of IoT applications.

Network infrastructures which support huge numbers of objects open up a range of opportunities for innovative services. Critically, these new opportunities rely on the ability to address the software engineering challenges of this new sector. We promote an approach that revolves around software frameworks. In areas such as mobile and web development, this approach has already been shown to facilitate software development by abstracting over implementation details and guiding the programmer.

4.4. Assistive Technology for Cognition

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 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 platform that is open-ended in terms of applications and entities. (1) An online 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 (whether hardware or software) can be added to a platform description to enhance its functionalities and extend the scope of future applications.

RAP Project-Team (section vide)

REGAL Project-Team (section vide)

RMOD Project-Team

4. Application Domains

4.1. Programming Languages and Tools

Many of the results of RMoD are improving programming languages or development tools for such languages. As such the application domain of these results is as varied as the use of programming languages in general. Pharo, the language that RMoD develops, is used for a very broad range of applications. From pure research experiments to real world industrial use (the Pharo Consortium has over 10 company members) <http://consortium.pharo.org> Examples are web applications, server backends for mobile applications or even graphical tools and embedded applications.

4.2. Software Reengineering

Moose is a language-independent environment for reverse- and re-engineering complex software systems. Moose provides a set of services including a common meta-model, metrics evaluation and visualization. As such Moose is used for analysing software systems to support understanding and continuous development as well as software quality analysis.

ROMA Team

4. Application Domains

4.1. 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 5.1) are: structural mechanics, biomechanics, medical image processing, tomography, geophysics, electromagnetism, fluid dynamics, econometric models, oil reservoir simulation, magneto-hydro-dynamics, chemistry, acoustics, glaciology, astrophysics, circuit simulation, and work on hybrid direct-iterative methods.

RUNTIME Team

4. Application Domains

4.1. Application Domains

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.).

SCALE Team

4. Application Domains

4.1. Simulation

4.1.1. Discrete-event simulation

Simulation is an example of an application with ever increasing computation needs that would benefit from the SCALE research results. In emergency planning and response, for example, users need to access the power of large scale distributed computing facilities to run faster than real-time simulations of the situations they face on the field; Such a computation can mix heterogeneous distributing computing platforms (PDA and laptops on the field, Cloud and HPC in background) and use a number of external services (eg. weather forecast).

Simulations made of multi-party contributed software models also demonstrate the need for a unifying and user-friendly programming model. Indeed, since the early 70's, the simulation field have been the subject of many efforts in order to abstract the computation models from their actual application domain. DEVS (Discrete Event Systems specification), is an example of such a popular formalism in the simulation community that breaks-down the representation of a simulation model into hierarchical components.

Our objective is to focus on the operational support of execution for such simulation models. For example, considering that the model of a single node of a Peer-to-peer network requires several (and possibly many) DEVS components, it is easy to see that running simulations of a realistic large-scale peer-to-peer network rapidly ends-up involving millions of DEVS components. In addition to the problems posed by the execution of a distributed simulation application made of millions of components, such a use-case is also challenging in terms of analytics, because when millions of components are instrumented to collect observations, it becomes a typical instance of a big-data analytics problem.

4.1.2. Stochastic simulation platform

Understanding how complex objects, as found in finance/insurance (option contracts), biology (proteins structure), etc. evolve is often investigated by stochastic simulations (e.g. Monte-Carlo based). These can be very computational intensive and the associated communities are always seeking adequate parallel computing infrastructures and simulation software. Being able to harness all the available computing power, while ensuring the simulation is at first performant but also robust, capable to self-adapt, e.g. to failures, is a real opportunity for research and validation of our approach. Many other simulation applications could also benefit from our models and techniques, and we may in the future set up specific collaborations, e.g. in biocomputing, data-center activity management, or other engineering domains. We have recently solved pricing of high-performance demanding financial products on heterogeneous GPUs and multicore CPUs clusters, mixing use of active objects and OpenCL codes. This kind of application could continue to serve as a benchmark for our multi-level programming model.

4.2. Big data

4.2.1. Big data analytics

The amount of data digitally produced is increasing at an exponential rate. Having a dedicated programming model and runtime, such as Hadoop-MapReduce, has proved very useful to build efficient big data mining and analysis applications albeit for very static environments. However, if we consider that not only the environment is dynamic (node sharing, failures...) but so are the data (variation in popularity, arrival rate...), it becomes a much more complex problem. This domain is thus a very good candidate as an application field for our work.

More precisely, we plan to contribute at the deployment level, runtime level, and at the analytics programming model for the end-user level. We already worked on close topics with the distributed P2P storage and publish/subscribe system for Semantic Web data (named *EventCloud*). However, expressing a particular interest about data through simple or even more complex subscriptions (CEP) is only a first step in data analytics. Going further requires the full expressivity of a programming language to express how to mine into the real-time data streams, aggregate intermediate analytics results, combine with past data when relevant, etc. We intend to enlarge this effort about extracting meaningful information by also creating tighter collaborations with groups specialized in data mining algorithms (e.g. the Mind team at I3S).

We think that the approach advocated in SCALE is particularly adapted to the programming and support of analytics. Indeed, the mix of computational aspects and of large amount of data make the computation of analytics the perfect target for our programming paradigms. We aim at illustrating the effectiveness of our approach by experimenting on different computations of analytics, but we will put a particular focus on the case of data streams, where the analysis is made of chains (even cyclic graphs) of parallel and distributed operators. These operators can naturally be expressed as coarse grained composition of fine grained parallel entities, both granularity levels featuring autonomic adaptation. Also, the underlying execution platform that supports this execution also has to feature autonomic adaptation in order to deal with an unstable and heterogeneous execution environment. Here autonomic adaptation is also crucial because the programmer of analytics is not expected to be an expert in distributed systems.

Overall, this second application domain target should illustrate the effectiveness of our runtime platform and of our methodology for dynamic and autonomic adaptation.

SOCRATE Project-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 [49] 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 [50].

Recent natural disasters have brought considerable attention to the need of enhanced public safety communication abroad [48]. 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 direction 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.

SPIRALS Team

4. Application Domains

4.1. Introduction

Although our research is general enough to be applied to many application domains, we currently focus on applications and distributed services for the retail industry and for the digital home. These two application domains are supported by a strong expertise in mobile computing and in cloud computing that are the two main target environments on which our research prototypes are build, for which we are recognized, and for which we have already established strong collaborations with the industrial ecosystem.

4.2. Distributed software services for the retail industry

This application domain is developed in relation with the **PICOM** (*Pôle de compétitivité Industries du Commerce*) cluster. We have established strong collaborations with local companies in the context of former funded projects, such as **Macchiato**, which focused on the development of a new generation of mobile computing platforms for e-commerce. We are also involved in the Datalyse and OCCIware funded projects that define cloud computing environments with applications for the retail industry. Finally, our activities in terms of crowdsensing and data gathering on mobile devices with the APISENSE[®] platform share also applications for the retail industry.

4.3. Distributed software services for the digital home

We are developing new middleware solutions for the digital home, in particular through our long standing collaboration with Orange Labs. We are especially interested in developing energy management and saving solutions with the POWERAPI software library for distributed environments such the ones that equip digital homes. We are also working to bridge the gap between distributed services hosted on home gateways and distributed services hosted on the cloud to be able to smoothly transition between both environments. This work is especially conducted with the SALOON platform.

TACOMA Team

4. Application Domains

4.1. Pervasive applications in Smart Home

A smart home is a residence equipped with information-and-communication-technology (ICT) devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

The idea of using the Ubiquitous Computing paradigm in the smart home domain is not new. However, the state-of-the-art solutions only partially adhere to its principles. Often the adopted approach consists in a heavy deployment of sensor nodes, which continuously send a lot of data to a central elaboration unit, in charge of the difficult task of extrapolating meaningful information using complex techniques. This is a *logical approach*. TACOMA proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicate processing compared to the current solutions. The result is a smart home that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a domestic environment, showing a solution for improving both comfort and energy savings.

4.2. Metamorphic House

The motivation for metamorphic houses is that many countries, including France, are going through socio-demographic evolutions, like growth of life expectancy and consequent increase in the number of elderly people, urbanization and resource scarcity. Households experience financial restrictions, while housing costs increase with the raise of real estate and energy prices [5].

Important questions arise concerning the future of housing policies and ways of living. We observe novel initiatives like participative housing and developing behaviors, including house-sharing, teleworking and longer stay of children in parents' homes.

To tackle the challenges raised by these emerging phenomena, future homes will have to be modular, upgradeable, comfortable, sparing of resources. They should be integrated in the urban context and exchange information with other homes, contribute to reducing the distances to be covered daily and respect the characteristics of the territory where they are located.

To reach these goals, metamorphic domestic environments will modify their shape and behavior to support activities and changes in life cycle of occupants, increase comfort and optimize the use of resources. Thanks to Information and Communication Technologies (ICT) and adaptive building elements, the same physical spaces will be transformed for different uses, giving inhabitants the illusion of living in bigger, more adapted and more comfortable places.

4.3. Pervasive applications in uncontrolled environments

Some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations than those usually encountered or expected for classical RFID systems.

RFID technology is to avoid missing tags when reading multiple objects, as reading reliability is affected by various effects such as shadowing or wave power absorption by some materials. The usual applications of RFID operate in a controlled environment in order to reduce the risk of missing tags while scanning objects.

In pervasive computing applications, a controlled reading environment is extremely difficult to achieve, as one of the principle is to enhance existing processes "in situ", unlike the controlled conditions that can be found in industrial processes. Consider for example a logistic application, where RFID tags could be used on items inside a package in order to check for its integrity along the shipping process. Tags would likely be placed randomly on items inside the package, and reading conditions would be variable depending on where the package is checked.

RFID operation in uncontrolled environments is challenging because RFID performance is affected by multiple parameters, in particular:

- Objects materials (on which tags are attached to),
- Materials in the surrounding environment,
- RFID frequency spectrum,
- Antenna nature and placement with respect to the tags.

In controlled environment, the difficulty to read tags can be limited by using the appropriate parameters to maximize the RFID performance for the application. But in many cases, it is needed to read large number of objects of various nature, arranged randomly in a given area or container. Most pervasive computing applications fall in this context. At the software level, RFID inventory reliability issue is usually addressed by anti-collisions mechanisms and redundancy mechanisms. Anti-collisions protocols limit the risk of data corruption when multiples tags have to reply to an inventory request. Redundancy is often implemented in RFID readers by aggregating the results of multiple inventory requests over a time frame, to give the tags multiple opportunities to reply. While useful, these strategies cannot ensure that a given inventory is valid or not (in other words, one or more tags may be missing without being noticed).

TYREX Project-Team

4. Application Domains

4.1. Web Programming Technologies

Despite the major social and economic impacts of the web revolution, current web programming methods and content representation are lagging behind and remain severely limited and in many respects archaic. Dangerously, designing web applications even becomes increasingly complex as it relies more and more on a jungle of programming languages, tools and data formats, each targeted toward a different application layer (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 and security holes. In addition, the communication aspects are often handled independently via remote service invocations and represent another source of complexity and vulnerability. We believe that we reached a level where there is an urgent need and a growing demand for alternative programming frameworks that capture the essence of web applications: advanced content, data and communication. Therefore, successful candidate frameworks must capture rich document formats, data models and communication patterns. A crucial aspect is 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. For all these reasons, we believe that a new generation of tools must be created and developed in order to overcome the aforementioned limitations of current web technologies.

4.2. Multimedia and Augmented Environments

The term Augmented Environments refers collectively to ubiquitous computing, context-aware computing, and intelligent environments. The goal of our research on these environments is to introduce personal Augmented Reality (AR) devices, taking advantage of their embedded sensors. We believe that personal AR devices such as mobile phones or tablets will play a central role in augmented environments. These environments offer the possibility of using ubiquitous computation, communication, and sensing to enable the presentation of context-sensitive information and services to the user. AR applications often rely on 3D content and employ specialized hardware and computer vision techniques for both tracking and scene reconstruction and exploration. Our approach tries to seek a balance between these traditional AR contexts and what has come to be known as mobile AR browsing. It first acknowledges that mobile augmented environment browsing does not require that 3D content be the primary means of authoring. It provides instead a method for HTML5 and audio content to be authored, positioned in the surrounding environments and manipulated as freely as in modern web browsers. The applications we develop to guide and validate our concepts are pedestrian navigation techniques and applications for cultural heritage visits. Features found in augmented environments are demanding for the other activities in the team. They require all kinds of multimedia information, that they have to combine. This information has to be processed efficiently and safely, often in real time, and it also, for a significant part, has to be created by human users.

URBANET Team

4. Application Domains

4.1. Smart urban infrastructure

Unlike the communication infrastructure that went through a continuous development in the last decades, the distribution networks in our cities including water, gas and 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 became 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 necessities of both consumers and suppliers

Another fundamental urban infrastructure is the transportation system. The progress made in 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 can be achieved through 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 architecture 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 participatory 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. Human-centric networks

Combining location awareness and data recovered from multiple sources like social networks or sensing devices can surface previously unknown characteristics of the urban environment, and enable important new services. As a few examples, one could think of informing citizens about often disobeyed (and thus risky) traffic signs, polluted neighborhoods, or queue waiting times at current exhibitions in the urban area.

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.

WHISPER Team

4. Application Domains

4.1. Linux

Linux is an open-source operating system that is used in settings ranging from embedded systems to supercomputers. The most recent release of the Linux kernel, v3.17, comprises over 12 million lines of code, and supports 29 different families of CPU architectures, 73 file systems, and thousands of device drivers. Linux is also in a rapid stage of development, with new versions being released roughly every 2.5 months. Recent versions have each incorporated around 13,500 commits, from around 1500 developers. These developers have a wide range of expertise, with some providing hundreds of patches per release, while others have contributed only one. Overall, the Linux kernel is critical software, but software in which the quality of the developed source code is highly variable. These features, combined with the fact that the Linux community is open to contributions and to the use of tools, make the Linux kernel an attractive target for software researchers. Tools that result from research can be directly integrated into the development of real software, where it can have a high, visible impact.

Starting from the work of Engler et al. [42], numerous research tools have been applied to the Linux kernel, typically for finding bugs [41], [56], [65], [72] or for computing software metrics [46], [74]. In our work, we have studied generic C bugs in Linux code, bugs in function protocol usage [50], [51], issues related to the processing of bug reports [20] and crash dumps [18], and the problem of backporting (work in progress), illustrating the variety of issues that can be explored on this code base. Unique among research groups working in this area, we have furthermore developed numerous contacts in the Linux developer community. These contacts provide insights into the problems actually faced by developers and serve as a means of validating the practical relevance of our work. Section 5.1.2 presents our dissemination efforts to the Linux community.

4.2. Device Drivers

Device drivers are essential to modern computing, to provide applications with access, via the operating system, to physical devices such as keyboards, disks, networks, and cameras. Development of new computing paradigms, such as the internet of things, is hampered because device driver development is challenging and error-prone, requiring a high level of expertise in both the targeted OS and the specific device. Furthermore, implementing just one driver is often not sufficient; today's computing landscape is characterized by a number of OSes, *e.g.*, Linux, Windows, MacOS, BSD and many real time OSes, and each is found in a wide range of variants and versions. All of these factors make the development, porting, backporting, and maintenance of device drivers a critical problem for device manufacturers, industry that requires specific devices, and even for ordinary users.

The last fifteen years have seen a number of approaches directed towards easing device driver development. Réveillère, who was supervised by G. Muller, proposes Devil [7], a domain-specific language for describing the low-level interface of a device. Chipounov *et al.* propose RevNic, [36] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [66], [67] an approach for synthesizing device driver code from a specification of an OS and a device. Currently, these approaches have been successfully applied to only a small number of toy drivers. Indeed, Kadav and Swift [47] observe that these approaches make assumptions that are not satisfied by many drivers; for example, the assumption that a driver involves little computation other than the direct interaction between the OS and the device. At the same time, a number of tools have been developed for finding bugs in driver code. These tools include SDV, [27] Coverity [42], CP-Miner, [55] PR-Miner [56], and Coccinelle [8]. These approaches, however, focus on analyzing existing code, and do not provide guidelines on structuring drivers.

In summary, there is still a need for a methodology that first helps the developer understand the software architecture of drivers for commonly used operating systems, and then provides guidelines and tools for the maintenance and the development of new drivers. Section 3.2 describes this research direction.

ALICE Project-Team

4. Application Domains

4.1. Numerical simulation

flow simulation for oil exploration: we co-advised three Ph.D. theses with the Gocad Consortium, that develops modeling algorithms for oil and gas exploration. We developed specialized meshing algorithms, well suited to represent geological layers at various resolutions [27], [19].

optimal transport: this is an active research topics in the mathematics community. Given two measures μ and ν , optimal transport defines a distance between μ and ν , as the minimum cost of “morphing” μ into ν . This distance (called the *Wasserstein distance*) structures the space of measures and offers new ways of solving some highly non-linear PDEs (Monge-Ampere, Fokker-Plank ...). This requires a numerical way of computing the Wasserstein distance and its gradients. We studied a semi-discrete technique [21] (conditionally accepted to ESAIM J. M2AN) that optimizes power diagrams. This is to our knowledge the first numerical implementation of optimal transport for volumetric densities (computes the Wasserstein distance between a sum of Dirac masses and a piece-wise linear density supported on a tetrahedral mesh).

Bose-Einstein condensates: Xavier Antoine (prof. in mathematics at the Université de Lorraine) joined the team on a “delegation” position (Sept. 2013 - Sept. 2014) to explore some common research topics. We are members of the BECASIM project, funded by the ANR (“French NSF”). In a certain sense, a Bose-Einstein condensate is a “Schroedinger cat” made of a few hundred atoms. By special physical means (low temperature and lasers), the probability waves of these atoms are intermixed, thus forming an alternative state of matter. The BECASIM project aims at developing numerical simulation methods for these complicated phenomena (that intermix fluid dynamics, electromagnetics and quantum physics).

4.2. Fabrication

Our work around fabrication and additive manufacturing finds applications in different fields. Our algorithms for fast geometric computations on solids (boolean operations, morphological operations) are useful to model a variety of shapes, from mechanical engineering parts to prosthetics for medical applications. Our techniques allow for simpler modelling and processing of very intricate geometries and therefore also find applications in art and design, for unusual shapes that would be very difficult to obtain otherwise.

ALPAGE Project-Team

4. Application Domains

4.1. Overview

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 four 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; related to this group lies the automatic correction of the output of OCR systems;

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 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 Villemonde de La Clergerie, 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. 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 wires 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 was a CIFRE PhD student at Alpage and at AFP, and worked in 2013 within the ANR project EDyLex).

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 5.5) and Alexina morphological lexicons. Several other parts of *vera* have been co-developed by Verbatim Analysis and by Inria.

4.4. Multilingual terminologies and lexical resources for companies

Participant: Éric Villemonte de La Clergerie.

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 Mikaël Morardo in 2012 as an engineer, funded by Inria's DTI. He pursued his work 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, including the WOLF.

4.5. Automatic and semi-automatic spelling correction in an industrial setting

Participants: Kata Gábor, Pierre Magistry, Benoît Sagot, Éric Villemonte de La Clergerie.

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. Alpage is working within the "Investissements d'avenir" project PACTE, headed by Numen, a company specialized in text digitalization, and three other partners. Kata Gábor and Pierre Magistry are doing post-docs funded by PACTE (see 6.3)

4.6. Experimental and quantitative linguistics

Participants: Benoît Crabbé, Benoît Sagot, Alexandra Simonenko, Sarah Beniamine, Kristina Gulordava.

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 [62], [139], the existence of graded grammaticality judgments [86].

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 [72] gave way on the one hand to more abstract characterizations of the language faculty [72], and on the other hand to the construction of detailed, formally explicit, and often implemented, alternative formulation of the generative approach [61], [103]. For French several grammars have been implemented in this trend, such as the tree adjoining grammars of [65], [76] 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 [121]. 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. In the perspective of better understanding which factors influence the relative ordering of post verbal complements across languages and through language evolution.

On the other hand we are also collaborating 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.

In parallel, quantitative methods are applied to computational morphology, in particular in the context of Sarah Beniamine's PhD, co-supervised by Benoît Sagot (Alpage) and Olivier Bonami (LLF, CNRS, U. Paris Diderot and U. Paris Sorbonne). Collaborative work in this area is also conducted in collaboration with descriptive linguists from CRLAO (CNRS and Inalco; Guillaume Jacques) and HTL (CNRS, U. Paris Diderot and U. Sorbonne Nouvelle; Aimée Lahaussais) and formal linguists from DDL (CNRS and Université Lyon 2; Géraldine Walther).

AVIZ Project-Team

4. Application Domains

4.1. Panorama

Research in visual analytics can profit from the challenges and requirements of real-world datasets. Aviz develops active collaboration with users from a range of 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.

We apply our techniques to important medical applications domains such as bioinformatics and brain studies. In particular, we are interested in helping neuroscientists make sense of evolving functional networks, in the form of weighted and/or dynamic graphs.

Other application domains include:

- *Digital Humanities* in general, and in particular with the Cendari European project with historians from most European countries, and the joint project with Microsoft Research and Inria on Graph Visualization;
- *Genealogy*, in cooperation with North Carolina State University;
- *Digital Libraries*, in cooperation with the French National Archives and the Wikipedia community.

AYIN Team

4. Application Domains

4.1. Remote sensing

With the development and launch of new instruments (for instance, GeoEye, Ikonos, Pleiades, COSMO-SkyMed, TerraSAR-X, and future missions EnMAP, PRISMA, HYPXIM, ...) capturing Earth images at very high spatial, spectral, and temporal resolutions, numerous new applications arise, such as precision agriculture, natural disaster management, monitoring of urban environments, and mineralogy. We apply our new methodologies to the analysis of SAR, multi- and hyper-spectral remote sensing images and temporal sequences. In particular, we address image segmentation and classification, change detection, the extraction of structures, and object tracking.

4.2. Skin care

The most recent sensors used in dermatology and cosmetology produce images with very high spatial, spectral, and temporal resolutions. As with remote sensing, numerous applications then arise that can make use of the new information. In the application to dermatology, we are particularly interested in hyperpigmentation detection and the evaluation of the severity of various disorders (for instance, for melasma, vitiligo, acne, melanoma, etc.). In the application to cosmetology, our main goals are the analysis, modeling, and characterization of the condition of human skin, especially as applied to the evaluation of methods designed to influence that condition.

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 project-team research applications have been oriented towards surveillance, monitoring and decision support. Our domains of application are:

- Agriculture and environment
- Health
- Exploitation of execution traces in an industrial setting

4.2. Environmental decision making

environment, decision methods

The need for decision support systems in the environmental domain is now well-recognized. It is especially true in the domain of water quality. 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. This work began in ANR projects like APPEAU and ACASSYA or the PSDR GO CLIMASTER project (Changement climatique, systèmes agricoles, ressources naturelles et développement territorial).

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 farmers' decisions. One of our main contributions 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 "unsatisfactory" situation.

Our main partners are the SAS INRA research group, located in Rennes and the BIA INRA and AGIR INRA research groups in Toulouse.

Ecosystem Management.

The objective of ecosystem management is to ensure sustainable ecosystems even when submitted to various stressors such as natural disturbances or human pressures. Several studies have already shown the interest of qualitative modelling for ecosystems [38]. In our case, we propose to couple a qualitative modelling with model-checking tools to explore marine ecosystems (as explained section 3.2). We applied our approach on a small-scale subsistence fishery in a coral reef lagoon (Uvea, New Caledonia). A well described foodweb model provides us with useful input data for steady-state biomass data and estimates of production and consumption. A timed automata model was developed using EcoMata to investigate the direct and indirect effects of various fishing strategies on a subset of the trophic network.

This work has been realized in collaboration with ecologists: Yves-Marie Bozec (today in position in Marine Spatial Ecology, University of Queensland, Australia) and Guy Fontenelle (Professeur at Agrocampus Ouest).

A second application has been studied in the dairy management area. Over an hybrid modelling on the grazing activities, four methods to generate the best grazing management activity has been proposed. The expert partners are researchers from the SAS INRA research group, located in Rennes.

4.3. Health

health-care, patient monitoring, medicament usage, pharmaco-immunology, health-care pathways, wireless sensors

Clinical monitoring, electronic patient records and computer supported disease management produce more and larger volumes of clinical data. This data is a strategic resource for healthcare institutions. Data mining brings the facility to discover patterns and correlation hidden within the data repository and assists professionals to uncover these patterns and to exploit them to improve medical care.

We are working on two aspects of health-care:

- exploitation of data from the french care insurance (Assurance Maladie) that contains records of medicament reimbursements for pharmaco-immunology purposes. Our goal is to reconstruct and mine patients' healthcare pathways in order to detect regularities and anomalies in the way patients take medicaments and alert medical authorities in case some problem is detected, such as non expected negative consequences of medicament intake. We are working in the framework of a project funded by the National Medicament Security Agency (ANSM - Agence Nationale de la Sécurité du Médicament) for building a platform enabling focused studies on specific medicaments as well as discovering potential problems with medicament usage. This means selecting from billions of patients records, patients sharing similar medical contexts and showing different consequences of medicament intake,
- veterinary monitoring of feedlot cattle in big farms from sensors recording behavioral and physiological data. As farms are becoming bigger and bigger, detecting ill animals by visual appraisal is becoming more and more difficult. With the advent of cheap wireless sensors, animals (i.e. cows or steers) may be monitored in quasi real time for detecting relevant changes in their behavior that could be related to specific diseases. We are exploring diverse methods for detecting changes on multivariate data, such as cusum charts, specific sequential patterns or distribution of frequent patterns. We are specifically working with veterinaries from the university of Calgary (Canada) for monitoring feedlot cattle in farms growing up to 50.000 animals.

4.4. Exploitation of execution traces

log analysis, data mining, embedded systems.

We have an ongoing collaborations with STMicroelectronics, which is one of the world top-5 electronic chip makers. Nowadays, set-top boxes, smartphones or onboard car computers are powered by highly integrated chips called *System-on-Chip (SoC)*. Such chips contain on a single die processing units, memories, IO units and specialized accelerators (such as audio and video encoding/decoding). Programming SoC is a hard task due to their inherent parallelism, leading to subtle bugs when several components do not deliver their results within a given time frame. Existing debuggers and profilers are ill-adapted in this case because of their high intrusivity that modifies the timings. Hence the most used technique is to capture a *trace* of the execution and analyze it post-mortem. While Alexandre Termier was in Grenoble he initiated several works for analyzing such traces with data mining techniques [55], [50], which he is now pursuing with his colleagues of the Dream project-team [14].

4.5. Software components monitoring

software components, web services, distributed diagnosis

Web-services 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. 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, becomes then a crucial concern.

We have been working on this problem within the WS-DIAMOND project [61], a large European funded project involving eight partners in Italy, France, Austria and Netherlands <http://wsdiamond.di.unito.it/>.

We do not work anymore on the diagnosis of web services, now we aim 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 processing when faults occur and propagate through the orchestration.

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*⁰ 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 industrials 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*.

⁰Advanced Driver Assistance Systems

EXMO Project-Team

4. Application Domains

4.1. Semantic web technologies

The main application context motivating our work is the “semantic web” infrastructure [18].

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 enables 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” [14] complements the web, with formal knowledge representation spanning across sites. Taking advantage of this semantic web requires 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 more specific uses of semantic web technologies in wider context (typically in the smart city context, §7.2.1.1).

FLOWERS Project-Team

4. Application Domains

4.1. Applications

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. Semantic Metadata

Semantic metadata are at the core of the applications we have been working on for several years. These three last years, we have switched from semantic annotations of documents to interlinking problems between individual references in annotations of documents. The main linkage problem in our current ANR project Qualinca (see Section 8.1) consists of identifying an authority (*i.e.*, an element of a referential described by metadata) in a bibliographic notice (*i.e.*, metadata describing a document). This problem is an instance of the intensively studied reference resolution problem. In the Semantic Web, it can be recast as the computation of `OWL:sameAs` links between two metadata bases, clearly a fundamental problem for the Linked Open Data. We use a knowledge-based approach to solve this problem, and this year we have especially studied key notions for building rules that conclude on coreference or difference links between entities.

4.2. Agronomy

Within this field, we investigate two different agronomy scenarios: (1) in the context of a public health controversy about bread making, choosing between different kinds of flour in function of nutritional, economic, health and other criteria and (2) designing ecoefficient and biodegradable packaging. The second scenario is part of a larger decision support system implemented within the EU FP7 project EcoBioCap (see Section 8.2).

Both scenarios rely upon different criteria which bring conflicting information for decision making. The aim is then twofold. First to properly model the knowledge using facts, rules and negative constraints. Then, in a second step, in the possibly inconsistent knowledge base thus obtained, to select maximally consistent subsets that will be used for decision making. We have chosen to use argumentation in this context (of reasoning in the presence of inconsistency) due to the fact that we aim to investigate, in the future, the explanation power of argumentation approaches (very useful in this context where the domain experts are not computer scientists).

HEPHAISTOS 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), 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 *modeling*, *optimal design* and *analysis* 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. Although these topics were new for us when initiating the project 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. 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 is a long term goal.

HYBRID Project-Team

4. Application Domains

4.1. Overview

The research program of Hybrid team aims at next generations of virtual reality and 3D user interfaces which could possibly address both the “body” and “mind” of the user. Novel interaction schemes are designed, for one or multiple users. We target better integrated systems and more compelling user experiences.

The applications of our research program correspond to the applications of virtual reality technologies which could benefit from the addition of novel body-based or mind-based interaction capabilities:

- **Industry:** with training systems, virtual prototyping, or scientific visualization;
- **Medicine:** with rehabilitation and reeducation systems, or surgical training simulators;
- **Entertainment:** with 3D web navigations, video games, or attractions in theme parks,
- **Construction:** with virtual mock-ups design and review, or historical/architectural visits.

IMAGINE Project-Team

4. Application Domains

4.1. Domain

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 modeling within the reach of the general public, enabling educators, media and other practitioners to author their own 3D content.

Our current application domains are:

- Visual arts
 - Modeling and animation for 3D films and games.
 - Virtual cinematography and tools for theatre directors.
- Engineering
 - Industrial design.
 - Mechanical & civil engineering.
- Natural Sciences
 - Virtual functional anatomy.
 - Virtual plants.
- Education and Creative tools
 - Sketch-based teaching.
 - Creative environments for novice users.

The diversity of users these domains bring, from digital experts to other professionals and novices, gives us excellent opportunities to validate our general methodology with different categories of users. Our ongoing projects in these various application domains are listed in Section 6.

IN-SITU Project-Team

4. Application Domains

4.1. Application 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. Application Domains

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 motion, 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 images⁰, 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).

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.

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.

⁰<http://www.sims.berkeley.edu/research/projects/how-much-info/summary.html>

LINKMEDIA Project-Team

4. Application Domains

4.1. Asset management in the entertainment business

Regardless of the ingestion and storage issues, media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l'Audiovisuel, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals.

4.3. Multiscreen TV

The convergence between television and the Internet has accelerated significantly over the past few years, with the democratization of TV on-demand and replay services and the emergence of social TV services and multiscreen applications. These evolutions and the ever growing number of innovative applications incurred offer a unique playground for multimedia technologies. Recommendation plays a major role in connected TV. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases, appears as a key feature in this context, at the core of rich TV and second screen applications.

4.4. E-learning

On-line courses are rapidly gaining interest with the recent movement for massive open on-line courses (MOOCs). Such courses usually aggregate multimedia material, such as a video of the course with handouts and potentially text books, exercises and other related resources. This setting is very similar to that of the media aggregation sites though in a different domain. Automatically analyzing and describing video and textual content, synchronizing all material available across modalities, creating and characterizing links between related material or between different courses are all necessary features for on-line courses authoring.

LINKS Team

3. Application Domains

3.1. Collective Intelligence

⁰ Links represented in the data are important for web users, who try to locate relevant information. They typically want to pose their queries locally and obtain the answers from both local and remote repositories. With the concept of linked data collections, the users are provided with a virtual collection of data and links. The answers to a query need to follow both explicit and implicit links to external repositories. Nevertheless, we argue that the benefits of links are not limited to casual users. In this paragraph, we briefly discuss two applications in which linked data collections need to be created. In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today's enterprises need to face the problem of information explosion, due to the Internet's instability to rapidly convey large amounts of information throughout the world via end-user applications and tools. A linked data collection thus represents a virtual knowledge repository, in which relevant data is collected and meaningful mappings between this data and external world are inferred. A linked data collection would ease the task of expert users to (i) process data, metadata and knowledge that derives from machine to machine processing and eventually make sense of it; (ii) assemble and disassemble pieces of data, metadata and knowledge to create aggregate opinions or to disaggregate opinions in a way that turns to be useful in decision making; (iii) continuously learn from user feedback to produce better knowledge and to enhance the semantics of data processing.

3.2. Linked Bibliographic Collections

The second example concerns scientists who want to quickly inspect relevant literature and datasets. In such a case, local knowledge that comes from a local repository of publications belonging to a research institute (e.g. HAL) need to be integrated with other Web-based repositories, such as DBLP, Google Scholar, ResearchGate and even Wikipedia. Indeed, the local repository may be incomplete or contain semantic ambiguities, such as mistaken or missing conference venues, mistaken long names for the publication venues and journals, missing explanation of research keywords, and opaque keywords. A linked data collection would lead to build both implicit and explicit links by (i) cleaning the errors with links to correct data e.g. via mappings from HAL to DBLP for the publications errors, and via mappings from HAL to Wikipedia for opaque keywords, (ii) thoroughly complete the list of publications of the research institute, and (iii) support complex queries on the corrected data combined with external links. Links are thus useful in all scenarios, in which massive data need to be understood, analyzed and processed. Finally, they can be processed in highly distributed scenarios, such as in the cloud, where their evaluation can be run on many different sites.

⁰Collective intelligence is a shared or group intelligence that emerges from the collaboration and competition of many individuals and appears in consensus decision making" (from Wikipedia - Collective Intelligence)

MAGNET Team

4. Application Domains

4.1. Overview

Our main targeted applications are browsing, monitoring and mining in information networks. Such discovered structures would also be beneficial to predicting links between users and texts which is at the core of recommender systems. All the learning tasks considered in the project such as node clustering, node and link classification and link prediction are likely to yield important improvements in these applications. Application domains cover social networks for cultural data and e-commerce, and biomedical informatics.

MAGRIT Project-Team

4. Application Domains

4.1. Augmented reality

We have a significant experience in AR that allowed good progress in building usable, reliable and robust AR systems. Our contributions cover the entire process of AR: matching, pose initialization, 3D tracking, in-situ modeling, handling interaction between real and virtual objects....

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. Contributions of the team focus about the developments of AR tools for neuro-navigation as well as the development of simulation tools of the interventional act for training or planning. Laparoscopic surgery is another field of interest with the development of methods for tracking deformable organs based on bio-mechanical models. Some of these projects are developed in collaboration with the EPC SHACRA.

4.3. Applied mechanics

In experimental solid mechanics, an important problem is to characterize properties of specimen subject to mechanical constraints, which makes it necessary to measure tiny strains. Contactless measurement techniques have emerged in the last few years and are spreading quickly. They are mainly based on images of the surface of the specimen on which a regular grid or a random speckle has been deposited. We are engaged since June 2012 in a transdisciplinary collaboration with Institut Pascal (Clermont-Ferrand Université). The aim is to characterize the metrological performances of these techniques limited by, e.g., the sensor noise, and to improve them by several dedicated image processing tools.

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 Personally Assisted 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);
- Active Sensing, where decisions have to be taken in order to gather information on a system. This can be applied to many fields, like for example monitoring the integrity of airplanes wings or the behavior of people in public areas.

4.2. Ambient intelligence

As the Nancy – Grand Est Research Center scientific strategy pushes the development of plateforms on Robotics and Smart Living Apartments, some members of the team have recentered their research toward “ambient intelligence and AI” . This choice is backed up by the Inria 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 coordinator is a member of MAIA Team. Within this new domain of research in MAIA, we explore how intelligent decentralized complex systems can help designing 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.

MANAO Project-Team (section vide)

MAVERICK Project-Team (section vide)

MIMETIC Project-Team

4. Application Domains

4.1. Autonomous Characters

Autonomous characters are becoming more and more popular as 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 kinds 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.

4.2. 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 has 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 practitioners 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 example, for the collision avoidance.

4.3. 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.

4.4. 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.

4.5. 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.

4.6. 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 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 literature.

4.7. VR and Ergonomics

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VE is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms ask for several improvements in the field of motion analysis and VR: the interactions have to be as fidelistic as possible to properly mimic the motions performed in real environments, the fidelity of the simulator need also to be correctly evaluated, and motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

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 [49]. 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.

MORPHEO Project-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-Flight 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 analyzes 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 cues 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 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.

MULTISPEECH Team

4. Application Domains

4.1. Introduction

Approaches and models developed in the MULTISPEECH project are intended to be used for facilitating oral-based communication in various situations through enhancements of the communication channels, either directly via automatic speech recognition or speech production technologies, or indirectly, thanks to computer assisted language learning. Applications also include the usage of speech technologies for helping people in handicapped situations or for improving their autonomy. Foreseen application domains are related to computer assisted learning, health and autonomy (more precisely aided communication and monitoring), annotation and processing of spoken documents, and multimodal computer interaction.

4.2. Computer assisted learning

Although speaking seems quite natural, learning foreign languages, or learning the mother tongue for people with language deficiencies, represent critical cognitive stages. Hence, many scientific activities have been devoted to these issues either from a production or a perception point of view.

The general guiding principle with respect to computer assisted mother or foreign language learning is to combine modalities or to augment speech to make learning easier. Also, the system should provide indications on what should be corrected, a guidance which is considered as necessary by specialists in the oral aspects of language learning. Consequently, based upon a comparison of the learner's production to a reference, automatic diagnoses of the learner's production can be considered, as well as perceptual feedback relying on an automatic transformation of the learner's voice. For example, with respect to prosody, the diagnosis provided through both a text and a visual display, comes from an evaluation of the melodic curve and of the phoneme durations of the learner's realization; and the perceptual feedback consists in a replacement of the learner's prosodic cues by those of the reference; i.e., the signal of the learner's utterance is modified in order to reflect the prosodic cues (duration and F0) of the reference in order to make the learner aware of the expected prosodic cues. The diagnosis step strongly relies on the studies on categorization of sounds and prosody in the mother tongue and in the second language, and also depends on the influence between them. Furthermore, reliable diagnosis on individual utterances is still a challenge, and elaboration of advanced automatic feedback requires a temporally accurate segmentation of speech utterances into phones and this explains why accurate segmentation of native and non-native speech is also an important topic in the field of acoustic speech modeling.

4.3. Aided communication and monitoring

Speech technologies provide ways of helping people in handicapped situations or improving their autonomy. The following applications are considered in the project.

The first one is related to the tuning of speech recognition technology for providing a means of communication between a speaking person and a hard-of-hearing or a deaf person, through an adequate display of the recognized words and/or syllables, which takes also into account the reliability of the recognized items.

The second application aims at improving pathological voices. In this context, the goal is typically to transform the pathological voice signal in order to make it more intelligible. Ongoing work deals with esophageal voices, i.e., substituted voice learned by a laryngectomized patient who has lost his/her vocal cords after surgery. Voice conversion techniques will be studied further to enhance such voice signals, in order to produce clean and intelligible speech signals in replacement of the pathological voice.

The third application aims at improving the autonomy of elderly or disabled people, and fit with smartrooms. In a first step, source separation techniques could be tuned and should help for locating and monitoring people through the detection of sound events inside apartments. In a longer perspective, adapting speech recognition technologies to the voice of elder people should also be useful for such applications, but this requires the recording of adequate databases. Sound monitoring in other application fields (security, environmental monitoring) could also be envisaged.

4.4. Annotation and processing of spoken documents

The first type of annotation consists in transcribing a spoken document in order to get the corresponding sequences of words, with possibly some complementary information, such as the structure (punctuation) or the modality (affirmation/question) of the utterances to make the reading and understanding easier. Typical applications of the automatic transcription of radio or TV shows, or of any other spoken document, include making possible their access by deaf people, as well as by text-based indexing tools.

The second type of annotation is related to speech-text alignment, which aims at determining the starting and ending times of the words, and possibly of the sounds (phonemes). This is of interest in several cases as for example, for annotating speech corpora for linguistic studies, and for synchronizing lip movements with speech sounds, for example for avatar-based communications. Although good results are currently achieved on clean data, automatic speech-text alignment needs to be improved for properly processing noisy spontaneous speech data and needs to be extended to handle overlapping speech.

Finally, there is also a need for speech signal processing techniques in the field of multimedia content creation and rendering. Relevant techniques include speech and music separation, speech equalization, prosody modification, and speaker conversion.

4.5. Multimodal computer interactions

Speech synthesis has tremendous application in facilitating communication in a human-machine interaction context to make machines more accessible. For example, it started to be widely common to use acoustic speech synthesis in smartphones to make possible the uttering of all the information. This is valuable in particular in the case of handicap, as for blind people. Audiovisual speech synthesis, when used in an application such as a talking head, i.e., virtual 3D animated face synchronized with acoustic speech, is beneficial in particular for hard-of-hearing individuals. This requires an audiovisual synthesis that is intelligible, both acoustically and visually. A talking head could be an intermediate between two persons communicating remotely when their video information is not available, and can also be used in language learning applications as vocabulary tutoring or pronunciation training tool. Expressive acoustic synthesis is of interest for the reading of story, such as audiobook, to facilitate the access to literature (for instance for blind people or illiterate people).

OAK Project-Team

4. Application Domains

4.1. Social Networks

We develop models and algorithms for efficiently exploiting, enhancing, and querying social network data, in particular based on structured content, semantic annotations, and user interaction networks. We pursue this research with many industrial partners within the ALICIA project (Section 7.2.1) as well as in the Structured, Social, and Semantic Search project (Section 7.2.2).

4.2. Computational Journalism

Modern journalism increasingly relies on content management technologies in order to represent, store, and query source data and media objects themselves. Writing news articles increasingly requires consulting several sources, interpreting their findings in context, and crossing links between related sources of information. OAKresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. We have launched a collaboration with Le Monde's "Les Décodateurs" team to investigate these topics.

4.3. Open Data Intelligence

The Web is a vast source of information, to which more is added every day either in unstructured form (Web pages) or, increasingly, as partially structured sources of information, in particular as Open Data sets, which can be seen as connected graphs of data, most frequently described in the RDF data format recommended by the W3C. Further, RDF data is also the most appropriate format for representing structured information extracted automatically from Web pages, such as the DBPedia database extracted from Wikipedia or Google's InfoBoxes. To intelligently exploit such Open Data collections, OAKhas developed a complete framework for RDF data analytics within the recently completed DW4RDF project and continues work on this topic within the ODIN project started this year.

4.4. Hybrid Data Warehousing

Increasingly many modern applications need to exploit data from a variety of formats, including relations, text, trees, graphs etc. The recent development of data management systems aimed at "Big Data", including NoSQL platforms, large-scale distributed systems etc. provides enterprise architects with many systems to chose from. This makes it hard to decide which part of the application data to handle in which system, especially given that each system is best at handling a specific kind of data and a certain class of operations. OAKinvestigates principled techniques for distributing an application's data sources across a variety of systems and data models, based on materialized views. We test our ideas in this area within the Datalyse project.

ORPAILLEUR Project-Team

4. Application Domains

4.1. Biology and Chemistry

Participants: Mehwish Alam, Aleksey Buzmakov, Adrien Coulet, Marie-Dominique Devignes, Elias Egho, Nicolas Jay, Bernard Maigret, Amedeo Napoli, Nicolas Pépin-Hermann, Gabin Personeni, David Ritchie, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

Keywords: knowledge discovery in life sciences, bioinformatics, biology, chemistry, genomics

One major application domain which is currently investigated by the 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 the developed solutions bring new research ideas or possibilities 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. Indeed, 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.

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 depending on domain knowledge.

On the same line as biological data, chemical data are presenting important challenges w.r.t. knowledge discovery, for example for mining collections of molecular structures and collections of chemical reactions in organic chemistry. The mining of such collections is an important task for various reasons among which the challenge of graph mining and the industrial needs (especially in drug design, pharmacology and toxicology). Molecules and chemical reactions are complex data that can be modeled as undirected labeled graphs. One objective for guiding computer-based synthesis in organic chemistry is to discover general synthesis methods (i.e. kinds of “meta-reactions”) from currently available chemical reaction databases for designing generic and reusable synthesis plans.

Graph mining methods may play an important role in this framework as illustrated in [125], but Formal Concept Analysis (FCA) can also be used in an efficient and well-founded way [101]. Combining supervised methods –with a training sets where objects are tagged– and unsupervised methods, “jumping emerging patterns” can be detected that characterize classes of interest, e.g. toxic molecules or inhibitors. Then, a hybrid classification method based on FCA can be used for building a concept lattice where some of the concepts can be used as reference classes for classifying unknown objects, for recognition and prediction tasks. Graph mining in the framework of FCA is a very important task on which we are actively working, whose results can be transferred to text mining as well.

4.2. Medicine

Participants: Aleksey Buzmakov, Adrien Coulet, Elias Egho, Nicolas Jay, Jean Lieber, Amedeo Napoli, Matthieu Osmuk, Chedy Raïssi, Yannick Toussaint, Mickaël Zehren.

Keywords: knowledge representation, description logics, classification-based reasoning, case-based reasoning, semantic web, formal concept analysis, sequence mining, text mining

We are working on several applications in medicine, mainly in knowledge management and analysis of patient trajectories as sequences. In the first case, the Kasimir research project is about decision support and knowledge management for the treatment of cancer. This is a multidisciplinary research project in which participate researchers in computer science (Orpailleur), experts in oncology (“Institut de Cancérologie de Lorraine 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, which is applied in 70% of the cases and provides a treatment. The 30% remaining cases are “out of the protocol”, e.g. contraindication, treatment impossibility, etc. and the protocol should be adapted, based on discussions among specialists. This adaptation process is modeled in Kasimir thanks to CBR, where the semantic Web technologies are used and adapted in the Kasimir project for several years.

Another work is in concern with the analysis of patient trajectories, i.e. the “path” of a patient during illness (chronic illnesses and cancer), considered as sequences. It is important to understand these sequence data and temporal data mining methods are good candidate tools for that. However, these methods should be adapted for addressing the complex nature of medical events. Thus, there is an ongoing work on the analysis of trajectories with different levels of granularity and w.r.t. external domain ontologies. In addition, it is also important to be able to compare and classify trajectories according to their content. This is why there is also a work on the definition of a similarity measure able to take into account the complex nature of trajectories and that can be efficiently implemented for allowing quick and reliable classifications.

4.3. Cooking

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: cooking, knowledge representation, knowledge discovery, case-based reasoning, semantic wiki

The origin of the Taaable project is the Computer Cooking Contest (CCC). A contestant to CCC is a system that answers queries about 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 based on various technologies from semantic web, knowledge discovery, knowledge representation and reasoning. From a research viewpoint the system enables to test scientific results and to study the complementarity of various research trends in an application domain which is simple to understand and which raises complex issues at the same time. Taaable has been at the origin of the ANR CONTINT project Kolflow, whose application domain is WikiTaaable, the semantic wiki of Taaable.

4.4. Agronomy

Participants: Sébastien Da Silva, Florence Le Ber [contact person], Jean-François Mari.

Keywords: simulation, Markov model, Formal Concept Analysis, graph

In September, Sébastien da Silva has defended his PhD thesis [13]. His research was conducted in the framework of an Inria-INRA collaboration, which takes place in the INRA research network PAYOTE about landscape modeling. The thesis, supervised both by Claire Lavigne (DR in ecology, INRA Avignon) and Florence Le Ber, was concerned with the characterization and the simulation of hedgerows structures in agricultural landscapes, based on Hilbert-Peano curves and Markov models [6] [13], [66], [98].

An on-going research work about the representation of peasant knowledge is involved within a collaboration with IRD in Madagascar [40]. Sketches drawn by peasants were transformed into graphs and compared thanks to Formal Concept Analysis.

PANAMA Project-Team

4. Application Domains

4.1. Acoustic scene capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

4.2. Audio signal separation in reverberant environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

4.3. Multimedia indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

4.4. Brain source imaging

Epilepsies constitute a common neurological disorder that affects about 1% of the world population. As the epileptic seizure is a dynamic phenomenon, imaging techniques showing static images of the brain (MRI, PET scan) are frequently not the best tools to identify the brain area of interest. Electroencephalography (EEG) is the technique most indicated to capture transient events directly related to the underlying epileptic pathology (like interictal spikes, in particular). EEG convey essential information regarding brain (patho-)physiological activity. In addition, recording techniques of surface signals have the major advantage of being noninvasive. For this reason, an increased use in the context of epilepsy surgery is most wanted. However, to

reach this objective, we have to solve an electromagnetic inverse problem, that is to say to estimate the current generators underlying noisy EEG data. Theoretically, a specific electromagnetic field pattern may be generated by an infinite number of current distributions. The considered inverse problem, called "brain source imaging problem", is then said to be ill-posed.

PERCEPTION Project-Team (section vide)

POTIOC Project-Team

4. Application Domains

4.1. Popularization of science, education, art, entertainment

Our project aims at providing 3D digital worlds to all, including the general public, to stimulate understanding, learning, communication and creation. Our scope of applications encompasses

- popularization of science
- education
- art
- entertainment

See "Objective 3: Exploring new applications and usages" for the detailed description [3.4](#).

PRIMA Project-Team (section vide)

REVES Project-Team

4. Application Domains

4.1. Domain

The application domain is vast. It ranges from audiovisual production, which typically requires long, offline computation to obtain high quality results, all the way to real-time applications such as computer games or virtual reality, for which the main consideration is to guarantee 60 frames per second frame rates, or, in general the reduction of latency to user reaction. The process of generation of images and sound, generally called *rendering* is our primary interest; our second main interest are virtual environments (VE's) as well as augmented (AE's) or mixed environments (ME's), that is scenes containing both real objects (often digitized) as well as purely synthetic objects. We are interested in both the generation and the interaction with these environments. We use the term virtual environments for scenes with a certain degree of interactivity, potentially in a semi-immersive (stereo and tracking, workbench) or immersive (CAVE, RealityCenter) context.

RITS 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 which have strong mutual dependencies: driving assistance, new transportation systems 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) paths. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component into 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. Automated vehicles

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. RITS 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 braking and steering systems. Regarding technical topics, several aspects of Cybercars have been developed at RITS 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. All our developments and demonstrations on our cybercars (cycabs, Yamaha AGV and new Cybus platforms) are using the RTMaps SDK development platform. 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. 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.

SEMAGRAMME Project-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. 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 and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

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, by 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

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, two applications are today more specifically targeted by the SMIS team. The first one deals with privacy preservation in EHR (Electronic Health Record) systems and PCEHR (Personally Controlled EHR). We are developing technologies tackling this issue and experiment them in the field. The second application area deals with privacy preservation in the context of personal Cloud, that is personal data hosted in dedicated servers staying under the holder's control (e.g., in a personal internet box or in a home automation box).

STARS Project-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 [6], [1], [8] (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 SUP 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 [67], [85] (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).

4.3.1. Topics

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longitudinal experimentation. 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 behavioral 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.

4.3.2. Ethical and Acceptability Issues

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. Now, Stars can benefit from the support of the COERLE (Comité Opérationnel d'Evaluation des Risques Légaux et Ethiques) to help it to respect ethical policies in its applications.

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.

TITANE Project-Team

4. Application Domains

4.1. Domains

In addition to tackling scientific challenges, our research on geometric modeling and processing is motivated by applications to computational engineering, reverse engineering, digital mapping and urban planning. The main deliverables of our research are algorithms with theoretical foundations. Ultimately we wish to contribute making geometry modeling and processing routine for practitioners who deal with real-world data. Our contributions may also be used as a sound basis for future software and technology developments.

Our ambition for technology transfer is to consolidate the components of our research experiments in the form of new software components for the CGAL (Computational Geometry Algorithms Library) library. Through CGAL we wish to contribute to the “standard geometric toolbox”, so as to provide a generic answer to application needs instead of fragmenting our contributions. We already cooperate with the Inria spin-off company Geometry Factory, which commercializes CGAL, maintains it and provide technical support.

We also started increasing our research momentum with companies through advising Cifre Ph.D. theses and postdoctoral fellows.

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. 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.

WIMMICS Project-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 licenses, 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: Viseo Joint Laboratory, 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. 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.
- **Personal health data analysis and privacy** The “Quantified Self” movement has gained a large popularity these past few years. Today, it is possible to acquire data on many domains related to personal data. For instance, one can collect data on her daily activities, habits or health. It is also possible to measure performances in sports. This can be done thanks to sensors, communicating devices or even connected glasses (as currently being developed by companies such as Google, for instance). Obviously, such data, once acquired, can lead to valuable knowledge for these domains. For people having a specific disease, it might be important to know if they belong to a specific category that needs particular care. For an individual, it can be interesting to find a category that corresponds to her performances in a specific sport and then adapt her training with an adequate program. Meanwhile, for privacy reasons, people will be reluctant to share their personal data and make them public. Therefore, it is important to provide them solutions that can extract such knowledge from everybody’s data, while guaranteeing that their private data won’t be disclosed to anyone.
- **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 systems. 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. 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.