

Inria

RESEARCH CENTER

FIELD

Activity Report 2019

Section Application Domains

Edition: 2020-03-21

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

1. ANTIQUE Project-Team	9
2. ARIC Project-Team	13
3. AROMATH Project-Team	14
4. CAIRN Project-Team	16
5. CAMBIUM Project-Team	17
6. CAMUS Project-Team	18
7. CARAMBA Project-Team	19
8. CASCADE Project-Team	21
9. CASH Project-Team (section vide)	22
10. CELTIQUE Project-Team (section vide)	23
11. CIDRE Project-Team (section vide)	24
12. COMETE Project-Team	25
13. CONVECS Project-Team	26
14. CORSE Project-Team	27
15. DATASHAPE Project-Team (section vide)	28
16. DATASPHERE Team	29
17. DEDUCTEAM Project-Team	30
18. GALLINETTE Project-Team (section vide)	31
19. GAMBLE Project-Team	32
20. GRACE Project-Team	33
21. HYCOMES Project-Team (section vide)	35
22. Kairos Project-Team	36
23. KOPERNIC Team	37
24. LFANT Project-Team (section vide)	38
25. MEXICO Project-Team	39
26. MOCQUA Team	41
27. OURAGAN Project-Team	43
28. PACAP Project-Team	46
29. PARKAS Project-Team	47
30. PARSIFAL Project-Team	48
31. PESTO Project-Team	49
32. PIR2 Project-Team (section vide)	50
33. POLSYS Project-Team (section vide)	51
34. PRIVATICS Project-Team	52
35. PROSECCO Project-Team	54
36. SECRET Project-Team	55
37. SPADES Project-Team	56
38. SPECFUN Project-Team	57
39. STAMP Project-Team	58
40. SUMO Project-Team	60

41. TAMIS Project-Team (section vide)	62
42. TEA Project-Team	63
43. TOCCATA Project-Team	65
44. VERIDIS Project-Team	66

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

45. ACUMES Project-Team	67
46. BONUS Project-Team	71
47. CAGE Project-Team	73
48. CAGIRE Project-Team	77
49. CARDAMOM Project-Team	79
50. CELESTE Project-Team	84
51. COMMANDS Project-Team	86
52. CQFD Project-Team	87
53. DEFI Project-Team	88
54. DISCO Project-Team	91
55. ECUADOR Project-Team	92
56. ELAN Project-Team	95
57. FACTAS Project-Team	97
58. GAMMA Project-Team (section vide)	102
59. GEOSTAT Project-Team	103
60. I4S Project-Team (section vide)	105
61. INOCS Project-Team	106
62. MATHERIALS Project-Team (section vide)	107
63. MATHRISK Project-Team	108
64. MCTAO Project-Team	109
65. MEMPHIS Project-Team	112
66. MEPHYSTO Team (section vide)	116
67. MINGUS Project-Team	117
68. MISTIS Project-Team	119
69. MODAL Project-Team	120
70. MOKAPLAN Project-Team (section vide)	121
71. NACHOS Project-Team	122
72. NANO-D Team (section vide)	127
73. NECS Team	128
74. POEMS Project-Team (section vide)	130
75. QUANTIC Project-Team	131
76. RANDOPT Project-Team	132
77. RAPSODI Project-Team	133
78. REALOPT Project-Team	135
79. SEQUEL Project-Team	139
80. SIERRA Project-Team	140

81. SIMSMART Project-Team	141
82. SPHINX Project-Team	142
83. TAU Project-Team	143
84. TOSCA Team	145
85. TRIPOP Project-Team	148
86. TROPICAL Project-Team	152
87. VALSE Project-Team	153

DIGITAL HEALTH, BIOLOGY AND EARTH

88. ABS Project-Team (section vide)	156
89. AIRSEA Project-Team	157
90. ANGE Project-Team	158
91. ARAMIS Project-Team	161
92. ATHENA Project-Team	162
93. BEAGLE Project-Team	163
94. BIGS Project-Team	164
95. BIOCORE Project-Team	165
96. BIOVISION Project-Team	167
97. CAMIN Project-Team (section vide)	168
98. CAPSID Project-Team	169
99. CARMEN Project-Team	171
100. CASTOR Project-Team (section vide)	172
101. COFFEE Project-Team	173
102. COMMEDIA Project-Team	175
103. DRACULA Project-Team (section vide)	176
104. DYLISS Project-Team	177
105. EMPENN Project-Team	178
106. EPIONE Project-Team (section vide)	181
107. ERABLE Project-Team	182
108. FLUMINANCE Project-Team (section vide)	183
109. GENSCALE Project-Team	184
110. IBIS Project-Team (section vide)	185
111. LEMON Project-Team (section vide)	186
112. LIFEWARE Project-Team	187
113. M3DISIM Project-Team (section vide)	189
114. MAGIQUE-3D Project-Team	190
115. MAMBA Project-Team	192
116. MATHNEURO Project-Team (section vide)	201
117. MIMESIS Team	202
118. MNEMOSYNE Project-Team	203
119. MONC Project-Team	204
120. MORPHEME Project-Team (section vide)	205

121. MOSAIC Project-Team (section vide)	206
122. NEUROSYS Project-Team	207
123. NUMED Project-Team (section vide)	208
124. OPIS Project-Team	209
125. PARIETAL Project-Team	212
126. PLEIADE Project-Team	214
127. REO Team	218
128. SERENA Project-Team	220
129. SERPICO Project-Team	221
130. SISTM Project-Team	223
131. STEEP Project-Team	225
132. TONUS Project-Team	229
133. XPOP Project-Team	230

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

134. AGORA Project-Team	235
135. ALPINES Project-Team	236
136. AVALON Project-Team	238
137. COAST Project-Team	240
138. COATI Project-Team	241
139. CTRL-A Project-Team	242
140. DANTE Project-Team	243
141. DATAMOVE Project-Team	245
142. DELYS Project-Team (section vide)	253
143. DIANA Project-Team (section vide)	254
144. DIONYSOS Project-Team	255
145. DIVERSE Project-Team (section vide)	256
146. DYOGENE Project-Team	257
147. EASE Project-Team	258
148. EVA Project-Team	260
149. FOCUS Project-Team	263
150. FUN Project-Team (section vide)	264
151. GANG Project-Team	265
152. HIEPACS Project-Team	266
153. INDES Project-Team	269
154. KERDATA Project-Team	270
155. MARACAS Team	271
156. MIMOVE Project-Team	273
157. Myriads Project-Team	276
158. NEO Project-Team	277
159. POLARIS Project-Team	278
160. RESIST Team	280

161. RMOD Project-Team	282
162. ROMA Project-Team	283
163. SOCRATE Project-Team (section vide)	284
164. SPIRALS Project-Team	285
165. STACK Project-Team	286
166. STORM Project-Team	288
167. TADAAM Project-Team	289
168. TRIBE Project-Team (section vide)	290
169. WHISPER Project-Team	291
170. WIDE Project-Team (section vide)	292

PERCEPTION, COGNITION AND INTERACTION

171. ALICE Team	293
172. ALMANACH Project-Team	294
173. Auctus Team	295
174. AVIZ Project-Team	296
175. CEDAR Project-Team	297
176. CHORALE Team	298
177. CHROMA Project-Team	299
178. COML Team	301
179. DEFROST Project-Team	302
180. EX-SITU Project-Team	305
181. FLOWERS Project-Team	306
182. GRAPHDECO Project-Team (section vide)	308
183. GRAPHIK Project-Team	309
184. HEPHAISTOS Project-Team (section vide)	310
185. HYBRID Project-Team	311
186. ILDA Project-Team	312
187. IMAGINE Project-Team	313
188. LACODAM Project-Team	314
189. LARSEN Project-Team	318
190. LINKMEDIA Project-Team	319
191. LINKS Project-Team	320
192. LOKI Project-Team	321
193. MAGNET Project-Team	322
194. MAGRIT Team	323
195. MANAO Project-Team	324
196. MAVERICK Project-Team	325
197. MFX Project-Team	326
198. MIMETIC Project-Team	327
199. MOEX Project-Team (section vide)	331
200. MORPHEO Project-Team	332

201. MULTISPEECH Project-Team	333
202. ORPAILLEUR Project-Team	335
203. PANAMA Project-Team	336
204. PERCEPTION Project-Team (section vide)	337
205. PERVASIVE Project-Team	338
206. PETRUS Project-Team	341
207. POTIOC Project-Team	342
208. RAINBOW Project-Team (section vide)	343
209. RITS Project-Team	344
210. SEMAGRAMME Project-Team	346
211. SIROCCO Project-Team	347
212. Stars Project-Team (section vide)	349
213. THOTH Project-Team	350
214. TITANE Project-Team	351
215. TYREX Project-Team	352
216. VALDA Project-Team	353
217. WILLOW Team	355
218. WIMMICS Project-Team	356
219. ZENITH Project-Team	359

ANTIQUÉ Project-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 the 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, and monitoring typically relies on a *parallel* structure, where several threads are executed concurrently, 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. Models of mechanistic interactions between proteins

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 one 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 analyses based on abstract interpretation, and 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 in order to identify key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify 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.

4.4. Consensus

Fault-tolerant distributed systems provide a dependable service on top of unreliable computers and networks. Famous examples are geo-replicated data-bases, distributed file systems, or blockchains. Fault-tolerant protocols replicate the system and ensure that all (unreliable) replicas are perceived from the outside as one single reliable machine. To give the illusion of a single reliable machine “consensus” protocols force replicas to agree on the “current state” before making this state visible to an outside observer. We are interested in (semi-)automatically proving the total correctness of consensus algorithms in the benign case (messages are lost or processes crash) or the Byzantine case (processes may lie about their current state). In order to do this, we first define new reduction theorems to simplify the behaviors of the system and, second, we introduce new static analysis methods to prove the total correctness of adequately simplified systems. We focus on static analysis based Satisfiability Modulo Theories (SMT) solvers which offers a good compromise between automation and expressiveness. Among our benchmarks are Paxos, PBFT (Practical Byzantine Fault-Tolerance), and blockchain algorithms (Red-Belly, Tendermint, Algorand). These are highly challenging benchmarks, with a lot of non-determinism coming from the interleaving semantics and from the adversarial environment in which correct processes execute, environment that can drop messages, corrupt them, etc. Moreover, these systems were originally designed for a few servers but today are deployed on networks with thousands of nodes. The “optimizations” for scalability can no longer be overlooked and must be considered as integral part of the algorithms, potentially leading to specifications weaker than the so much desired consensus.

4.5. Models of growth

In systems and synthetic biology (engineered systems) one would like study the environment of a given cellular process (such as signaling pathways mentioned earlier) and the ways in which that process interacts with different resources provided by the host. To do this, we have built coarse-grained models of cellular physiology which summarize fundamental processes (transcription, translation, transport, metabolism). such models describe global growth in mechanistic way and allow one to plug the model of one’s process of interest into a simplified and yet realistic and reactive model of the process interaction with its immediate environment. A first ODE-based deterministic version of this model [26] explaining the famous bacterial growth laws and how the allocation of resources to different genomic sectors depends on the growth conditions- was published in 2015 and has already received nearly 150 citations. The model also allows one to bridge between population genetic models which describe cells in terms of abstract features and fitness and intra-cellular models. For instance, we find that fastest growing strategies are not evolutionary stable in competitive experiments. We also find that vastly different energy storage strategies exist [24]. In a recent article [25] in *Nature Communications* we build a stochastic version of the above model. We predict the empirical size and doubling time distributions as a function of growth conditions. To be able to fit the parameters of the model to available single-cell data (note that the fitting constraints are far tighter than in the deterministic case), we introduce new techniques for the approximation of reaction-division systems which generalize continuous approximations of Langevin type commonly used for pure reaction systems. We also use cross-correlations to visualize causality and modes in noise propagation in the model (in a way reminiscent to abstract computational traces mentioned earlier). In other work, we show how to connect our new class of models to more traditional ones stemming from “flux balance analysis” by introducing an allocation vector which allows one to assign a formal growth rate to a class of reaction systems [20].

4.6. Static analysis of data science software

Nowadays, thanks to advances in machine learning and the availability of vast amounts of data, computer software plays an increasingly important role in assisting or even autonomously performing tasks in our daily lives. As data science software becomes more and more widespread, we become increasingly vulnerable to programming errors. In particular, programming errors that do not cause failures can have serious consequences since code that produces an erroneous but plausible result gives no indication that something went wrong. This issue becomes particularly worrying knowing that machine learning software, thanks to its ability

to efficiently approximate or simulate more complex systems, is slowly creeping into mission critical scenarios. However, programming errors are not the only concern. Another important issue is the vulnerability of machine learning models to adversarial examples, that is, small input perturbations that cause the model to misbehave in unpredictable ways. More generally, a critical issue is the notorious difficulty to interpret and explain machine learning software. Finally, as we are witnessing widespread adoption of software with far-reaching societal impact — i.e., to automate decision-making in fields such as social welfare, criminal justice, and even health care — a number of recent cases have evidenced the importance of ensuring software fairness as well as data privacy. Going forward, data science software will be subject to more and more legal regulations (e.g., the European General Data Protection Regulation adopted in 2016) as well as administrative audits. It is thus paramount to develop method and tools that can keep up with these developments and enhance our understanding of data science software and ensure it behaves correctly and reliably. In particular, we are interesting in developing new static analyses specifically tailored to the idiosyncrasies of data science software. This makes it a new and exciting area for static analysis, offering a wide variety of challenging problems with huge potential impact on various interdisciplinary application domains [13].

ARIC Project-Team

4. Application Domains

4.1. 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.

4.2. Cryptography, Cryptology, Communication Theory

Lattice reduction algorithms have direct applications in

- public-key cryptography;
- diophantine equations;
- communications theory.

AROMATH Project-Team

4. Application Domains

4.1. Geometric modeling for Design and Manufacturing.

The main domain of applications that we consider for the methods we develop is Computer Aided Design and Manufacturing.

Computer-Aided Design (CAD) involves creating digital models defined by mathematical constructions, from geometric, functional or aesthetic considerations. Computer-aided manufacturing (CAM) uses the geometrical design data to control the tools and processes, which lead to the production of real objects from their numerical descriptions.

CAD-CAM systems provide tools for visualizing, understanding, manipulating, and editing virtual shapes. They are extensively used in many applications, including automotive, shipbuilding, aerospace industries, industrial and architectural design, prosthetics, and many more. They are also widely used to produce computer animation for special effects in movies, advertising and technical manuals, or for digital content creation. Their economic importance is enormous. Their importance in education is also growing, as they are more and more used in schools and educational purposes.

CAD-CAM has been a major driving force for research developments in geometric modeling, which leads to very large software, produced and sold by big companies, capable of assisting engineers in all the steps from design to manufacturing.

Nevertheless, many challenges still need to be addressed. Many problems remain open, related to the use of efficient shape representations, of geometric models specific to some application domains, such as in architecture, naval engineering, mechanical constructions, manufacturing ...Important questions on the robustness and the certification of geometric computation are not yet answered. The complexity of the models which are used nowadays also appeals for the development of new approaches. The manufacturing environment is also increasingly complex, with new type of machine tools including: turning, 5-axes machining and wire EDM (Electrical Discharge Machining), 3D printer. It cannot be properly used without computer assistance, which raises methodological and algorithmic questions. There is an increasing need to combine design and simulation, for analyzing the physical behavior of a model and for optimal design.

The field has deeply changed over the last decades, with the emergence of new geometric modeling tools built on dedicated packages, which are mixing different scientific areas to address specific applications. It is providing new opportunities to apply new geometric modeling methods, output from research activities.

4.2. Geometric modeling for Numerical Simulation and Optimization

A major bottleneck in the CAD-CAM developments is the lack of interoperability of modeling systems and simulation systems. This is strongly influenced by their development history, as they have been following different paths.

The geometric tools have evolved from supporting a limited number of tasks at separate stages in product development and manufacturing, to being essential in all phases from initial design through manufacturing.

Current Finite Element Analysis (FEA) technology was already well established 40 years ago, when CAD-systems just started to appear, and its success stems from using approximations of both the geometry and the analysis model with low order finite elements (most often of degree ≤ 2).

There has been no requirement between CAD and numerical simulation, based on Finite Element Analysis, leading to incompatible mathematical representations in CAD and FEA. This incompatibility makes interoperability of CAD/CAM and FEA very challenging. In the general case today this challenge is addressed by expensive and time-consuming human intervention and software developments.

Improving this interaction by using adequate geometric and functional descriptions should boost the interaction between numerical analysis and geometric modeling, with important implications in shape optimization. In particular, it could provide a better feedback of numerical simulations on the geometric model in a design optimization loop, which incorporates iterative analysis steps.

The situation is evolving. In the past decade, a new paradigm has emerged to replace the traditional Finite Elements by B-Spline basis element of any polynomial degree, thus in principle enabling exact representation of all shapes that can be modeled in CAD. It has been demonstrated that the so-called isogeometric analysis approach can be far more accurate than traditional FEA.

It opens new perspectives for the interoperability between geometric modeling and numerical simulation. The development of numerical methods of high order using a precise description of the shapes raises questions on piecewise polynomial elements, on the description of computational domains and of their interfaces, on the construction of good function spaces to approximate physical solutions. All these problems involve geometric considerations and are closely related to the theory of splines and to the geometric methods we are investigating. We plan to apply our work to the development of new interactions between geometric modeling and numerical solvers.

CAIRN Project-Team

4. Application Domains

4.1. Panorama

keywords: Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography, Machine Learning, Deep Learning, Image and Signal Processing.

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.

Wireless Communication is our privileged application domain. Our research includes the prototyping of (subsets of) such applications on reconfigurable and programmable platforms. For this application domain, the high computational complexity of the 5G Wireless Communication Systems calls for the design of high-performance and energy-efficient architectures. In **Wireless Sensor Networks** (WSN), where each wireless node is expected to operate without battery replacement for significant periods of time, energy consumption is the most important constraint. 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.

Other important fields are also considered: hardware cryptographic and security modules, high-rate optical communications, machine learning, data mining, and multimedia processing.

CAMBIUM Project-Team

4. Application Domains

4.1. Formal methods

We develop techniques and tools for the formal verification of critical software:

- program logics based on CFML and Iris for the deductive verification of software, including concurrency and algorithmic complexity aspects;
- verified development tools such as the CompCert verified C compiler, which extends properties established by formal verification at the source level all the way to the final executable code.

Some of these techniques have already been used in the nuclear industry (MTU Friedrichshafen uses CompCert to develop emergency diesel generators) and are under evaluation in the aerospace industry.

4.2. High-assurance software

Software that is not critical enough to undergo formal verification can still benefit greatly, in terms of reliability and security, from a functional, statically-typed programming language. The OCaml type system offers several advanced tools (generalized algebraic data types, abstract types, extensible variant and object types) to express many data structure invariants and safety properties and have them automatically enforced by the type-checker. This makes OCaml a popular language to develop high-assurance software, in particular in the financial industry. OCaml is the implementation language for the Tezos blockchain and cryptocurrency. It is also used for automated trading at Jane Street and for modeling and pricing of financial contracts at Bloomberg, Lexifi and Simcorp. OCaml is also widely used to implement code verification and generation tools at Facebook, Microsoft, CEA, Esterel Technologies, and many academic research groups, at Inria and elsewhere.

4.3. Design and test of microprocessors

The **diy** tool suite and the underlying methodology is in use at ARM Ltd to design and test the memory model of ARM architectures. In particular, the internal reference memory model of the ARMv8 (or AArch64) architecture has been written “in house” in Cat, our domain-specific language for specifying and simulating memory models. Moreover, our test generators and runtime infrastructure are used routinely at ARM to test various implementations of their architectures.

4.4. Teaching programming

Our work on the OCaml language family has an impact on the teaching of programming. OCaml 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. The MOOC “Introduction to Functional Programming in OCaml”, developed at University Paris Diderot, is available on the France Université Numérique platform and comes with an extensive platform for self-training and automatic grading of exercises, developed in OCaml itself.

CAMUS Project-Team

4. Application Domains

4.1. Application Domains

Computational performance being our main objective, our 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.

CARAMBA Project-Team

4. Application Domains

4.1. Better Awareness and Avoidance of Cryptanalytic Threats

Our study of the Number Field Sieve family of algorithms aims at showing how the threats underlying various supposedly hard problems are real. Our record computations, as well as new algorithms, contribute to having a scientifically accurate assessment of the feasibility limit for these problems, given academic computing resources. The data we provide in this way is a primary ingredient for government agencies whose purpose includes guidance for the choice of appropriate cryptographic primitives. For example the French ANSSI⁰, German BSI, or the NIST⁰ in the United States base their recommendations on such computational achievements.

The software we make available to achieve these cryptanalytic computations also allows us to give cost estimates for potential attacks to cryptographic systems that are taking the security/efficiency/legacy compatibility trade-offs too lightly. Attacks such as LogJam [26] are understood as being serious concerns thanks to our convincing proof-of-concepts. In the LogJam context, this impact has led to rapid worldwide security advisories and software updates that eventually defeat some potential intelligence threats and improve confidentiality of communications.

4.2. Promotion of Better Cryptography

We also promote the switch to algebraic curves as cryptographic primitives. Those offer nice speed and excellent security, while primitives based on elementary number theory (integer factorization, discrete logarithm in finite fields), which underpin e.g., RSA, are gradually forced to adopt unwieldy key sizes so as to comply with the desired security guarantees of modern cryptography. Our contributions to the ultimate goal of having algebraic curves eventually take over the cryptographic landscape lie in our fast arithmetic contributions, our contributions to the point counting problem, and more generally our expertise on the diverse surrounding mathematical objects, or on the special cases where the discrete logarithm problem is not hard enough and should be avoided.

We also promote cryptographically sound electronic voting, for which we develop the Belenios prototype software (licensed under the AGPL). It depends on research made in collaboration with the PESTO team, and provides stronger guarantees than current state of the art.

4.3. Key Software Tools

The vast majority of our work is eventually realized as software. We can roughly categorize it in two groups. Some of our software covers truly fundamental objects, such as the GNU MPFR, GNU MPC, GF2X, or MPFQ packages. To their respective extent, these software packages are meant to be included or used in broader projects. For this reason, it is important that the license chosen for this software allows proper reuse, and we favor licenses such as the LGPL, which is not restrictive. We can measure the impact of this software by the way it is used in e.g., the GNU Compiler Collection (GCC), in Victor Shoup's Number Theory Library (NTL), or in the Sage computer algebra system. The availability of these software packages in most Linux distributions is also a good measure for the impact of our work.

⁰In [27], the minimal recommended RSA key size is 2048 bits for usage up to 2030. See also Annex B, in particular Section B.1 "Records de calculs cryptographiques".

⁰The work [31] is one of only two academic works cited by NIST in the initial version (2011) of the report [33].

We also develop more specialized software. Our flagship software package is Cado-NFS [15], and we also develop some others with various levels of maturity, such as GMP-ECM, CMH, or Belenios, aiming at quite diverse targets. Within the lifespan of the CARAMBA project, we expect more software packages of this kind to be developed, specialized towards tasks relevant to our research targets: important mathematical structures attached to genus 2 curves, generation of cryptographically secure curves, or tools for attacking cryptographically hard problems. Such software both illustrates our algorithms, and provides a base on which further research work can be established. Because of the very nature of these specialized software packages as research topics in their own right, needing both to borrow material from other projects, and being possible source of inspiring material for others, it is again important that these be developed in a free and open-source development model.

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.

CASH Project-Team (section vide)

CELTIQUE Project-Team (section vide)

CIDRE Project-Team (section vide)

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Catuscia Palamidessi, Konstantinos Chatzikokolakis, Ehab Elsalamouny, Ali Kassem, Anna Pazzi, Marco Romanelli, Natasha Fernandes.

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.

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 § 7.5) illustrates the diversity of applications:

- *Bioinformatics*: genetic regulatory networks, nutritional stress response, metabolic pathways,
- *Component-based systems*: Web services, peer-to-peer networks,
- *Cloud computing*: self-deployment protocols, dynamic reconfiguration protocols,
- *Fog and IoT*: stateful IoT applications in the fog,
- *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, train supervision 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.

CORSE Project-Team

4. Application Domains

4.1. Transfer

The main industrial sector related to the research activities of CORSE 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 run-time technology. These applications are based over numerical kernels (linear algebra, FFT, convolution...) that can be adapted on a large spectrum of architectures. More specifically, an important activity concerns the optimization of machine learning applications for some high-performance accelerators. Members of CORSE already maintain fruitful and strong collaborations with several companies such as STMICROELECTRONICS, Atos/Bull, Kalray.

DATASHAPE Project-Team (section vide)

DATASPHERE Team

4. Application Domains

4.1. Governance

- City governance, local democracy and interaction with citizens.
- Local governance versus global norms and control.
- Strategy beyond public open data.
- Smart city governance.

4.2. CyberSecurity

- Cyber-strategy, defense and security in an evolving world shaped by the digital in particular China/Russia/US cyber-strategy.
- Data strategy for the digital economy, cross border intermediation, platform strategie.
- Strategy of Artificial Intelligence, transparency/acceptability/explainability of AI.
- Cartography of the cyberspace.
- Network, BGP security.

4.3. Anthropocene

- Adaptation to the conditions of the anthropocene, digital control of resources and homeostasis.
- Geopolitics of the environmental challenges, adaptation and mitigation.
- Contemporaneity of the digital revolution and global warming.

DEDUCTEAM Project-Team

4. Application Domains

4.1. Interoperability

Our main impact applications, for instance to proofs of programs, or to air traffic control, are through our cooperation with other teams.

As a matter of fact, we view our work on interoperability and on the design of a formal proof encyclopedia as a service to the formal proof community.

GALLINETTE Project-Team (section vide)

GAMBLE Project-Team

4. Application Domains

4.1. Applications of computational geometry

Many domains of science can benefit from the results developed by GAMBLE. Curves and surfaces are ubiquitous in all sciences to understand and interpret raw data as well as experimental results. Still, the non-linear problems we address are rather basic and fundamental, and it is often difficult to predict the impact of solutions in that area. The short-term industrial impact is likely to be small because, on basic problems, industries have used ad hoc solutions for decades and have thus got used to it. The example of our work on quadric intersection is typical: even though we were fully convinced that intersecting 3D quadrics is such an elementary/fundamental problem that it ought to be useful, we were the first to be astonished by the scope of the applications of our software ⁰ (which was the first and still is the only one —to our knowledge— to compute robustly and efficiently the intersection of 3D quadrics) which has been used by researchers in, for instance, photochemistry, computer vision, statistics, and mathematics. Our work on certified drawing of plane (algebraic) curves falls in the same category. It seems obvious that it is widely useful to be able to draw curves correctly (recall also that part of the problem is to determine where to look in the plane) but it is quite hard to come up with specific examples of fields where this is relevant. A contrario, we know that certified meshing is critical in mechanical-design applications in robotics, which is a non-obvious application field. There, the singularities of a manipulator often have degrees higher than 10 and meshing the singular locus in a certified way is currently out of reach. As a result, researchers in robotics can only build physical prototypes for validating, or not, the approximate solutions given by non-certified numerical algorithms.

The fact that several of our pieces of software for computing non-Euclidean triangulations had already been requested by users long before they become public in CGAL is a good sign for their wide future impact. This will not come as a surprise, since most of the questions that we have been studying followed from discussions with researchers outside computer science and pure mathematics. Such researchers are either users of our algorithms and software, or we meet them in workshops. Let us only mention a few names here. Rien van de Weijgaert [55], [69] (astrophysicist, Groningen, NL) and Michael Schindler [66] (theoretical physicist, ENSPCI, CNRS, France) used our software for 3D periodic weighted triangulations. Stephen Hyde and Vanessa Robins (applied mathematics and physics at Australian National University) used our package for 3D periodic meshing. Olivier Faugeras (neuromathematics, Inria Sophia Antipolis) had come to us and mentioned his needs for good meshes of the Bolza surface [45] before we started to study them. Such contacts are very important both to get feedback about our research and to help us choose problems that are relevant for applications. These problems are at the same time challenging from the mathematical and algorithmic points of view. Note that our research and our software are generic, i.e., we are studying fundamental geometric questions, which do not depend on any specific application. This recipe has made the success of the CGAL library.

Probabilistic models for geometric data are widely used to model various situations ranging from cell phone distribution to quantum mechanics. The impact of our work on probabilistic distributions is twofold. On the one hand, our studies of properties of geometric objects built on such distributions will yield a better understanding of the above phenomena and has potential impact in many scientific domains. On the other hand, our work on simulations of probabilistic distributions will be used by other teams, more maths oriented, to study these distributions.

⁰QI: [web](#).

GRACE Project-Team

4. Application Domains

4.1. Application Domain: cybersecurity

Participants: Guénaél Renault, Benjamin Smith, François Morain, Alexis Challande, Simon Montoya, Maxime Anvari.

We are interesting in developing some interactions between cryptography and cybersecurity. In particular, we develop some researches in embedded security (side channels and fault attack), software security (finding vulnerability efficiently) and privacy (security of TOR).

4.2. Application Domain: blockchains

Participants: Daniel Augot, Sarah Bordage, Matthieu Rambaud, Lucas Benmouffok, Hanna-Mae Bissierier.

The huge interest shown by companies for blockchains and cryptocurrencies have attracted the attention of mainstream industries for new, advanced uses of cryptographic, beyond confidentiality, integrity and authentication. In particular, zero-knowledge proofs, computation with encrypted data, etc. are now revealing their potential in the blockchain context. Team Grace is investigating two topics in these areas: secure multiparty computation and so-called “STARKS”.

Secure multiparty computation enables several participants to compute a common function of data they each secretly own, without each participant revealing his data to the other participants. This area has seen great progress in recent years, and the cryptographic protocols are now mature enough for practical use. This topic is new to project-team Grace, and we will investigate it in the context of blockchains, through the lenses of use for private “smart contracts”. A PhD student has been hired since October, funded by IRT System-X.

Daniel Augot is involved in blockchains from the point of view of cryptography for better blockchains, mainly for improving privacy. A PhD student has been enrolled at IRT System-X, to study practical use cases of Secure Multiparty Computation.

Also Daniel Augot, together with Julian Prat (economist, ENSAE), is leading a Polytechnique teaching and research “chair”, funded by CapGemini, for blockchains in the industry, B2B platforms, supply chains, etc.

4.3. Cloud storage

The team is concerned with several aspect of reliability and security of cloud storage, obtained mainly with tools from coding theory. On the privacy side, we build protocols for so-called Private Information Retrieval which enable a user to query a remote database for an entry, while not revealing his query. For instance, a user could query a service for stock quotes without revealing with company he is interested in. On the availability side, we study protocols for proofs of retrievability, which enable a user to get assurance that a huge file is still available on a remote server, with a low bandwidth protocol which does not require to download the whole file. For instance, in a peer-to-peer distributed storage system, where nodes could be rewarded for storing data, they can be audited with proof of retrievability protocols to make sure they indeed hold the data.

We investigate these problems with algebraic coding theory, mainly codes with locality (locally decodable codes, locally recoverable codes, and so on).

An M2 intern, Maxime Roméas, Bordeaux university, studied the constructive cryptography model, "A study of the Constructive Cryptography model of Maurer et. al." 5 months, followed by a PhD grant from IP Paris/Ecole Polytechnique for a 3-year doctorate (Oct 2019-Sept 2022): "The Constructive Cryptography paradigm applied to Interactive Cryptographic Proofs".

The Constructive Cryptography framework redefines basic cryptographic primitives and protocols starting from discrete systems of three types (resources, converters, and distinguishers). This not only permits to construct them effectively, but also lighten and sharpen their security proofs. One strength of this model is its composability. The purpose of the PhD is to apply this model to rephrase existing interactive cryptographic proofs so as to assert their genuine security, as well as to design new proofs. The main concern here is security and privacy in Distributed Storage settings.

HYCOMES Project-Team (section vide)

Kairos Project-Team

4. Application Domains

4.1. Cyber-Physical and Embedded Systems

We have historical contacts with industrial and academic partners in the domains of avionics and embedded electronics (Airbus, Thales, Safran). We have new collaborations in the fields of satellites (Thales Alenia Space) and connected cars (Renault Software Labs). These provide for use case and new issues in CPS co-modeling and co-design (Digital Twins) further described in New Results section.

4.2. Connected Objects in the Internet Of Things

Due to increasing collaborations with local partners, we have recently considered Smart Contracts (as popularized in Blockchain frameworks), as a way to formally established specification of behavioral system traces, applied to connected objects in a IoT environment. The new ANR project SIM is based on this.

KOPERNIC Team

4. Application Domains

4.1. Avionics

This work is based on a direct collaboration between Airbus and Inria, complementary to collaborative projects like PIA LEOC Capacites and CIFRE thesis. The time critical solutions in this context are based on temporal and spatial isolation of the programs and the understanding of multicore interferences is crucial. Our contributions belong mainly to the solutions space for the objective identified in Section 3.1 .

4.2. Railway

This work is based on a direct collaboration with Clearsy and SNCF, complementary to collaborative projects like PIA BGLE Departs and FUI 21 Waruna. The time critical solutions in this context concern both the proposition of an appropriate scheduler and associated schedulability analyses. Our contributions belong to the solutions space of problems dealt within the objectives identified in Section 3.1 .

4.3. Autonomous cars

This work is based on a direct collaboration with RITS (Inria project team). The time critical solutions in this context concern the interaction between programs executed on multicore processors and messages transmitted through wireless communication channels. Our contributions belong to the solutions space of all three classes of problems dealt within the objectives identified in Section 3.2 .

4.4. Drones

This work is based on the collaborative project FUI/FEDER 22 Ceos. As in the case of autonomous cars, there is an interaction between programs and messages, suggesting that our contributions in this context belong to the solutions space of all three classes of problems dealt within the objectives identified in Section 3.2 .

LFANT Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

MEXICO's research is motivated by problems of system management in several domains, such as:

- 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.

Currently, we have no active cooperation on these subjects.

4.2. Biological Regulation Networks

Participants: Thomas Chatain, Matthias Fuegger, Stefan Haar, Serge Haddad, Juraj Kolcak, Hugues Mandon, Stefan Schwoon.

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. We have applied 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. 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. Our current research focusses cellular reprogramming on the one hand, and distributed algorithms in wild or synthetic biological systems on the other. The latter is a distributed algorithms' view on microbiological systems, both with the goal to model and analyze existing microbiological systems as distributed systems, and to design and implement distributed algorithms in synthesized microbiological systems. Envisioned

major long-term goals are drug production and medical treatment via synthesized bacterial colonies. We are approaching our goal of a distributed algorithm's view of microbiological systems from several directions: (i) Timing plays a crucial role in microbiological systems. Similar to modern VLSI circuits, dominating loading effects and noise render classical delay models unfeasible. In previous work we showed limitations of current delay models and presented a class of new delay models, so called involution channels. In [26] we showed that involution channels are still in accordance with Newtonian physics, even in presence of noise. (ii) In [7] we analyzed metastability in circuits by a three-valued Kleene logic, presented a general technique to build circuits that can tolerate a certain degree of metastability at its inputs, and showed the presence of a computational hierarchy. Again, we expect metastability to play a crucial role in microbiological systems, as similar to modern VLSI circuits, loading effects are pronounced. (iii) We studied agreement problems in highly dynamic networks without stability guarantees [28], [27]. We expect such networks to occur in bacterial cultures where bacteria communicate by producing and sensing small signal molecules like AHL. Both works also have theoretically relevant implications: The work in [27] presents the first approximate agreement protocol in a multidimensional space with time complexity independent of the dimension, working also in presence of Byzantine faults. In [28] we proved a tight lower bound on convergence rates and time complexity of asymptotic and approximate agreement in dynamic and classical static fault models. (iv) We are currently working with Manish Kushwaha (INRA), and Thomas Nowak (LRI) on biological infection models for E. coli colonies and M13 phages.

4.3. Metabolic Networks

Participant: Philippe Dague.

Analysis of metabolic networks in presence of biological (thermodynamical, kinetic, gene regulatory) constraints has been studied achieving a complete mathematical characterization of the solutions space at steady state (generalization of the elementary flux modes) and investigating related computing methods.

4.4. Transportation Systems

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

- **Autonomous Vehicles.** The validation of safety properties is a crucial concern for the design of computer guided systems, in particular for automated transport systems. Our approach consists in analyzing the interactions of a randomized environment (roads, cross-sections, etc.) with a vehicle controller.
- **Multimodal Transport Networks.** We are interested in predicting and harnessing the propagation of perturbations across different transportation modes.

MOCQUA Team

4. Application Domains

4.1. Quantum Computing

Quantum Computing is currently the most promising technology to extend Moore's law, whose end is expected with the engraving at 7 nm, in less than 5 years. Thanks to the exponential computational power it will bring, it will represent a decisive competitive advantage for those who will control it.

Quantum Computing is also a major security issue, since it allows us to break today's asymmetric cryptography. Hence, mastering quantum computing is also of the highest importance for national security concerns. Recent scientific and technical advances suggest that the construction of the first quantum computers will be possible in the coming years.

As a result, the major US players in the IT industry have embarked on a dramatic race, mobilizing huge resources: IBM, Microsoft, Google and Intel have each invested between 20 and 50 million euros, and are devoting significant budgets to attract and hire the best scientists on the planet. Some states have launched ambitious national programs, including Great Britain, the Netherlands, Canada, China, Australia, Singapore, and very recently Europe, with the upcoming 10-year FET Flagship program in Quantum Engineering.

While a large part of these resources are going towards R-&-D in quantum hardware, there is still an important need and real opportunities for leadership in the field of quantum software.

The Mocqua team contributes to the computer science approach to quantum computing, aka the quantum software approach. We aim at a better understanding of the power and limitations of the quantum computer, and therefore of its impact on society. We also contribute to ease the development of the quantum computer by filling the gap between the theoretical results on quantum algorithms and complexity and the recent progresses in quantum hardware.

4.2. Higher-Order Computing

The idea of considering functions as first-class citizens and allowing programs to take functions as inputs has emerged since the very beginning of theoretical computer science through Church's λ -calculus and is nowadays at the core of functional programming, a paradigm that is used in modern software and by digital companies (Google, Facebook, ...). In the meantime higher-order computing has been explored in many ways in the fields of logic and semantics of programming languages.

One of the central problems is to design programming languages that capture most of, if not all, the possible ways of computing with functions as inputs. There is no Church thesis in higher-order computing and many ways of taking a function as input can be considered: allowing parallel or only sequential computations, querying the input as a black-box or via an interactive dialog, and so on.

The Kleene-Kreisel computable functionals are arguably the broadest class of higher-order continuous functionals that could be computed by a machine. However their complexity is such that no current programming language can capture all of them. Better understanding this class of functions is therefore fundamental in order to identify the features that a programming language should implement to make the full power of higher-order computation expressible in such a language.

4.3. Simulation of Dynamical Systems by Cellular Automata

We aim at developing various tools to simulate and analyse the dynamics of spatially-extended discrete dynamical systems such as cellular automata. The emphasis of our approach is on the evaluation of the robustness of the models under study, that is, their capacity to resist various perturbations.

In the framework of pure computational questions, various examples of such systems have already been proposed for solving complex problems with a simple bio-inspired approach (e.g. the decentralized gathering problem [40]). We are now working on their transposition to various real-world situations. For example when one needs to understand the behaviour of large-scale networks of connected components such as wireless sensor networks. In this direction of research, a first work has been presented on how to achieve a decentralized diagnosis of networks made of simple interacting components and the results are rather encouraging [5]. Nevertheless, there are various points that remain to be studied in order to complete this model for its integration in a real network.

We have also tackled the question of the evaluation of the robustness of a swarming model proposed by A. Deutsch to mimic the self-organization process observed in various natural systems (birds, fishes, bacteria, etc.) [2]. We now wish to develop our simulation tools to apply them to various biological phenomena where a great number of agents are implied.

We are also currently extending the range of applications of these techniques to the field of economy. We have started a collaboration with Massimo Amato, a professor in economy at the Bocconi University in Milan. Our aim is to examine how to propose a decentralized view of a business-to-business market and propose agent-oriented and totally decentralized models of such markets. Various banks and large businesses have already expressed their interest in such modelling approaches.

OURAGAN Project-Team

4. Application Domains

4.1. Security of cryptographic systems

The study of the security of asymmetric cryptographic systems comes as an application of the work carried out in algorithmic number theory and revolves around the development and the use of a small number of general purpose algorithms (lattice reduction, class groups in number fields, discrete logarithms in finite fields, ...). For example, the computation of generators of principal ideals of cyclotomic fields can be seen as one of these applications since these are used in a number of recent public key cryptosystems.

The cryptographic community is currently very actively assessing the threat coming for the development of quantum computers. Indeed, such computers would permit tremendous progress on many number theoretic problems such as factoring or discrete logarithm computations and would put the security of current cryptosystem under a major risk. For this reason, there is a large global research effort dedicated to finding alternative methods of securing data. For example, the US standardization agency called NIST has recently launched a standardization process around this issue. In this context, OURAGAN is part of the competition and has submitted a candidate (which has not been selected) [40]. This method is based on number-theoretic ideas involving a new presumably difficult problem concerning the Hamming distance of integers modulo large numbers of Mersenne.

4.2. Robotics

Algebraic computations have tremendously been used in Robotics, especially in kinematics, since the last quarter of the 20th century [93]. For example, one can find algebraic proofs for the 40 possible solutions to the direct kinematics problem [117] for Stewart platforms and companion experiments based on Gröbner basis computations [83]. On the one hand, hard general kinematics problems involve too many variables for pure algebraic methods to be used in place of existing numerical or semi-numerical methods everywhere and everytime, and on the other hand, global algebraic studies allow to propose exhaustive classifications that cannot be reached by other methods, for some quite large classes.

Robotics is a long-standing collaborative work with LS2N (Laboratory of Numerical Sciences of Nantes). Work has recently focused on the offline study of mechanisms, mostly parallel, their singularities or at least some types of singularities (cuspidal robots [140]).

For most parallel or serial manipulators, pose variables and joints variables are linked by algebraic equations and thus lie on an algebraic variety. The two-kinematics problems (the direct kinematics problem - DKP- and the inverse kinematics problem - IKP) consist in studying the preimage of the projection of this algebraic variety onto a subset of unknowns. Solving the DKP remains to computing the possible positions for a given set of joint variables values while solving the IKP remains to computing the possible joints variables values for a given position. Algebraic methods have been deeply used in several situations for studying parallel and serial mechanisms, but finally their use stays quite confidential in the design process. Cylindrical Algebraic Decomposition coupled with variable's eliminations by means of Gröbner based computations can be used to model the workspace, the joint space and the computation of singularities. On the one hand, such methods suffer immediately when increasing the number of parameters or when working with imprecise data. On the other hand, when the problem can be handled, they might provide full and exhaustive classifications. The tools we use in that context [60], [59], [95], [97], [96] depend mainly on the resolution of parameter-based systems and therefore of study-dependent curves or flat algebraic surfaces (2 or 3 parameters), thus joining our thematic *Computational Geometry*.

4.3. Control theory

Certain problems studied in mathematical systems theory and control theory can be better understood and finely studied by means of algebraic structures and methods. Hence, the rich interplay between algebra, computer algebra, and control theory has a long history. For instance, the first main paper on Gröbner bases written by their creators, Buchberger, was published in Bose's book [46] on control theory of multidimensional systems. Moreover, the differential algebra approach to nonlinear control theory (see [72], [73] and the references therein) was a major motivation for the algorithmic study of differential algebra [47], [76]. Finally, the behaviour approach to linear systems theory [141], [119] advocates for an algorithmic study of algebraic analysis (see Section 2.2.4). More generally, control theory is porous to computer algebra since one finds algebraic criteria of all kinds in the literature even if the control theory community has a very few knowledge in computer algebra.

OURAGAN has a strong interest in the computer algebra aspects of mathematical systems theory and control theory related to both functional and polynomial systems, particularly in the direction of robust stability analysis and robust stabilization problems for multidimensional systems [46], [119] and infinite-dimensional systems [66] (such as, e.g., differential time-delay systems).

Let us shortly state a few points of our recent interests in this direction.

In control theory, stability analysis of linear time-invariant control systems is based on the famous Routh-Hurwitz criterion (late 19th century) and its relation with Sturm sequences and Cauchy index. Thus, stability tests were only involving tools for univariate polynomials [102]. While extending those tests to multidimensional systems or differential time-delay systems, one had to tackle multivariate problems recursively with respect to the variables [46]. Recent works use a mix of symbolic/numeric strategies, Linear Matrix Inequalities (LMI), sums of squares, etc. But still very few practical experiments are currently involving certified algebraic computations based on general solvers for polynomial equations. We have recently started to study certified stability tests for multidimensional systems or differential time-delay systems with an important observation: with a correct modelization, some recent algebraic methods – derived from our work in algorithmic geometry and shared with applications in robotics – can now handle previously impossible computations and lead to a better understanding of the problems to be solved [52], [54], [55]. The previous approaches seem to be blocked on a recursive use of one-variable methods, whereas our approach involves the direct processing of the problem for a larger number of variables.

The structural stability of n -D discrete linear systems (with $n \geq 2$) is a good source of problems of several kinds ranging from solving univariate polynomials to studying algebraic systems depending on parameters. For instance, we show [53], [54], [55] that the standard characterization of the structural stability of a multivariate rational transfer function (namely, the denominator of the transfer function does not have solutions in the unit polydisc of \mathbb{C}^n) is equivalent to deciding whether or not a certain system of polynomial equations has real solutions. The use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems has been validated in several situations from toy examples with parameters to state-of-the-art examples involving, e.g., the resolution of bivariate systems [51], [50].

The rich interplay between control theory, algebra, and computer algebra is also well illustrated with our recent work on robust stabilization problems for multidimensional and finite/infinite-dimensional systems [48], [123], [129], [132], [130], [131].

4.4. Signal processing

Due to numerous applications (e.g. sensor network, mobile robots), sources and sensors localization has intensively been studied in the literature of signal processing. The *anchor position self calibration problem* is a well-known problem which consists in estimating the positions of both the moving sources and a set of fixed sensors (anchors) when only the distance information between the points from the different sets is available. The position self-calibration problem is a particular case of the *Multidimensional Unfolding* (MDU) problem for the Euclidean space of dimension 3. In the signal processing literature, this problem is attacked by means of optimization problems (see [65] and the references therein). Based on computer algebra methods for

polynomial systems, we have recently developed a new approach for the MDU problem which yields closed-form solutions and a very efficient algorithm for the estimation of the positions [68] based only on linear algebra techniques. This first result, done in collaboration with Dagher (Inria Chile) and Zheng (DEFROST, Inria Lille), yielded a recent patent [67]. This result advocates for the study of other localization problems based on the computational polynomial techniques developed in OURAGAN.

In collaboration with *Safran Tech* (Barau, Hubert) and Dagher (Inria Chile), a symbolic-numeric study of the new *multi-carrier demodulation method* [92] has recently been initiated. *Gear fault diagnosis* is an important issue in aeronautics industry since a damage in a gearbox, which is not detected in time, can have dramatic effects on the safety of a plane. Since the vibrations of a spur gear can be modeled as a product of two periodic functions related to the gearbox kinematic, it is proposed to recover each function from the global signal by means of an optimal reconstruction problem which, based on Fourier analysis, can be rewritten as $\operatorname{argmin}_{u \in \mathbb{C}^n, v_1, v_2 \in \mathbb{C}^m} \|M - u v_1^{\star} - D u v_2^{\star}\|_F$, where $M \in \mathbb{C}^{n \times m}$ (resp. $D \in \mathbb{C}^{n \times n}$) is a given matrix with a special shape (resp. diagonal matrix), $\|\cdot\|_F$ is the Frobenius norm, and v^{\star} is the Hermitian transpose of v . We have recently obtained closed-form solutions for the exact problem, i.e., $M = u v_1^{\star} + D u v_2^{\star}$, which is a polynomial system with parameters. This first result gives interesting new insides for the study of the non-exact case, i.e. for the above optimization problem.

Our expertise on *algebraic parameter estimation problem*, developed in the former NON-A project-team (Inria Lille), will be further developed. Following this work [84], the problem consists in estimating a set θ of parameters of a signal $x(\theta, t)$ — which satisfies a certain dynamics — when the signal $y(t) = x(\theta, t) + \gamma(t) + \varpi(t)$ is observed, where γ denotes a structured perturbation and ϖ a noise. It has been shown that θ can sometimes be explicitly determined by means of closed-form expressions using iterated integrals of y . These integrals are used to filter the noise ϖ . Based on a combination of algebraic analysis techniques (rings of differential operators), differential elimination theory (Gröbner basis techniques for Weyl algebras), and operational calculus (Laplace transform, convolution), an algorithmic approach to algebraic parameter estimation problem has been initiated in [125] for a particular type of structured perturbations (i.e. bias) and was implemented in the Maple prototype NonA. The case of a general structured perturbation is still lacking.

PACAP Project-Team

4. Application Domains

4.1. Domains

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impact on any application domain that requires high performance execution (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.

PARKAS Project-Team

4. Application Domains

4.1. Embedded Control Software

Embedded control software defines the interactions of specialized hardware with the physical world. It normally ticks away unnoticed inside systems like medical devices, trains, aircraft, satellites, and factories. This software is complex and great effort is required to avoid potentially serious errors, especially over many years of maintenance and reuse.

Engineers have long designed such systems using block diagrams and state machines to represent the underlying mathematical models. One of the key insights behind synchronous programming languages is that these models can be executable and serve as the base for simulation, validation, and automatic code generation. This approach is sometimes termed Model-Based Development (MBD). The SCADE language and associated code generator allow the application of MBD in safety-critical applications. They incorporate ideas from LUSTRE, LUCID SYNCHRONE, and other programming languages.

4.2. Hybrid Systems Design and Simulation

Modern embedded systems are increasingly conceived as rich amalgams of software, hardware, networking, and physical processes. The terms Cyberphysical System (CPS) or Internet-of-Things (IoT) are sometimes used as labels for this point of view.

In terms of modeling languages, the main challenges are to specify both discrete and continuous processes in a single *hybrid* language, give meaning to their compositions, simulate their interactions, analyze the behavior of the overall system, and extract code either for target control software or more efficient, possibly online, simulation. Languages like Simulink and Modelica are already used in the design and analysis of embedded systems; it is more important than ever to understand their underlying principles and to propose new constructs and analyses.

PARSIFAL Project-Team

4. Application Domains

4.1. Automated Theorem Proving

The Parsifal team studies the structure of mathematical proofs, in ways that often makes them more amenable to automated theorem proving – automatically searching the space of proof candidates for a statement to find an actual proof – or a counter-example.

(Due to fundamental computability limits, fully-automatic proving is only possible for simple statements, but this field has been making a lot of progress in recent years, and is in particular interested with the idea of generating verifiable evidence for the proofs that are found, which fits squarely within the expertise of Parsifal.)

4.2. Proof-assistants

The team work on the structure of proofs also suggests ways that they could be presented to a user, edited and maintained, in particular in “proof assistants”, automated tool to assist the writing of mathematical proofs with automatic checking of their correctness.

4.3. Programming language design

Our work also gives insight on the structure and properties of programming languages. We can improve the design or implementation of programming languages, help programmers or language implementors reason about the correctness of the programs in a given language, or reason about the cost of execution of a program.

PESTO Project-Team

4. Application Domains

4.1. Cryptographic protocols

Security protocols, such as TLS, Kerberos, ssh or AKA (mobile communication), are the main tool for securing our communications. The aim of our work is to improve their security guarantees. For this, we propose models that are expressive enough to formally represent protocol executions in the presence of an adversary, formal definitions of the security properties to be satisfied by these protocols, and automated tools able to analyse them and possibly exhibit design flaws.

4.2. Automated reasoning

Many techniques for symbolic verification of security are rooted in automated reasoning. A typical example is equational reasoning used to model the algebraic properties of a cryptographic primitive. Our work therefore aims to improve and adapt existing techniques or propose new ones when needed for reasoning about security.

4.3. Electronic voting

Electronic elections have in the last years been used in several countries for politically binding elections. The use in professional elections is even more widespread. The aim of our work is to increase our understanding of the security properties needed for secure elections, propose techniques for analysing e-voting protocols, design of state-of-the-art voting protocols, but also to highlight the limitations of e-voting solutions.

4.4. Privacy in social networks

The treatment of information released by users on social networks can violate a user's privacy. The goal of our work is to allow users to control the information released while guaranteeing their privacy.

PI.R2 Project-Team (section vide)

POLSYS Project-Team (section vide)

PRIVATICS Project-Team

3. Application Domains

3.1. Domain 1: Privacy in smart environments

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. Cryptographic primitives

Our major application domain is the design of cryptographic primitives, especially for platforms with restricting implementation requirements. For instance, we aim at recommending (or designing) low-cost (or extremely fast) encryption schemes, or primitives which remain secure against quantum computers.

4.2. Code Reconstruction

To evaluate the quality of a cryptographic algorithm, it is usually assumed that its specifications are public, as, in accordance with Kerckhoffs principle, it would be dangerous to rely, even partially, on the fact that the adversary does not know those specifications. However, this fundamental rule does not mean that the specifications are known to the attacker. In practice, before mounting a cryptanalysis, it is necessary to strip off the data. This reverse-engineering process is often subtle, even when the data formatting is not concealed on purpose. A typical case is interception: some raw data, not necessarily encrypted, is observed out of a noisy channel. To access the information, the whole communication system has first to be disassembled and every constituent reconstructed. A transmission system actually corresponds to a succession of elements (symbol mapping, scrambler, channel encoder, interleaver...), and there exist many possibilities for each of them. In addition to the “preliminary to cryptanalysis” aspect, there are other links between those problems and cryptology. They share some scientific tools (algorithmics, discrete mathematics, probability...), but beyond that, there are some very strong similarities in the techniques.

SPADES Project-Team

4. Application Domains

4.1. Industrial Applications

Our applications are in the embedded system area, typically: transportation, energy production, robotics, telecommunications, the Internet of things (IoT), 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. Current Industrial Cooperations

Regarding applications and case studies with industrial end-users of our techniques, we cooperate with Orange Labs on software architecture for cloud services.

SPECFUN Project-Team

4. Application Domains

4.1. Computer Algebra in Mathematics

Our expertise in computer algebra and complexity-driven design of algebraic algorithms has applications in various domains, including:

- combinatorics, especially the study of combinatorial walks,
- theoretical computer science, like by the study of automatic sequences,
- number theory, by the analysis of the nature of so-called periods.

STAMP Project-Team

4. Application Domains

4.1. Mathematical Components

The Mathematical Components is the main by-product of an effort started almost two decades ago to provide a formally verified proof for a major theorem in group theory. Because this major theorem had a proof published in books of several hundreds of pages, with elements coming from character theory, other coming from algebra, and some coming from real analysis, it was an exercise in building a large library, with results in many domains, and in establishing clear guidelines for further increase and data search.

This library has proved to be a useful repository of mathematical facts for a wide area of applications, so that it has a growing community of users in many countries (Denmark, France, Germany, Japan, Singapore, Spain, Sweden, UK, USA, at the time of writing these lines in 2019) and for a wide variety of topics (transcendental number theory, elliptic curve cryptography, articulated robot kinematics, recently block chain foundations).

Interesting questions on this library range around the importance of decidability and proof irrelevance, the way to structure knowledge to automatically inherit theorems from one topic to another, the way to generate infrastructure to make this automation efficient and predictable. In particular, we want to concentrate on adding a new mathematical topic to this library: real analysis and then complex analysis (Mathematical Components Analysis).

On the front of automation, we are convinced that a higher level language is required to describe similarities between theories, to generate theorems that are immediate consequences of structures, etc, and for this reason, we invest in the development of a new language on top of the proof assistant (ELPI).

4.2. Proofs in cryptography

When we work on cryptography, we are interested in the formal verification of proofs showing that some cryptographic primitives provide good guarantees against unwanted access to information. Over the years we have developed a technique for this kind of reasoning that relies on a programming logic (close to Hoare logic) with probabilistic aspects and the capability to establish relations between several implementations of a problem. The resulting programming logic is called *probabilistic relational Hoare logic*. In more recent work, we have also started to study questions of *side-channel* attacks, where we wish to guarantee that opponents cannot gain access to protected knowledge, even if they observe specific features of execution, like execution time (to which the answer lies in *constant-time* execution) or partial access to memory bits (to which the answer lies in *masking*).

For this domain of application, we choose to work with a specific proof tool (EasyCrypt), which combines powerful first-order reasoning and uses of automatic tools, with a specific support for probabilistic relational Hoare Logic. The development of this EasyCrypt proof tool is one of the objectives of our team.

When it comes to formal proofs of resistance to side-channel attack, we contend that it is necessary to verify formally that the compiler used in the production of actually running code respects the resistance properties that were established in formally verified proofs. One of our objectives is to describe such a compiler (Jasmin) and show its strength on a variety of applications.

4.3. Proofs for robotics

Robots are man-made artifacts where numerous design decisions can be argued based on logical or mathematical principles. For this reason, we wish to use this domain of application as a focus for our investigations. The questions for which we are close to providing answers involve precision issues in numeric computation, obstacle avoidance and motion planning (including questions of graph theory), articulated limb cinematics and dynamics, and balance and active control.

From the mathematical perspective, these topics require that we improve our library to cover real algebraic geometry, computational geometry, real analysis, graph theory, and refinement relations between abstract algorithms and executable programs.

In the long run, we hope to exhibit robots where pieces of software and part of the design has been subject to formal verification.

SUMO Project-Team

4. Application Domains

4.1. Smart transportation systems

The smart-city trend aims at optimizing all functions of future cities with the help of digital technologies. We focus on the segment of urban trains, which will evolve from static and scheduled offers to reactive and eventually on-demand transportation offers. We address two challenges in this field. The first one concerns the optimal design of robust subway lines. The idea is to be able to evaluate, at design time, the performance of time tables and of different regulation policies. In particular, we focus on robustness issues: how can small perturbations and incidents be accommodated by the system, how fast will return to normality occur, when does the system become unstable? The second challenge concerns the design of new robust regulation strategies to optimize delays, recovery times, and energy consumption at the scale of a full subway line. These problems involve large-scale discrete-event systems, with temporal and stochastic features, and translate into robustness assessment, stability analysis and joint numerical/combinatorial optimization problems on the trajectories of these systems.

4.2. Management of telecommunication networks and of data centers

Telecommunication-network management is a rich provider of research topics for the team, and some members of SUMO have a long background of contacts and transfer with industry in this domain. Networks are typical examples of large distributed dynamic systems, and their management raises numerous problems ranging from diagnosis (or root-cause analysis), to optimization, reconfiguration, provisioning, planning, verification, etc. They also bring new challenges to the community, for example on the modeling side: building or learning a network model is a complex task, specifically because these models should reflect features like the layering, the multi-resolution view of components, the description of both functions, protocols and configuration, and they should also reflect dynamically-changing architectures. Besides modeling, management algorithms are also challenged by features like the size of systems, the need to work on abstractions, on partial models, on open systems, etc. The networking technology is now evolving toward software-defined networks, virtualized-network functions, multi-tenant systems, etc., which reinforces the need for more automation in the management of such systems.

Data centers are another example of large-scale modular dynamic and reconfigurable systems: they are composed of thousands of servers, on which virtual machines are activated, migrated, resized, etc. Their management covers issues like troubleshooting, reconfiguration, optimal control, in a setting where failures are frequent and mitigated by the performance of the management plane. We have a solid background in the coordination of the various autonomic managers that supervise the different functions/layers of such systems (hardware, middleware, web services, ...) Virtualization technologies now reach the domain of networking, and telecommunication operators/vendors evolve towards providers of distributed open clouds. This convergence of IT and networking strongly calls for new management paradigms, which is an opportunity for the team.

4.3. Collaborative workflows

A current trend is to involve end-users in collection and analysis of data. Examples of this trend are contributive science, crisis-management systems, and crowd sourcing applications. All these applications are data-centric and user-driven. They are often distributed and involve complex, and sometimes dynamic workflows. In many cases, there are strong interactions between data and control flows: indeed, decisions taken regarding the next tasks to be launched highly depend on collected data. For instance, in an epidemic-surveillance system, the aggregation of various reported disease cases may trigger alerts. Another example is crowd sourcing applications where user skills are used to complete tasks that are better performed by humans than computers.

In return, this requires addressing imprecise and sometimes unreliable answers. We address several issues related to complex workflows and data. We study declarative and dynamic models that can handle workflows, data, uncertainty, and competence management.

Once these models are mature enough, we plan to build prototypes to experiment them on real use cases from contributive science, health-management systems, and crowd sourcing applications. We also plan to define abstraction schemes allowing formal reasoning on these systems.

TAMIS Project-Team (section vide)

TEA Project-Team

4. Application Domains

4.1. Automotive and Avionics

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. Yet, the research focuses on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, 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.

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

Together with the importance of open-source software, one of these projects, the FUI Project P (section 8), demonstrated that a centralized 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 centralized 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 modeler on top of SCADE, which uses the more model-engineering flavored 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 synthesized, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, based on the data-flow language Signal and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This ‘hardware in the loop’ simulator was used to test equipments supplied by other manufacturers for virtual prototyping of cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

4.2. Factory Automation

In collaboration with Mitsubishi R&D, we explore another application domain where time and domain heterogeneity are prime concerns: factory automation. In factory automation alone, a system is conventionally built from generic computing modules: PLCs (Programmable Logic Controllers), connected to the environment with actuators and detectors, and linked to a distributed network. Each individual, physically distributed, PLC module must be timely programmed to perform individually coherent actions and fulfill the global physical, chemical, safety, power efficiency, performance and latency requirements of the whole production chain. Factory chains are subject to global and heterogeneous (physical, electronic, functional) requirements whose enforcement must be orchestrated for all individual components.

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electro-mechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either the software or chemistry experts. As a consequence, cross-domain analysis issues are discovered very late during system integration and lead to costly fixes. This is particularly prevalent in multi-tier industries, such as avionic, automotive, factories, where systems are prominently integrated from independently-developed parts.

TOCCATA Project-Team

4. Application Domains

4.1. Safety-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. Currently our industrial collaborations or impact mainly belong to the domain of transportation: aerospace, aviation, railway, automotive.

Transfer to aeronautics: Airbus France Development of the control software of Airbus planes historically includes advanced usage of formal methods. A first aspect is the usage of the CompCert verified compiler for compiling C source code. Our work in cooperation with Gallium team for the safe compilation of floating-point arithmetic operations [2] is directly in application in this context. A second aspect is the usage of the Frama-C environment for static analysis to verify the C source code. In this context, both our tools Why3 and Alt-Ergo are indirectly used to verify C code.

Transfer to the community of Atelier B In the former ANR project BWare, we investigated 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 <https://www.atelierb.eu/en/>. The transfer effort continues nowadays through the FUI project LCHIP.

ProofInUse joint lab: transfer to the community of Ada development Through the creation of the ProofInUse joint lab (<https://www.adacore.com/proofinuse>) in 2014, with AdaCore company (<https://www.adacore.com/>), we have a growing impact on the community of industrial development of safety-critical applications written in Ada. See the web page <https://www.adacore.com/industries> for an overview of AdaCore's customer projects, in particular those involving the use of the SPARK Pro tool set. This impact involves both the use of Why3 for generating VCs on Ada source codes, and the use of Alt-Ergo for performing proofs of those VCs.

The impact of ProofInUse can also be measured in term of job creation: the first two ProofInUse engineers, D. Hauzar and C. Fumex, employed initially on Inria temporary positions, have now been hired on permanent positions in AdaCore company. It is also interesting to notice that this effort allowed AdaCore company to get new customers, in particular the domains of application of deductive formal verification went beyond the historical domain of aerospace: application in automotive (<https://www.adacore.com/customers/toyota-itc-japan>) cyber-security (<https://www.adacore.com/customers/multi-level-security-workstation>), health (artificial heart, <https://www.adacore.com/customers/total-artificial-heart>).

Extension of ProofInUse joint lab The current plans for continuation of the ProofInUse joint lab (<https://why3.gitlabpages.inria.fr/proofinuse/>) include extension at a larger perimeter than Ada applications. We started to collaborate with the TrustInSoft company (<https://trust-in-soft.com/>) for the verification of C and C++ codes, including the use of Why3 to design verified and reusable C libraries (ongoing CIFRE PhD thesis). We also started to collaborate with Mitsubishi Electric in Rennes (<http://www.mitsubishielectric-rce.eu/xindex.php>) for a specific usage of Why3 for verifying embedded devices (logic controllers).

Generally speaking, we believe that our increasing industrial impact is a representative success for our general goal of spreading deductive verification methods to a larger audience, and we are firmly engaged into continuing such kind of actions in the future.

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 the verification of algorithms that are developed for supporting novel computing paradigms, including ad-hoc networks that underly mobile and low-power computing or overlay networks, peer-to-peer networks 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 we work together with domain experts on designing formal models of these protocols, and on verifying their properties. These cooperations help us focus on concrete algorithms and ensure that our work is relevant to the distributed algorithm community.

Our work on symbolic procedures for solving polynomial constraints finds applications beyond verification. In particular, we have been working in interdisciplinary projects with researchers from mathematics, computer science, system biology, and system medicine on the analysis of molecular interaction networks in order to infer the principal qualitative properties of models. Our techniques complement numerical analysis techniques and are validated against collections of models from computational biology.

ACUMES Project-Team

4. Application Domains

4.1. Active flow control for vehicles

The reduction of CO₂ emissions represents a great challenge for the automotive and aeronautic industries, which committed respectively a decrease of 20% for 2020 and 75% for 2050. This goal will not be reachable, unless a significant improvement of the aerodynamic performance of cars and aircrafts is achieved (e.g. aerodynamic resistance represents 70% of energy losses for cars above 90 km/h). Since vehicle design cannot be significantly modified, due to marketing or structural reasons, active flow control technologies are one of the most promising approaches to improve aerodynamic performance. This consists in introducing micro-devices, like pulsating jets or vibrating membranes, that can modify vortices generated by vehicles. Thanks to flow non-linearities, a small energy expense for actuation can significantly reduce energy losses. The efficiency of this approach has been demonstrated, experimentally as well as numerically, for simple configurations [167].

However, the lack of efficient and flexible numerical tools, that allow to simulate and optimize a large number of such devices on realistic configurations, is still a bottleneck for the emergence of this technology in industry. The main issue is the necessity of using high-order schemes and complex models to simulate actuated flows, accounting for phenomena occurring at different scales. In this context, we intend to contribute to the following research axes:

- *Sensitivity analysis for actuated flows.* Adjoint-based (reverse) approaches, classically employed in design optimization procedure to compute functional gradients, are not well suited to this context. Therefore, we propose to explore the alternative (direct) formulation, which is not so much used, in the perspective of a better characterization of actuated flows and optimization of control devices.
- *Isogeometric simulation of control devices.* To simulate flows perturbed by small-scale actuators, we investigate the use of isogeometric analysis methods, which allow to account exactly for CAD-based geometries in a high-order hierarchical representation framework. In particular, we try to exploit the features of the method to simulate more accurately complex flows including moving devices and multiscale phenomena.

4.2. Vehicular and pedestrian traffic flows

Intelligent Transportation Systems (ITS) is nowadays a booming sector, where the contribution of mathematical modeling and optimization is widely recognized. In this perspective, traffic flow models are a commonly cited example of "complex systems", in which individual behavior and self-organization phenomena must be taken into account to obtain a realistic description of the observed macroscopic dynamics [125]. Further improvements require more advanced models, keeping into better account interactions at the microscopic scale, and adapted control techniques, see [71] and references therein. In particular, we will focus on the following aspects:

- *Junction models.* We are interested in designing a general junction model both satisfying basic analytical properties guaranteeing well-posedness and being realistic for traffic applications. In particular, the model should be able to overcome severe drawbacks of existing models, such as restrictions on the number of involved roads and prescribed split ratios [86], [114], which limit their applicability to real world situations. Hamilton-Jacobi equations could be also an interesting direction of research, following the recent results obtained in [130].
- *Data assimilation.* In traffic flow modeling, the capability of correctly estimating and predicting the state of the system depends on the availability of rich and accurate data on the network. Up to now, the most classical sensors are fixed ones. They are composed of inductive loops (electrical wires) that are installed at different spatial positions of the network and that can measure the traffic flow,

the occupancy rate (i.e. the proportion of time during which a vehicle is detected to be over the loop) and the speed (in case of a system of two distant loops). These data are useful / essential to calibrate the phenomenological relationship between flow and density which is known in the traffic literature as the Fundamental Diagram. Nowadays, thanks to the wide development of mobile internet and geolocalization techniques and its increasing adoption by the road users, smartphones have turned into perfect mobile sensors in many domains, including in traffic flow management. They can provide the research community with a large database of individual trajectory sets that are known as Floating Car Data (FCD), see [127] for a real field experiment. Classical macroscopic models, say (hyperbolic systems of) conservation laws, are not designed to take into account this new kind of microscopic data. Other formulations, like Hamilton-Jacobi partial differential equations, are most suited and have been intensively studied in the past five years (see [79], [80]), with a stress on the (fixed) Eulerian framework. Up to our knowledge, there exist a few studies in the time-Lagrangian as well as space-Lagrangian frameworks, where data coming from mobile sensors could be easily assimilated, due to the fact that the Lagrangian coordinate (say the label of a vehicle) is fixed.

- *Control of autonomous vehicles.* Traffic flow is usually controlled via traffic lights or variable speed limits, which have fixed space locations. The deployment of autonomous vehicles opens new perspectives in traffic management, as the use of a small fraction of cars to optimize the overall traffic. In this perspective, the possibility to track vehicles trajectories either by coupled micro-macro models [94], [115] or via the Hamilton-Jacobi approach [79], [80] could allow to optimize the flow by controlling some specific vehicles corresponding to internal conditions.

4.3. Virtual Fractional Flow Reserve in coronary stenting

Atherosclerosis is a chronic inflammatory disease that affects the entire arterial network and especially the coronary arteries. It is an accumulation of lipids over the arterial surface due to a dysfunction of this latter. The objective of clinical intervention, in this case, is to establish a revascularization using different angioplasty techniques, among which the implantation of stents is the most widespread. This intervention involves introducing a stent into the damaged portion in order to allow the blood to circulate in a normal way over all the vessels. Revascularization is based on the principle of remedying ischemia, which is a decrease or an interruption of the supply of oxygen to the various organs. This anomaly is attenuated by the presence of several lesions (multivessel disease patients), which can lead to several complications. The key of a good medical intervention is the fact of establishing a good diagnosis, in order to decide which lesion requires to be treated. In the diagnosis phase, the clinician uses several techniques, among which angiography is the most popular. Angiography is an X-ray technique to show the inside (the lumen) of blood vessels, in order to identify vessel narrowing: stenosis. Despite its widespread use, angiography is often imperfect in determining the physiological significance of coronary stenosis. If the problem remains simple for non significant lesions ($\leq 40\%$) or very severe ($\geq 70\%$), a very important category of intermediate lesions must benefit from a functional evaluation which will determine the strategy of treatment [90].

The technique of the Fractional Flow Reserve (FFR) has derived from the initial coronary physical approaches decades ago. Since then, many studies have demonstrated its effectiveness in improving the patients prognosis, by applying the appropriate approach. Its contribution in the reduction of mortality was statistically proved by the FAME (Fractional Flow Reserve Versus Angiography for Multivessel Evaluation) study [169]. It is established that the FFR can be easily measured during coronary angiography by calculating the ratio of distal coronary pressure P_d to aortic pressure P_a . These pressures are measured simultaneously with a special guide-wire. FFR in a normal coronary artery equals to 1.0. FFR value of 0.80 or less identifies ischemia-causing coronary lesions with an accuracy of more than 90% [169].

Obviously, from an interventional point of view, the FFR is binding since it is invasive. It should also be noted that this technique induces additional costs, which are not covered by insurances in several countries. For these reasons, it is used only in less than 10% of the cases.

In this perspective, a new virtual version of the FFR, entitled VFFR, has emerged as an attractive and non-invasive alternative to standard FFR, see [161], [144]. VFFR is based on computational modeling, mainly fluid and fluid-structural dynamics. However, there are key scientific, logistic and commercial challenges that need to be overcome before VFFR can be translated into routine clinical practice.

While most of the studies related to VFFR use Navier-Stokes models, we focus on the non-newtonian case, starting with a generalized fluid flow approach. These models are more relevant for the coronary arteries, and we expect that the computation of the FFR should then be more accurate. We are also leading numerical studies to assess the impact (on the FFR) of the interaction of the physical devices (catheter, optical captors, spheroids) with the blood flow.

4.4. Other application fields

Besides the above mentioned axes, which constitute the project's identity, the methodological tools described in Section have a wider range of application. We currently carry on also the following research actions, in collaboration with external partners.

- **Modeling cell dynamics.** Migration and proliferation of epithelial cell sheets are the two keystone aspects of the collective cell dynamics in most biological processes such as morphogenesis, embryogenesis, cancer and wound healing. It is then of utmost importance to understand their underlying mechanisms.

Semilinear reaction-diffusion equations are widely used to give a phenomenological description of the temporal and spatial changes occurring within cell populations that undergo scattering (moving), spreading (expanding cell surface) and proliferation. We have followed the same methodology and contributed to assess the validity of such approaches in different settings (cell sheets [122], dorsal closure [58], actin organization [57]). However, epithelial cell-sheet movement is complex enough to undermine most of the mathematical approaches based on *locality*, that is mainly traveling wavefront-like partial differential equations. In [109] it is shown that Madin-Darby Canine Kidney (MDCK) cells extend cryptic lamellipodia to drive the migration, several rows behind the wound edge. In [149] MDCK monolayers are shown to exhibit similar non-local behavior (long range velocity fields, very active border-localized leader cells).

Our aim is to start from a mesoscopic description of cell interaction: considering cells as independent anonymous agents, we plan to investigate the use of mathematical techniques adapted from the mean-field game theory. Otherwise, looking at them as interacting particles, we will use a multi-agent approach (at least for the actin dynamics). We intend also to consider approaches stemming from compartment-based simulation in the spirit of those developed in [106], [111], [113].

- **Game strategies for thermoelastography.** Thermoelastography is an innovative non-invasive control technology, which has numerous advantages over other techniques, notably in medical imaging [142]. Indeed, it is well known that most pathological changes are associated with changes in tissue stiffness, while remaining isoechoic, and hence difficult to detect by ultrasound techniques. Based on elastic waves and heat flux reconstruction, thermoelastography shows no destructive or aggressive medical sequel, unlike X-ray and comparables techniques, making it a potentially prominent choice for patients.

Physical principles of thermoelastography originally rely on dynamical structural responses of tissues, but as a first approach, we only consider static responses of linear elastic structures.

The mathematical formulation of the thermoelasticity reconstruction is based on data completion and material identification, making it a harsh ill posed inverse problem. In previous works [123], [132], we have demonstrated that Nash game approaches are efficient to tackle ill-posedness. We intend to extend the results obtained for Laplace equations in [123], and the algorithms developed in Section 3.1.2.4 to the following problems (of increasing difficulty):

- Simultaneous data and parameter recovery in linear elasticity, using the so-called Kohn and Vogelius functional (ongoing work, some promising results obtained).

- Data recovery in coupled heat-thermoelasticity systems.
- Data recovery in linear thermoelasticity under stochastic heat flux, where the imposed flux is stochastic.
- Data recovery in coupled heat-thermoelasticity systems under stochastic heat flux, formulated as an incomplete information Nash game.
- Application to robust identification of cracks.

- **Constraint elimination in Quasi-Newton methods.** In single-objective differentiable optimization, Newton's method requires the specification of both gradient and Hessian. As a result, the convergence is quadratic, and Newton's method is often considered as the target reference. However, in applications to distributed systems, the functions to be minimized are usually "functionals", which depend on the optimization variables by the solution of an often complex set of PDE's, through a chain of computational procedures. Hence, the exact calculation of the full Hessian becomes a complex and costly computational endeavor.

This has fostered the development of *quasi-Newton's methods* that mimic Newton's method but use only the gradient, the Hessian being iteratively constructed by successive approximations inside the algorithm itself. Among such methods, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is well-known and commonly employed. In this method, the Hessian is corrected at each new iteration by rank-one matrices defined from several evaluations of the gradient only. The BFGS method has "super-linear convergence".

For constrained problems, certain authors have developed so-called *Riemannian BFGS*, e.g. [152], that have the desirable convergence property in constrained problems. However, in this approach, the constraints are assumed to be known formally, by explicit expressions.

In collaboration with ONERA-Meudon, we are exploring the possibility of representing constraints, in successive iterations, through local approximations of the constraint surfaces, splitting the design space locally into tangent and normal sub-spaces, and eliminating the normal coordinates through a linearization, or more generally a finite expansion, and applying the BFGS method through dependencies on the coordinates in the tangent subspace only. Preliminary experiments on the difficult Rosenbrock test-case, although in low dimensions, demonstrate the feasibility of this approach. On-going research is on theorizing this method, and testing cases of higher dimensions.

- **Multi-objective optimization for nanotechnologies.** Our team takes part in a larger collaboration with CEA/LETI (Grenoble), initiated by the Inria Project-Team Nachos, and related to the Maxwell equations. Our component in this activity relates to the optimization of nanophotonic devices, in particular with respect to the control of thermal loads. We have first identified a gradation of representative test-cases of increasing complexity:
 - infrared micro-source;
 - micro-photoacoustic cell;
 - nanophotonic device.

These cases involve from a few geometric parameters to be optimized to a functional minimization subject to a finite-element solution involving a large number of dof's. CEA disposes of such codes, but considering the computational cost of the objective functions in the complex cases, the first part of our study is focused on the construction and validation of meta-models, typically of RBF-type. Multi-objective optimization will be carried out subsequently by MGDA, and possibly Nash games.

BONUS Project-Team

4. Application Domains

4.1. Introduction

For the validation of our findings we obviously use standard benchmarks to facilitate the comparison with related works. In addition, we also target real-world applications in the context of our collaborations and industrial contracts. From the *application* point of view two classes are targeted: *complex scheduling* and *engineering design*. The objective is twofold: proposing new models for complex problems and solving efficiently BOPs using jointly the three lines of our research program. In the following, are given some use cases that are the focus of our current industrial collaborations.

4.2. Big optimization for complex scheduling

Three application domains are targeted: energy, health and transport and logistics. In the **energy** field, with the smart grid revolution (multi-)house energy management is gaining a growing interest. The key challenge is to make elastic with respect to the energy market the (multi-)house energy consumption and management. *This kind of demand-side management will be of strategic importance for energy companies in the near future.* In collaboration with the EDF energy company we are working on the formulation and solving of optimization problems on demand-side management in smart micro-grids for single- and multi-user frameworks. These complex problems require taking into account multiple conflicting objectives and constraints and many (deterministic/uncertain, discrete/continuous) parameters. A representative example of such BOPs that we are addressing is the scheduling of the activation of a large number of electrical and thermal appliances for a set of homes optimizing at least three criteria: maximizing the user's confort, minimizing its energy bill and minimizing peak consumption situations. In the **health** care domain, we are collaborating with the Beckman & Coulter company on the design and planning of large medical laboratories. This is a hot topic resulting from the mutualisation phenomenon which makes these laboratories bigger. As a consequence, being responsible for analyzing medical tests ordered by physicians on patient's samples, these laboratories receive large amounts of prescriptions and tubes making their associated workflow more complex. Our aim is therefore to design and plan any medical laboratory to minimize the costs and time required to perform the tests. More exactly, the focus is put on the multi-objective modeling and solving of large (e.g. dozens of thousands of medical test tubes to be analyzed) strategic, tactical and operational problems such as the layout design, machine selection and configuration, assignment and scheduling. Finally, in **transport and logistics**, within the context of our potential collaboration (being set up) with the EXOTEC company we target the optimization of the robotic logistics of 3D warehouses. More exactly, the problem consists in efficient complex scheduling without collision of thousands of missions realized by a fleet of dozens of robots and several operators in a 3D logistics warehouse. The problem is identified in the literature as the parts-to-picker based order processing in a rack-moving mobile robots environment.

4.3. Big optimization for engineering design

The focus is for now put on the aerospace vehicle design, a complex multidisciplinary optimization process, we are exploring in collaboration with ONERA. The objective is to find the vehicle architecture and characteristics that provide the optimal performance (flight performance, safety, reliability, cost etc.) while satisfying design requirements [39]. A representative topic we are investigating, and will continue to investigate throughout the lifetime of the project given its complexity, is the design of launch vehicles that involves at least four tightly coupled disciplines (aerodynamics, structure, propulsion and trajectory). Each discipline may rely on time-demanding simulations such as Finite Element analyses (structure) and Computational Fluid Dynamics analyses (aerodynamics). Surrogate-assisted optimization is highly required to reduce the time complexity. In addition, the problem is high-dimensional (dozens of parameters and more than three

objectives) requiring different decomposition schemas (coupling vs. local variables, continuous vs. discrete even categorial variables, scalarization of the objectives). Another major issue arising in this area is the non-stationarity of the objective functions which is generally due to the abrupt change of a physical property that often occurs in the design of launch vehicles. In the same spirit than deep learning using neural networks, we use Deep Gaussian Processes to deal with non-stationary multi-objective functions. Finally, the resolution of the problem using only one objective takes one week using a multi-core processor. Therefore, *in addition to surrogates ultra-scale computing is required at different levels to speed up the search and improve the reliability which is a major requirement in aerospace design.* This example shows that we need to use the synergy between the three lines of our research program to tackle such BOPs.

CAGE Project-Team

4. Application Domains

4.1. First axis: Geometry of vision

A suggestive application of sub-Riemannian geometry and in particular of hypoelliptic diffusion comes from a model of geometry of vision describing the functional architecture of the primary visual cortex V1. In 1958, Hubel and Wiesel (Nobel in 1981) observed that the visual cortex V1 is endowed with the so-called **pinwheel structure**, characterized by neurons grouped into orientation columns, that are sensible both to positions and directions [110]. The mathematical rephrasing of this discovery is that the visual cortex lifts an image from \mathbf{R}^2 into the bundle of directions of the plane [97], [136], [138], [109].

A simplified version of the model can be described as follows: neurons of V1 are grouped into orientation columns, each of them being sensitive to visual stimuli at a given point of the retina and for a given direction on it. The retina is modeled by the real plane, i.e., each point is represented by a pair $(x, y) \in \mathbf{R}^2$, while the directions at a given point are modeled by the projective line, i.e. an element θ of the projective line P^1 . Hence, the primary visual cortex V1 is modeled by the so called projective tangent bundle $\text{PTR}^2 = \mathbf{R}^2 \times \mathbf{P}^1$. From a neurological point of view, orientation columns are in turn grouped into hypercolumns, each of them being sensitive to stimuli at a given point (x, y) with any direction.

Orientation columns are connected between them in two different ways. The first kind of connections are the vertical (inhibitory) ones, which connect orientation columns belonging to the same hypercolumn and sensible to similar directions. The second kind of connections are the horizontal (excitatory) connections, which connect neurons belonging to different (but not too far) hypercolumns and sensible to the same directions. The resulting metric structure is sub-Riemannian and the model obtained in this way provides a convincing explanation in terms of sub-Riemannian geodesics of gestalt phenomena such as Kanizsa illusory contours.

The sub-Riemannian model for image representation of V1 has a great potential of yielding powerful bio-inspired image processing algorithms [104], [90]. Image inpainting, for instance, can be implemented by reconstructing an incomplete image by activating orientation columns in the missing regions in accordance with sub-Riemannian non-isotropic constraints. The process intrinsically defines an hypoelliptic heat equation on PTR^2 which can be integrated numerically using non-commutative Fourier analysis on a suitable semidiscretization of the group of roto-translations of the plane [88].

We have been working on the model and its software implementation since 2012. This work has been supported by several project, as the ERC starting grant GeCoMethods and the ERC Proof of Concept ARTIV1 of U. Boscain, and the ANR GCM.

A parallel approach that we will pursue and combine with this first one is based on **pattern matching in the group of diffeomorphisms**. We want to extend this approach, already explored in the Riemannian setting [151], [127], to the general sub-Riemannian framework. The paradigm of the approach is the following: consider a distortable object, more or less rigid, discretized into a certain number of points. One may track its distortion by considering the paths drawn by these points. One would however like to know how the object itself (and not its discretized version) has been distorted. The study in [151], [127] shed light on the importance of Riemannian geometry in this kind of problem. In particular, they study the Riemannian submersion obtained by making the group of diffeomorphisms act transitively on the manifold formed by the points of the discretization, minimizing a certain energy so as to take into account the whole object. Settled as such, the problem is Riemannian, but if one considers objects involving connections, or submitted to nonholonomic constraints, like in medical imaging where one tracks the motions of organs, then one comes up with a sub-Riemannian problem. The transitive group is then far bigger, and the aim is to lift curves submitted to these nonholonomic constraints into curves in the set of diffeomorphisms satisfying the corresponding constraints, in a unique way and minimizing an energy (giving rise to a sub-Riemannian structure).

4.2. Second axis: Quantum control

The goal of quantum control is to design efficient protocols for tuning the occupation probabilities of the energy levels of a system. This task is crucial in atomic and molecular physics, with applications ranging from photochemistry to nuclear magnetic resonance and quantum computing. A quantum system may be controlled by exciting it with one or several external fields, such as magnetic or electric fields. The goal of quantum control theory is to adapt the tools originally developed by control theory and to develop new specific strategies that tackle and exploit the features of quantum dynamics (probabilistic nature of wavefunctions and density operators, measure and wavefunction collapse, decoherence, ...). A rich variety of relevant models for controlled quantum dynamics exist, encompassing low-dimensional models (e.g., single-spin systems) and PDEs alike, with deterministic and stochastic components, making it a rich and exciting area of research in control theory.

The controllability of quantum system is a well-established topic when the state space is finite-dimensional [102], thanks to general controllability methods for left-invariant control systems on compact Lie groups [92], [112]. When the state space is infinite-dimensional, it is known that in general the bilinear Schrödinger equation is not exactly controllable [154]. Nevertheless, weaker controllability properties, such as approximate controllability or controllability between eigenstates of the internal Hamiltonian (which are the most relevant physical states), may hold. In certain cases, when the state space is a function space on a 1D manifold, some rather precise description of the set of reachable states has been provided [77]. A similar description for higher-dimensional manifolds seems intractable and at the moment only approximate controllability results are available [128], [134], [113]. The most widely applicable tests for controllability of quantum systems in infinite-dimensional Hilbert spaces are based on the **Lie–Galerkin technique** [96], [85], [86]. They allow, in particular, to show that the controllability property is generic among this class of systems [125].

A family of algorithms which are specific to quantum systems are those based on adiabatic evolution [158], [157], [116]. The basic principle of adiabatic control is that the flow of a slowly varying Hamiltonian can be approximated (up to a phase factor) by a quasi-static evolution, with a precision proportional to the velocity of variation of the Hamiltonian. The advantage of the **adiabatic approach** is that it is constructive and produces control laws which are both smooth and robust to parameter uncertainty. The paradigm is based on the adiabatic perturbation theory developed in mathematical physics [83], [133], [150], where it plays an important role for understanding molecular dynamics. Approximation theory by adiabatic perturbation can be used to describe the evolution of the occupation probabilities of the energy levels of a slowly varying Hamiltonian. Results from the last 15 years, including those by members of our team [64], [89], have highlighted the effectiveness of control techniques based on adiabatic path following.

4.3. Third axis: Stability and uncertain dynamics

Switched and hybrid systems constitute a broad framework for the description of the heterogeneous aspects of systems in which continuous dynamics (typically pertaining to physical quantities) interact with discrete/logical components. The development of the switched and hybrid paradigm has been motivated by a broad range of applications, including automotive and transportation industry [142], energy management [135] and congestion control [126].

Even if both controllability [146] and observability [114] of switched and hybrid systems have attracted much research efforts, the central role in their study is played by the problem of stability and stabilizability. The goal is to determine whether a dynamical or a control system whose evolution is influenced by a time-dependent signal is uniformly stable or can be uniformly stabilized [119], [147]. Uniformity is considered with respect to all signals in a given class. Stability of switched systems lead to several interesting phenomena. For example, even when all the subsystems corresponding to a constant switching law are exponentially stable, the switched systems may have divergent trajectories for certain switching signals [118]. This fact illustrates the fact that stability of switched systems depends not only on the dynamics of each subsystem but also on the properties of the class of switching signals which is considered.

The most common class of switching signals which has been considered in the literature is made of all piecewise constant signals. In this case uniform stability of the system is equivalent to the existence of a common quadratic Lyapunov function [129]. Moreover, provided that the system has finitely many modes, the Lyapunov function can be taken polyhedral or polynomial [80], [81], [103]. 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 the surveys [121], [141] and the references therein). It is known, however, that the existence of a common quadratic Lyapunov function is not necessary for the global uniform exponential stability of a linear switched system with finitely many modes. Moreover, there exists no uniform upper bound on the minimal degree of a common polynomial Lyapunov function [124]. More refined tools rely on multiple and non-monotone Lyapunov functions [91]. Let us also mention linear switched systems technics based on the analysis of the Lie algebra generated by the matrices corresponding to the modes of the system [67].

For systems evolving in the plane, more geometrical tests apply, and yield a complete characterization of the stability [84], [73]. Such a geometric approach also yields sufficient conditions for uniform stability in the linear planar case [87].

In many situations, it is interesting for modeling purposes to specify the features of the switched system by introducing **constrained switching rules**. A typical constraint is that each mode is activated for at least a fixed minimal amount of time, called the dwell-time. Switching rules can also be imposed, for instance, by a timed automata. When constraints apply, the common Lyapunov function approach becomes conservative and new tools have to be developed to give more detailed characterizations of stable and unstable systems.

Our approach to constrained switching is based on the idea of relating the analytical properties of the classes of constrained switching laws (shift-invariance, compactness, closure under concatenation, ...) to the stability behavior of the corresponding switched systems. One can introduce **probabilistic uncertainties** by endowing the classes of admissible signals with suitable probability measures. One then looks at the corresponding Lyapunov exponents, whose existence is established by the multiplicative ergodic theorem. The interest of this approach is that probabilistic stability analysis filters out highly ‘exceptional’ worst-case trajectories. Although less explicitly characterized from a dynamical viewpoint than its deterministic counterpart, the probabilistic notion of uniform exponential stability can be studied using several reformulations of Lyapunov exponents proposed in the literature [78], [98], [155].

4.4. Joint theoretical core

The theoretical questions raised by the different applicative area will be pooled in a research axis on the transversal aspects of geometric control theory and sub-Riemannian structures.

We recall that sub-Riemannian geometry is a generalization of Riemannian geometry, whose birth dates back to Carathéodory’s seminal paper on the foundations of Carnot thermodynamics [94], followed by E. Cartan’s address at the International Congress of Mathematicians in Bologna [95]. In the last twenty years, sub-Riemannian geometry has emerged as an independent research domain, with a variety of motivations and ramifications in several parts of pure and applied mathematics. Let us mention geometric analysis, geometric measure theory, stochastic calculus and evolution equations together with applications in mechanics and optimal control (motion planning, robotics, nonholonomic mechanics, quantum control) [62], [63].

One of the main open problems in sub-Riemannian geometry concerns the regularity of length-minimizers [65], [131]. Length-minimizers are solutions to a variational problem with constraints and satisfy a first-order necessary condition resulting from the Pontryagin Maximum Principle (PMP). Solutions of the PMP are either *normal* or *abnormal*. Normal length-minimizers are well-known to be smooth, i.e., C^∞ , as it follows by the Hamiltonian nature of the PMP. The question of regularity is then reduced to abnormal length-minimizers. If the sub-Riemannian structure has step 2, then abnormal length-minimizers can be excluded and thus every length-minimizer is smooth. For step 3 structures, the situation is already more complicated and smoothness of length-minimizers is known only for Carnot groups [115], [149]. The question of regularity of length-minimizers is not restricted to the smoothness in the C^∞ sense. A recent result prove that length-minimizers, for sub-Riemannian structures of any step, cannot have corner-like singularities [108]. When the

sub-Riemannian structure is analytic, more is known on the size of the set of points where a length-minimizer can lose analyticity [148], regardless of the rank and of the step of the distribution.

An interesting set of recent results in sub-Riemannian geometry concerns the extension to such a setting of the Riemannian notion of sectional curvature. The curvature operator can be introduced in terms of the symplectic invariants of the Jacobi curve [69], [117], [66], a curve in the Lagrange Grassmannian related to the linearization of the Hamiltonian flow. Alternative approaches to curvatures in metric spaces are based either on the associated heat equation and the generalization of the curvature-dimension inequality [74], [75] or on optimal transport and the generalization of Ricci curvature [145], [144], [123], [71].

CAGIRE Project-Team

4. Application Domains

4.1. Aeronautics

Cagire is presently involved in studies mainly related to:

- The combustion chamber wall: the modelling, the simulation and the experimentation of the flow around a multiperforated plate representative of a real combustion chamber wall are the three axes we have been developing during the recent period. The continuous improvement of our in-house test facility Maveric is also an important ingredient to produce our own experimental validation data for isothermal flows. For non-isothermal flows, our participation in the EU funded program Soprano will be giving us access to non-isothermal data produced by Onera.
- The flow around airfoils: the modelling of the turbulent boundary layer has been for almost a century a key issue in the aeronautics industry. However, even the more advanced RANS models face difficulties in predicting the influence of pressure gradients on the development of the boundary layer. A main issue is the reliability of the modelling hypotheses, which is crucial for less conservative design. One of the technological barriers is the prediction of the flow in regimes close to the edge of the flight domain (stall, buffeting, unsteady loads) when the boundary layer is slowed down by an adverse pressure gradient. This is the subject of the CIFRE PhD thesis of Gustave Sporschill, started in 2018, in collaboration with Dassault Aviation.

4.2. Power stations

R. Manceau has established a long term collaboration (4 CIFRE PhD theses in the past, 2 ongoing) with the R & D center of EDF of Chatou, for the development of refined turbulence models in the in-house CFD code of EDF, Code_Saturne :

- The prediction of heat transfer in fluid and solid components is of major importance in power stations, in particular, nuclear power plants. Either for the thermohydraulics of the plenum or in the study of accidental scenarii, among others, the accurate estimation of wall heat transfer, mean temperatures and temperature fluctuations are necessary for the evaluation of relevant thermal and mechanical design criteria. The PhD thesis (CIFRE EDF) of G. Mangeon is dedicated to the development of relevant RANS models for these industrial applications.
- Moreover, the prediction of unsteady hydrodynamic loadings is a key point for operating and for safety studies of PWR power plants. Currently, the static loading is correctly predicted by RANS computations but when the flow is transient (as, for instance, in Reactor Coolant Pumps, due to rotor/stator interactions, or during operating transients) or in the presence of large, energetic, coherent structures in the external flow region, the RANS approach is not sufficient, whereas LES is still too costly for a wide use in industry. This issue constitutes the starting point of the just-started PhD thesis (CIFRE EDF) of Vladimir Duffal.

4.3. Automotive propulsion

- The engine (underhood) compartment is a key component of vehicle design, in which the temperature is monitored to ensure the effectiveness and safety of the vehicle, and participates in 5 to 8% of the total drag and CO₂ emissions. Dimensioning is an aerodynamic and arothermal compromise, validated on a succession of road stages at constant speed and stopped phases (red lights, tolls, traffic jam). Although CFD is routinely used for forced convection, state-of-the-art turbulence models are not able to reproduce flows dominated by natural convection during stopped phases, with a Rayleigh

number of the order of 10^{10} , such that the design still relies on costly, full-scale, wind tunnel experiments. This technical barrier must be lifted, since the ambition of the PSA group is to reach a *full digital design of their vehicles in the 2025 horizon*, i.e., to almost entirely rely on CFD. This issue is the focus of the ongoing PhD thesis (CIFRE PSA) of S. Jameel, supervised by R. Manceau, and also a part of the ANR project MONACO_2025 described in section 9.2.2 .

- The Power & Vehicles Division of IFPEN co-develops a CFD code to simulate the internal flow in a spark-ignition engine, in order to provide the automotive industry with tools to optimize the design of combustion engines. The RANS method, widely used in the industry, is not sufficiently reliable for quantitative predictions, and is only used as a tool to qualitatively compare different geometries. On the other hand, LES provides more detailed and accurate information, but at the price of a CPU cost unaffordable for daily use in the industry. Therefore, IFPEN aims at developing the hybrid RANS/LES methodology, in order to combine the strengths of the two approaches. The PhD thesis of Hassan Afaïlal, co-supervised by Rémi Manceau, is focused on this issue.

CARDAMOM Project-Team

4. Application Domains

4.1. De-anti icing systems

Impact of large ice debris on downstream aerodynamic surfaces and ingestion by aft mounted engines must be considered during the aircraft certification process. It is typically the result of ice accumulation on unprotected surfaces, ice accretions downstream of ice protected areas, or ice growth on surfaces due to delayed activation of ice protection systems (IPS) or IPS failure. This raises the need for accurate ice trajectory simulation tools to support pre-design, design and certification phases while improving cost efficiency. Present ice trajectory simulation tools have limited capabilities due to the lack of appropriate experimental aerodynamic force and moment data for ice fragments and the large number of variables that can affect the trajectories of ice particles in the aircraft flow field like the shape, size, mass, initial velocity, shedding location, etc... There are generally two types of model used to track shed ice pieces. The first type of model makes the assumption that ice pieces do not significantly affect the flow. The second type of model intends to take into account ice pieces interacting with the flow. We are concerned with the second type of models, involving fully coupled time-accurate aerodynamic and flight mechanics simulations, and thus requiring the use of high efficiency adaptive tools, and possibly tools allowing to easily track moving objects in the flow. We will in particular pursue and enhance our initial work based on adaptive immersed boundary capturing of moving ice debris, whose movements are computed using basic mechanical laws.

In [31] it has been proposed to model ice shedding trajectories by an innovative paradigm that is based on Cartesian grids, PEnalization and LLevel Sets (LESCAPE code). Our objective is to use the potential of high order unstructured mesh adaptation and immersed boundary techniques to provide a geometrically flexible extension of this idea. These activities will be linked to the development of efficient mesh adaptation and time stepping techniques for time dependent flows, and their coupling with the immersed boundary methods we started developing in the FP7 EU project STORM [24], [76]. In these methods we compensate for the error at solid walls introduced by the penalization by using anisotropic mesh adaptation [50], [68], [69]. From the numerical point of view one of the major challenges is to guarantee efficiency and accuracy of the time stepping in presence of highly stretched adaptive and moving meshes. Semi-implicit, locally implicit, multi-level, and split discretizations will be explored to this end.

Besides the numerical aspects, we will deal with modelling challenges. One source of complexity is the initial conditions which are essential to compute ice shedding trajectories. It is thus extremely important to understand the mechanisms of ice release. With the development of next generations of engines and aircraft, there is a crucial need to better assess and predict icing aspects early in design phases and identify breakthrough technologies for ice protection systems compatible with future architectures. When a thermal ice protection system is activated, it melts a part of the ice in contact with the surface, creating a liquid water film and therefore lowering ability of the ice block to adhere to the surface. The aerodynamic forces are then able to detach the ice block from the surface [33]. In order to assess the performance of such a system, it is essential to understand the mechanisms by which the aerodynamic forces manage to detach the ice. The current state of the art in icing codes is an empirical criterion. However such an empirical criterion is unsatisfactory. Following the early work of [34], [30] we will develop appropriate asymptotic PDE approximations allowing to describe the ice formation and detachment, trying to embed in this description elements from damage/fracture mechanics. These models will constitute closures for aerodynamics/RANS and URANS simulations in the form of PDE wall models, or modified boundary conditions.

In addition to this, several sources of uncertainties are associated to the ice geometry, size, orientation and the shedding location. In very few papers [78], some sensitivity analysis based on Monte Carlo method have been conducted to take into account the uncertainties of the initial conditions and the chaotic nature of the ice particle motion. We aim to propose some systematic approach to handle every source of uncertainty in an efficient way relying on some state-of-art techniques developed in the Team. In particular, we will perform an uncertainty propagation of some uncertainties on the initial conditions (position, orientation, velocity,...) through a low-fidelity model in order to get statistics of a multitude of particle tracks. This study will be done in collaboration with ETS (Ecole de Technologies Supérieure, Canada). The longterm objective is to produce footprint maps and to analyse the sensitivity of the models developed.

4.2. Energy

We will develop modelling and design tools, as well as dedicated platforms, for Rankine cycles using complex fluids (organic compounds), and for wave energy extraction systems.

Organic Rankine Cycles (ORCs) use heavy organic compounds as working fluids. This results in superior efficiency over steam Rankine cycles for source temperatures below 900 K. ORCs typically require only a single-stage rotating component making them much simpler than typical multi-stage steam turbines. The strong pressure reduction in the turbine may lead to supersonic flows in the rotor, and thus to the appearance of shocks, which reduces the efficiency due to the associated losses. To avoid this, either a larger multi stage installation is used, in which smaller pressure drops are obtained in each stage, or centripetal turbines are used, at very high rotation speeds (of the order of 25,000 rpm). The second solution allows to keep the simplicity of the expander, but leads to poor turbine efficiencies (60-80%) - w.r.t. modern, highly optimized, steam and gas turbines - and to higher mechanical constraints. The use of *dense-gas working fluids*, *i.e.* operating close to the saturation curve, in properly chosen conditions could increase the turbine critical Mach number avoiding the formation of shocks, and increasing the efficiency. Specific shape optimization may enhance these effects, possibly allowing the reduction of rotation speeds. However, dense gases may have significantly different properties with respect to dilute ones. Their dynamics is governed by a thermodynamic parameter known as the fundamental derivative of gas dynamics

$$\Gamma = 1 + \frac{\rho}{c} \left(\frac{\partial c}{\partial \rho} \right)_s, \quad (1)$$

where ρ is the density, c is the speed of sound and s is the entropy. For ideal gas $\Gamma = (\gamma + 1)/2 > 1$. For some complex fluids and some particular conditions of pressure and temperature, Γ may be lower than one, implying that $(\partial c / \partial \rho)_s < 0$. This means that the acceleration of pressure perturbations through a variable density fluids may be reversed and become a deceleration. It has been shown that, for $\Gamma \ll 1$, compression shocks are strongly reduced, thus alleviating the shock intensity. This has great potential in increasing the efficiency. This is why so much interest is put on dense gas ORCs.

The simulation of these gases requires accurate thermodynamic models, such as Span-Wagner or Peng-Robinson (see [43]). The data to build these models is scarce due to the difficulty of performing reliable experiments. The related uncertainty is thus very high. Our work will go in the following directions:

1. develop deterministic models for the turbine and the other elements of the cycle. These will involve multi-dimensional high fidelity, as well as intermediate and low fidelity (one- and zero-dimensional), models for the turbine, and some 0D/1D models for other element of the cycle (pump, condenser, etc) ;
2. validation of the coupling between the various elements. The following aspects will be considered: characterization of the uncertainties on the cycle components (e.g. empirical coefficients modelling the pump or the condenser), calibration of the thermodynamic parameters, model the uncertainty of each element, and the influence of the unsteady experimental data ;
3. demonstrate the interest of a specific optimization of geometry, operating conditions, and the choice of the fluid, according to the geographical location by including local solar radiation data. Multi-objective optimization will be considered to maximize performance indexes (e.g. Carnot efficiency, mechanical work and energy production), and to reduce the variability of the output.

This work will provide modern tools for the robust design of ORCs systems. It benefits from the direct collaboration with the SME EXOES (ANR LabCom VIPER), and from a collaboration with LEMMA.

Wave energy conversion is an emerging sector in energy engineering. The design of new and efficient Wave Energy Converters (WECs) is thus a crucial activity. As pointed out by Weber [93], it is more economical to raise the technology performance level (TPL) of a wave energy converter concept at low technology readiness level (TRL). Such a development path puts a greater demand on the numerical methods used. The findings of Weber also tell us that important design decisions as well as optimization should be performed as early in the development process as possible. However, as already mentioned, today the wave energy sector relies heavily on the use of tools based on simplified linear hydrodynamic models for the prediction of motions, loads, and power production. Our objective is to provide this sector, and especially SMEs, with robust design tools to minimize the uncertainties in predicted power production, loads, and costs of wave energy.

Following our initial work [53], we will develop, analyse, compare, and use for multi-fidelity optimization, non-linear models of different scales (fidelity) ranging from simple linear hydrodynamics over asymptotic discrete nonlinear wave models, to non-hydrostatic anisotropic Euler free surface solvers. We will not work on the development of small scale models (VOF-RANS or LES) but may use such models, developed by our collaborators, for validation purposes. These developments will benefit from all our methodological work on asymptotic modelling and high order discretizations. As shown in [53], asymptotic models for WECs involve an equation for the pressure on the body inducing a PDE structure similar to that of incompressible flow equations. The study of appropriate stable and efficient high order approximations (coupling velocity-pressure, efficient time stepping) will be an important part of this activity. Moreover, the flow-floating body interaction formulation introduces time stepping issues similar to those encountered in fluid structure interaction problems, and require a clever handling of complex floater geometries based on adaptive and ALE techniques. For this application, the derivation of fully discrete asymptotics may actually simplify our task.

Once available, we will use this hierarchy of models to investigate and identify the modelling errors, and provide a more certain estimate of the cost of wave energy. Subsequently we will look into optimization cycles by comparing time-to-decision in a multi-fidelity optimization context. In particular, this task will include the development and implementation of appropriate surrogate models to reduce the computational cost of expensive high fidelity models. Here especially artificial neural networks (ANN) and Kriging response surfaces (KRS) will be investigated. This activity on asymptotic non-linear modelling for WECs, which has had very little attention in the past, will provide entirely new tools for this application. Multi-fidelity robust optimization is also an approach which has never been applied to WECs.

This work is the core of the EU OCEANer-net MIDWEST project, which we coordinate. It will be performed in collaboration with our European partners, and with a close supervision of European SMEs in the sector, which are part of the steering board of MIDWEST (WaveDragon, Waves4Power, Tecnalia).

4.3. Materials engineering

Because of their high strength and low weight, ceramic-matrix composite materials (CMCs) are the focus of active research for aerospace and energy applications involving high temperatures, either military or civil. Though based on brittle ceramic components, these composites are not brittle due to the use of a fibre/matrix interphase that preserves the fibres from cracks appearing in the matrix. Recent developments aim at implementing also in civil aero engines a specific class of Ceramic Matrix Composite materials (CMCs) that show a self-healing behaviour. Self-healing consists in filling cracks appearing in the material with a dense fluid formed in-situ by oxidation of part of the matrix components. Self-healing (SH) CMCs are composed of a complex three-dimensional topology of woven fabrics containing fibre bundles immersed in a matrix coating of different phases. The oxide seal protects the fibres which are sensitive to oxidation, thus delaying failure. The obtained lifetimes reach hundreds of thousands of hours [81].

The behaviour of a fibre bundle is actually extremely variable, as the oxidation reactions generating the self-healing mechanism have kinetics strongly dependent on temperature and composition. In particular, the lifetime of SH-CMCs depends on: (i) temperature and composition of the surrounding atmosphere; (ii) composition and topology of the matrix layers; (iii) the competition of the multidimensional diffusion/oxidation/volatilization processes; (iv) the multidimensional flow of the oxide in the crack; (v) the inner topology of fibre bundles; (vi) the distribution of critical defects in the fibres. Unfortunately, experimental investigations on the full materials are too long (they can last years) and their output too qualitative (the coupled effects can only be observed a-posteriori on a broken sample). Modelling is thus essential to study and to design SH-CMCs.

In collaboration with the LCTS laboratory (a joint CNRS-CEA-SAFRAN-Bordeaux University lab devoted to the study of thermo-structural materials in Bordeaux), we are developing a multi-scale model in which a structural mechanics solver is coupled with a closure model for the crack physico chemistry. This model is obtained as a multi-dimensional asymptotic crack averaged approximation for the transport equations (Fick's laws) with chemical reactions sources, plus a potential model for the flow of oxide [46], [51], [79]. We have demonstrated the potential of this model in showing the importance of taking into account the multi-dimensional topology of a fibre bundle (distribution of fibres) in the rupture mechanism. This means that the 0-dimensional model used in most of the studies (see e.g. [41]) will underestimate appreciably the lifetime of the material. Based on these recent advances, we will further pursue the development of multi-scale multi-dimensional asymptotic closure models for the parametric design of self healing CMCs. Our objectives are to provide: (i) new, non-linear multi-dimensional mathematical model of CMCs, in which the physico-chemistry of the self-healing process is more strongly coupled to the two-phase (liquid gas) hydro-dynamics of the healing oxide; (ii) a model to represent and couple crack networks; (iii) a robust and efficient coupling with the structural mechanics code; (iv) validate this platform with experimental data obtained at the LCTS laboratory. The final objective is to set up a multi-scale platform for the robust prediction of lifetime of SH-CMCs, which will be a helpful tool for the tailoring of the next generation of these materials.

4.4. Coastal and civil engineering

Our objective is to bridge the gap between the development of high order adaptive methods, which has mainly been performed in the industrial context and environmental applications, with particular attention to coastal and hydraulic engineering. We want to provide tools for adaptive non-linear modelling at large and intermediate scales (near shore, estuarine and river hydrodynamics). We will develop multi-scale adaptive models for free surface hydrodynamics. Beside the models and codes themselves, based on the most advanced numerics we will develop during this project, we want to provide sufficient know how to control, adapt and optimize these tools.

We will focus our effort in the understanding of the interactions between asymptotic approximations and numerical approximations. This is extremely important in at least two aspects. The first is the capability of a numerical model to handle highly dispersive wave propagation. This is usually done by high accuracy asymptotic PDE expansions. Here we plan to make heavily use of our results concerning the relations between vertical asymptotic expansions and standard finite element approximations. In particular, we will invest some effort in the development of $xy+z$ adaptive finite element approximations of the incompressible Euler equations. Local p -adaptation of the vertical approximation may provide a "variable depth" approximation exploiting numerics instead of analytical asymptotics to control the physical behaviour of the model.

Another important aspect which is not understood well enough at the moment is the role of dissipation in wave breaking regions. There are several examples of breaking closure, going from algebraic and PDE-based eddy viscosity methods [64], [83], [77], [48], to hybrid methods coupling dispersive PDEs with hyperbolic ones, and trying to mimic wave breaking with travelling bores [87], [88], [86], [62], [54]. In both cases, numerical dissipation plays an important role and the activation or not of the breaking closure, as the quantitative contribution of numerical dissipation to the flow has not been properly investigated. These elements must be clarified to allow full control of adaptive techniques for the models used in this type of applications.

Another point we want to clarify is how to optimize the discretization of asymptotic PDE models. In particular, when adding mesh size(s) and time step, we are in presence of at least 3 (or even more) small parameters. The relations between physical ones have been more or less investigated, as have been the ones between purely numerical ones. We plan to study the impact of numerics on asymptotic PDE modelling by reverting the usual process and studying asymptotic limits of finite element discretizations of the Euler equations. Preliminary results show that this does allow to provide some understanding of this interaction and to possibly propose considerably improved numerical methods [32].

CELESTE Project-Team

4. Application Domains

4.1. Neglected tropical diseases

CELESTE collaborates with Anavaj Sakuntabhai and Philippe Dussart (Pasteur Institute) on predicting dengue severity using only low-dimensional clinical data obtained at hospital arrival. Further collaborations are underway in dengue fever and encephalitis with researchers at the Pasteur Institute, including with Jean-David Pommier.

4.2. Pharmacovigilance

In pharmacovigilance, the goal is to detect, as soon as possible, potential associations between certain drugs and adverse effects, which appeared after the authorized marketing of these drugs. Preceding works showed the importance of defining an adapted methodology to deal with the size of the individual data (around 250000 reports, 2000 drugs, 4000 adverse effects) and their sparsity. We will explore several aspects from software point of view to statistical strategies as sub-sampling.

4.3. Electricity load consumption: forecasting and control

CELESTE has a long-term collaboration with EDF R&D about electricity consumption. An important problem is to forecast consumption. We currently work on an approach involving back and forth disaggregation (of the total consumption into the consumptions of well-chosen groups/regions) and aggregation of local estimates. We also work on consumption control by price incentives sent to specific users (volunteers), seeing it as a bandit problem.

4.4. Reliability

Collected product lifetime data is often non-homogeneous, affected by production variability and differing real-world usage. Usually, this variability is not controlled or observed in any way, but needs to be taken into account for reliability analysis. Latent structure models are flexible models commonly used to model unobservable causes of variability.

CELESTE currently collaborates with PSA Group. To dimension its vehicles, the PSA Group uses a reliability design method called Strength-Stress, which takes into consideration both the statistical distribution of part strength and the statistical distribution of customer load (called Stress). In order to minimize the risk of in-service failure, the probability that a severe customer will encounter a weak part must be quantified. Severity quantification is not simple since vehicle use and driver behaviour can be severe for some types of materials and not for others. The aim of the study is then to define a new and richer notion of the severity from the PSA databases resulting either from tests or client usages. This will lead to a more robust and accurate parts dimensioning method.

4.5. Spectroscopic imaging analysis of ancient materials

Ancient materials, encountered in archaeology and paleontology are often complex, heterogeneous and poorly characterized before physico-chemical analysis. A popular technique is to gather as much physico-chemical information as possible, is spectro-microscopy or spectral imaging, where a full spectra, made of more than a 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 requires the combination of an "image" approach with a "curve analysis" approach. Since 2010 CELESTE (previously SELECT) collaborates with Serge Cohen (IPANEMA) on clustering problems, taking spatial constraints into account.

4.6. Forecast of dwell time during train parking at stations

This is a Cifre PhD in collaboration with SNCF.

One of the factors in the punctuality of trains in dense areas (and management crisis in the event of an incident on a line) is the respect of both the travel time between two stations and the parking time in a station. These depend, among other things, on the train, its mission, the schedule, the instantaneous charge, and the configuration of the platform or station. Preliminary internal studies at SNCF have shown that the problem is complex. From a dataset concerning line E of the Transilien in Paris, we will address prediction (machine learning) and modeling (statistics): (1) construct a model of station-hours, station-hours-type of train, by example using co-clustering techniques; (2) study the correlations between the number of passengers (load), up and down flows, and parking times, and possibly other variables to be defined; (3) model the flows or loads (within the same station, or the same train) as a stochastic process; (4) develop a realistic digital simulator of passenger flows and test different scenarios of incidents and resolution, in order to propose effective solutions.

4.7. Fatigue aided-design

This is a Cifre PhD in collaboration with PSA.

The digitalization of design is at the heart of the processes of automotive manufacturers departments, to enable them to reduce costs and development time. This also applies to reliability studies of certain components of the chassis of a vehicle, and the will is to drastically reduce the number of physical tests to tend towards an almost entirely digital design having only one phase of validation. Deterministic models, although developed from detailed design drawings, can predict behaviors different from those observed on the structure during testing. These deviations can be due to the more or less faithful discretization of the geometry, the uncertainties on some parameters of the model (such as the properties of the materials, the boundary conditions), or the random loadings undergone by the structure (Beck and Katafygiotis, 1998). It is important to make available new methods in addition to the classical finite element (FE) deterministic modeling, to enable the exploitation of the accumulated data over the years for all the projects: computation results, measurements and test data.

One of the objectives of this project is to propose a probabilistic modeling of the behavior of a structure starting from a FE model, taking into account the non assignable fluctuations of the model, in order to define a probabilistic criterion of rupture and its margins of confidence. The following three steps are envisaged: (1) Define relevant prior information using business experience feedback (REX) and use a Bayesian estimation to calibrate the parameters. This REX is consequent and will require advanced statistical processing of machine learning, and in particular in clustering to identify similarities or similar patterns among several models. The estimation will use Bayesian non-iterative methods (Celeux and Pamphile, 2019), which are less expensive and less unstable than conventional methods. This will test their effectiveness in this context. (2) Select important parameters (physical or modeling). (3) Define a probabilistic criterion of coaxial fatigue taking into account both the random behavior of the structure and the material (Fouchereau et al., 2014) extending the existing deterministic criteria (Dang-Van, 1993).

COMMANDS Project-Team

4. Application Domains

4.1. Energy management for hybrid vehicles

In collaboraton with Ifpen and in the framework of A. Le Rhun's thesis, we have developped a methodology for the optimal energy management for hybrid vehicles, based on a statistical analysis of the traffic. See [12], [12], [7].

4.2. Biological cells culture

In collaboration with the Inbio team (Inst. Pasteur and Inria) we started to study the optimization of protein production based on cell culture.

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 is the collaboration that started in September 2014 with THALES Optronique. The goal of this project is the optimization of the maintenance of an onboard system equipped with a HUMS (Health Unit Monitoring Systems). The physical system under consideration is modeled by a piecewise deterministic Markov process. In the context of impulse control, we propose a dynamic maintenance policy, adapted to the state of the system and taking into account both random failures and those related to the degradation phenomenon.

The spectrum of applications of the topics that the team can address is large and can 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.

4.5. Fluid flow applications

Specific actions are devoted to the problem of atmospheric reentry simulations. We focus on several aspects : i) on the development of innovative algorithms improving the prediction of hypersonic flows and including system uncertainties, ii) on the application of these methods to the atmospheric reentry of space vehicles for the control and the optimization of the trajectory, iii) on the debris reentry, which is of fundamental importance for NASA, CNES and ESA. Several works are already initiated with funding from CNES, Thales, and ASL. An ongoing activity concerns the design of the Thermal Protection System (TPS) that shields the spacecraft from aerothermal heating, generated by friction at the surface of the vehicle. The TPS is usually composed of different classes of materials, depending on the mission and the planned trajectory. One major issue is to model accurately the material response to ensure a safe design. High-fidelity material modeling for ablative materials has been developed by NASA, but a lot of work is still needed concerning the assessment of physical and modeling uncertainties during the design process. Our objective is to set up a predictive numerical tool to reliably estimate the response of ablative materials for different aerothermal conditions.

An important effort is dedicated to the simulation of fluids featuring complex thermodynamic behavior, in the context of two distinct projects: the VIPER project, funded by Aquitaine Region, and a project with CWI (Scientific Computing Group). Dense gases (DGs) are defined as single-phase vapors operating at temperatures and pressures conditions close to the saturation curve. The interest in studying complex dynamics of compressible dense gas flows comes from the potential technological advantages of using these fluids in energy conversion cycles, such as in Organic Rankine Cycles (ORCs) which used dense gases as energy converters for biomass fuels and low-grade heat from geothermal or industrial waste heat sources. Since these fluids feature large uncertainties in their estimated thermodynamic properties (critical properties, acentric factor, etc.), a meaningful numerical prediction of the performance must necessarily take into account these uncertainties. Other sources of uncertainties include, but are not limited to, the inlet boundary conditions which are often unknown in dense gases applications. Moreover, a robust optimization must also include the more generic uncertainty introduced by the machining tolerance in the construction of the turbine blades.

DISCO Project-Team

4. Application Domains

4.1. Analysis and Control of life sciences systems

The team is involved in life sciences applications. The two main lines are the analysis of bioreactors models (microorganisms; bacteria, microalgae, yeast, etc..) and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML) in collaboration with St Antoine Hospital in Paris. A recent subject is the modelling of epidemics for tropical diseases.

4.2. Energy Management

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

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,
- adaption of parameters for classification tools such as Machine Learning systems, in which Adjoint Differentiation is also known as *backpropagation*.
- 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 criterion with respect to all inputs, and possibly additional gradients when there are constraints. Adjoint 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 of the altitude and velocities of the surface. Figure 1 shows an example of an inverse problem using the glaciology code ALIF (a pure C version of ISSM [30]) and its AD-adjoint produced by Tapenade.

One particular case of inverse problems is *data assimilation* [31] 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 though AD [33]. 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.

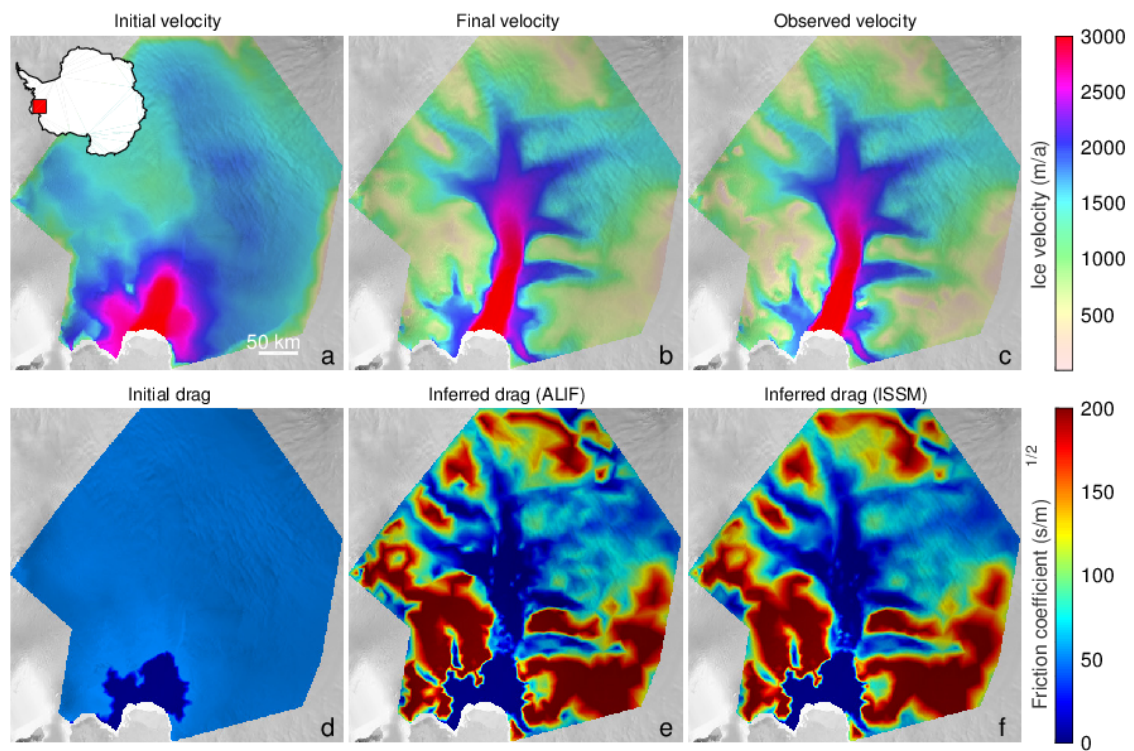


Figure 1. Assimilation of the basal friction under Pine Island glacier, West Antarctica. The final simulated surface velocity (b) is made to match the observed surface velocity (c), by estimation of the basal friction (e). A reference basal friction (f) is obtained by another data assimilation using the hand-written adjoint of ISSM

4.4. Linearization

Simulating a complex system often requires solving a system of Partial Differential Equations. This can be too expensive, in particular for real-time simulations. 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.

ELAN Project-Team

4. Application Domains

4.1. Mechanical Engineering

Many physicists and mathematicians have strived for centuries to understand the principles governing those complex mechanical phenomena, providing a number of continuous models for slender structures, granular matter, and frictional contact. In the XXth century, industrial applications such as process automatization and new ways of transportation have boosted the fields of Mechanical Engineering and Computer-Aided Design, where material strength, reliability of mechanisms, and safety, stood for the main priorities. Instead, large displacements of structures, buckling, tearing, or entanglement, and even dynamics, were long considered as undesirable behaviors, thus restraining the search for corresponding numerical models.

Only recently, the engineering industry has shown some new and growing interest into the modeling of dynamic phenomena prone to large displacements, contact and friction. For instance, the cosmetology industry is more and more interested in understanding the nonlinear deformation of hair and skin, with the help of simulation. Likewise, auto and aircraft manufacturers are facing new challenges involving buckling or entanglement of thin structures such as carbon or optical fibers; they clearly lack predictive, robust and efficient numerical tools for simulating and optimizing their new manufacturing process, which share many common features with the large-scale simulation scenarii traditionally studied in Computer Graphics applications.

4.2. Computer Graphics

In contrast, Computer Graphics, which has emerged in the 60's with the advent of modern computers, was from the very beginning eager to capture such peculiar phenomena, with the sole aim to produce spectacular images and create astonishing stories. At the origin, Computer Graphics thus drastically departed from other scientific fields. Everyday-life phenomena such as cloth buckling, paper tearing, or hair fluttering in the wind, mostly ignored by other scientists at that time, became actual topics of interest, involving a large set of new research directions to be explored, both in terms of modelling and simulation. Nowadays, although the image production still remains the core activity of the Computer Graphics community, more and more research studies are directed through the virtual and real prototyping of mechanical systems, notably driven by a myriad of new applications in the virtual try on industry (e.g., hairstyling and garment fitting). Furthermore, the advent of additive fabrication is currently boosting research in the free design of new mechanisms or systems for various applications, from architecture design and fabrication of metamaterials to the creation of new locomotion modes in robotics. Some obvious common interests and approaches are thus emerging between Computer Graphics and Mechanical Engineering, yet the two communities remain desperately compartmentalized.

4.3. Soft Matter Physics

From the physics-based viewpoint, since a few decades a new generation of physicists became interested again in the understanding of such visually fascinating phenomena, and started investigating the tight links between geometry and elasticity⁰. Common objects such as folded or torn paper, twined plants, coiled honey threads, or human hair have thus regained some popularity among the community in Nonlinear Physics⁰. In consequence, phenomena of interest have become remarkably close to those of Computer Graphics, since scientists in both places share the common goal to model complex and integrated mechanical phenomena at the macroscopic

⁰In France this new trend was particularly stimulated by the work of Yves Pomeau, who convinced many young scientists to study the nonlinear physics of common objects such as paper, plants, or hair [27].

⁰It is however amusing to observe that research in these areas is quite successful in obtaining the IG Nobel prize [8], [30], thus still being considered as an exotic research topic by physicists.

scale. Of course, the goals and employed methodologies differ substantially from one community to the other, but showcase some evident complementarity: while computer scientists are eager to learn and understand new physical models, physicists get more and more interested in the numerical tools, in which they perceive not only a means to confirm predictions afterwards, but also a support for testing new hypothesis and exploring scenarios that would be too cumbersome or even impossible to investigate experimentally. Besides, numerical exploration starts becoming a valuable tool for getting insights into the search for analytic solutions, thus fully participating to the modeling stage and physical understanding. However, physicists may be limited to a blind usage of numerical black boxes, which may furthermore not be dedicated to their specific needs. According to us, promoting a science of modeling in numerical physics would thus be a promising and rich avenue for the two research fields. Unfortunately, very scarce cooperation currently exists between the two communities, and large networks of collaboration still need to be set up.

FACTAS 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 magnetization problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Konstantinos Mavreas.

Generally speaking, inverse potential problems, similar to the one appearing in Section 4.3 , 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. Though this Associate Team reached the end of its term in 2018, the collaborations it has generated are still active. A joint work with Cerege (CNRS, Aix-en-Provence), in the framework of the ANR-project MagLune, completes this picture, see Sections 6.1.2 , 8.2.1 .

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. The associate team IMPINGE 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 planar region at small distance from the slab.

Mathematically speaking, both inverse source problems for EEG from Section 4.3 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 potential:

$$V(x) = \int_{\Omega} \frac{\operatorname{div} m(x') dx'}{|x-x'|}, \quad (2)$$

outside the volume Ω of the object. Depending on the geometry of models, the magnetization distribution m may lie in a volume or spread out on a surface. This results in quite different identifiability properties, see [37] and Section 6.1.1 , but the two situations share a substantial mathematical common core.

Another timely instance of inverse magnetization problems lies with geomagnetism. Satellites orbiting around the Earth measure the magnetic field at many points, and nowadays it is a challenge to extract global information from those measurements. In collaboration with C. Gerhards (Geomathematics and Geoinformatics Group, Technische Universität Bergakademie Freiberg, Germany), we started to work on the problem of separating the magnetic field due to the magnetization of the globe's crust from the magnetic field due to convection in the liquid metallic core. The techniques involved are variants, in a spherical context, from those developed within the IMPINGE associate team for paleomagnetism, see Section 6.1.1 .

4.3. Inverse source problems in EEG

Participants: Paul Asensio, Laurent Baratchart, Juliette Leblond, Jean-Paul Marmorat, Masimba Nemaire.

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, see [9]. 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 [8], [42]. 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. Our belief is that such a localization process can add a geometric, valuable piece of information to the standard temporal analysis of EEG signal records.

Numerical experiments obtained with our software FindSources3D give very good results on simulated data and we are now engaged in the process of handling real experimental data, simultaneously recorded by EEG and MEG devices, in collaboration with our partners at INS, hospital la Timone, Marseille (see Section 6.1.3).

Furthermore, another approach is being studied for EEG, that consists in regularizing the inverse source problem by a total variation constraint on the source term (a measure), added to the quadratic data approximation criterion. It is similar to the path that is taken for inverse magnetization problems (see Sections 4.2 and 6.1.1), and it presently focuses on surface-distributed models.

4.4. Identification and design of microwave devices

Participants: Laurent Baratchart, Sylvain Chevillard, Jean-Paul Marmorat, Martine Olivi, Fabien Seyfert.

This is joint work with Stéphane Bila (Xlim, Limoges).

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).

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

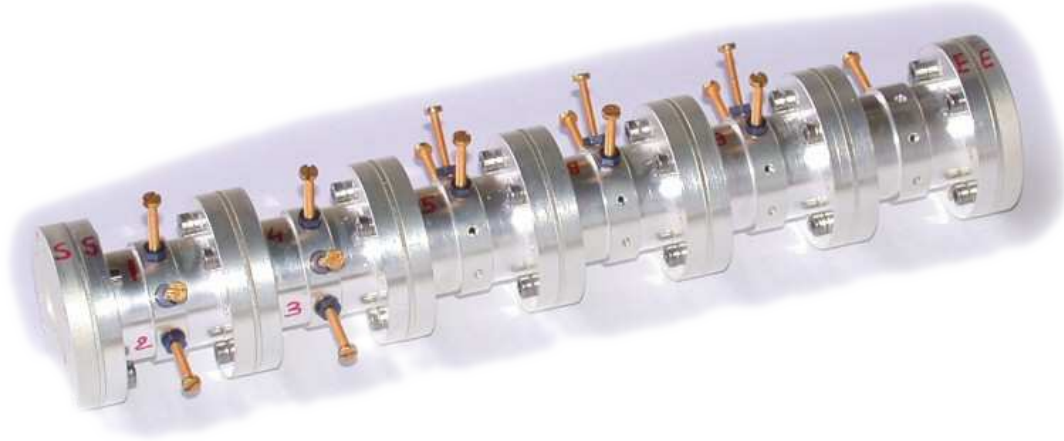


Figure 1. Picture of a 6-cavities dual mode filter. Each cavity (except the last one) has 3 screws to couple the modes within the cavity, so that 16 quantities must be optimized. Quantities such as the diameter and length of the cavities, or the width of the 11 slits are fixed during the design phase.

each cavity, and the width of each iris. Subsequently, the filter can be constructed and tuned by adjusting the screws. Finally, the screws are glued once a satisfactory response has been obtained. 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 to the Helmholtz equations is given by 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 (surface currents) in the cavities.

This way, the filter can be seen as a quadripole, with two ports, when plugged onto 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 matrix S is approximately 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 a 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) in order to get a model for the scattering matrix as an analytic matrix-valued function. This provides us with a scattering matrix known to be close to a rational matrix of order roughly 1/4 of the number of data points.
- Approximating this scattering matrix by a true rational transfer-function of appropriate degree (8 in this example) via the Endymion or RARL2 software (cf. Section 3.3.2.2).
- A state space realization of S , viewed as a transfer function, can then be obtained, where additional symmetry constraints coming from the reciprocity law and possibly other physical features of the device have to be 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).

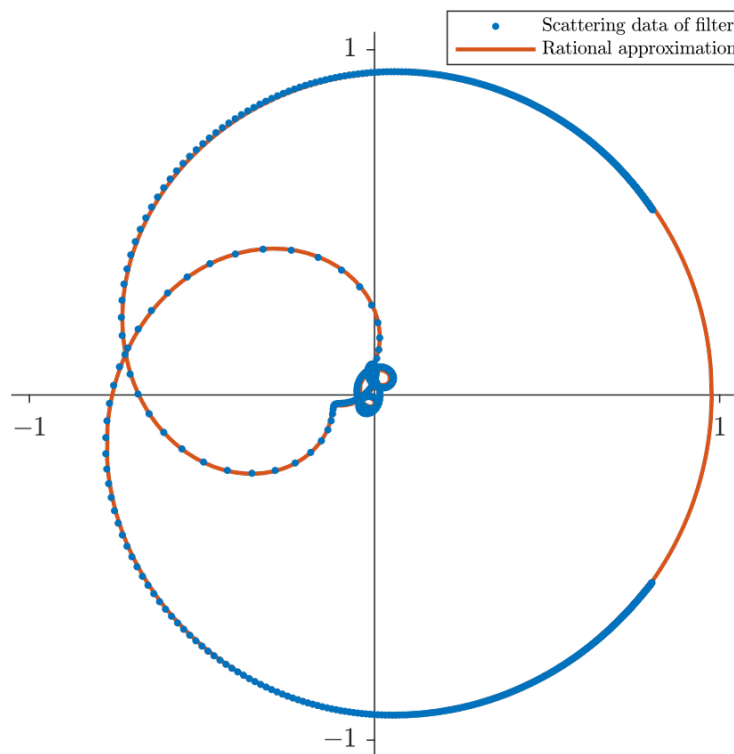


Figure 2. Nyquist Diagram of S_{22} . The rational approximation is of degree 8.

The final approximation is of high quality. This can be interpreted as a confirmation of the linearity assumption on the system: the relative L^2 error is less than 10^{-3} . This is illustrated by a reflection diagram (Figure 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 further 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 [12]. 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 may be chosen freely. The denominator q is then 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 no longer the case when considering the synthesis N -port devices for $N > 3$, like multiplexers, routers and power dividers, or when considering the synthesis of filters under matching conditions. The efficient derivation of multiplexers responses is the subject of recent investigation by Factas, using techniques based on constrained Nevanlinna-Pick interpolation (see Section 6.2).

Through contacts with CNES (Toulouse) and UPV (Bilbao), Apics got additionally involved in 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 non-linearities made it possible to compute the harmonic response of a (possibly nonlinear and unstable) device [79]. This has had tremendous impact on design, and there is a growing demand for software analyzers. The team is also becoming active in this area.

In this connection, 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). Applications by the team so far have been concerned with the first type of stability, and emphasis is put on defining and extracting the "unstable part" of the response, see Section 6.3. The stability check for limit cycles has made important theoretical advances, and numerical algorithms are now under investigation.

GAMMA Project-Team (section vide)

GEOSTAT Project-Team

4. Application Domains

4.1. Sparse signals & optimization

This research topic involves **Geostat** team and is used to set up an InnovationLab with **I2S company**

Sparsity can be used in many ways and there exist various sparse models in the literature; for instance minimizing the l_0 quasi-norm is known to be an NP-hard problem as one needs to try all the possible combinations of the signal's elements. The l_1 norm, which is the convex relation of the l_0 quasi-norm results in a tractable optimization problem. The l_p pseudo-norms with $0 < p < 1$ are particularly interesting as they give closer approximation of l_0 but result in a non-convex minimization problem. Thus, finding a global minimum for this kind of problem is not guaranteed. However, using a non-convex penalty instead of the l_1 norm has been shown to improve significantly various sparsity-based applications. Nonconvexity has a lot of statistical implications in signal and image processing. Indeed, natural images tend to have a heavy-tailed (kurtotic) distribution in certain domains such as wavelets and gradients. Using the l_1 norm comes to consider a Laplacian distribution. More generally, the hyper-Laplacian distribution is related to the l_p pseudo-norm ($0 < p < 1$) where the value of p controls how the distribution is heavy-tailed. As the hyper-Laplacian distribution for $0 < p < 1$ represents better the empirical distribution of the transformed images, it makes sense to use the l_p pseudo-norms instead of l_1 . Other functions that better reflect heavy-tailed distributions of images have been used as well such as Student-t or Gaussian Scale Mixtures. The internal properties of natural images have helped researchers to push the sparsity principle further and develop highly efficient algorithms for restoration, representation and coding. Group sparsity is an extension of the sparsity principle where data is clustered into groups and each group is sparsified differently. More specifically, in many cases, it makes sense to follow a certain structure when sparsifying by forcing similar sets of points to be zeros or non-zeros simultaneously. This is typically true for natural images that represent coherent structures. The concept of group sparsity has been first used for simultaneously shrinking groups of wavelet coefficients because of the relations between wavelet basis elements. Lastly, there is a strong relationship between sparsity, nonpredictability and scale invariance.

We have shown that the two powerful concepts of sparsity and scale invariance can be exploited to design fast and efficient imaging algorithms. A general framework has been set up for using non-convex sparsity by applying a first-order approximation. When using a proximal solver to estimate a solution of a sparsity-based optimization problem, sparse terms are always separated in subproblems that take the form of a proximal operator. Estimating the proximal operator associated to a non-convex term is thus the key component to use efficient solvers for non-convex sparse optimization. Using this strategy, only the shrinkage operator changes and thus the solver has the same complexity for both the convex and non-convex cases. While few previous works have also proposed to use non-convex sparsity, their choice of the sparse penalty is rather limited to functions like the l_p pseudo-norm for certain values of $p \geq 0.5$ or the Minimax Concave (MC) penalty because they admit an analytical solution. Using a first-order approximation only requires calculating the (super)gradient of the function, which makes it possible to use a wide range of penalties for sparse regularization. This is important in various applications where we need a flexible shrinkage function such as in edge-aware processing. Apart from non-convexity, using a first-order approximation makes it easier to verify the optimality condition of proximal operator-based solvers via fixed-point interpretation. Another problem that arises in various imaging applications but has attracted less works is the problem of multi-sparsity, when the minimization problem includes various sparse terms that can be non-convex. This is typically the case when looking for a sparse solution in a certain domain while rejecting outliers in the data-fitting term. By using one intermediate variable per sparse term, we show that proximal-based solvers can be efficient. We give a detailed study of the Alternating Direction Method of Multipliers (ADMM) solver for multi-sparsity and study its properties. The following subjects are addressed and receive new solutions:

1. **Edge aware smoothing:** given an input image g , one seeks a smooth image u "close" to g by minimizing:

$$\operatorname{argmin}_u \frac{\lambda}{2} \|u - g\|_2^2 + \psi(\nabla u)$$

where ψ is a sparsity-inducing non-convex function and λ a positive parameter. Splitting and alternate minimization lead to the sub-problems:

$$\begin{aligned} \text{(sp1)} & : v^{(k+1)} \leftarrow \operatorname{argmin}_v \psi(v) + \frac{\beta}{2} \|\nabla u^{(k)} - v\|_2^2 \\ \text{(sp2)} & : u^{(k+1)} \leftarrow \operatorname{argmin}_u \lambda \|u - g\|_2^2 + \beta \|\nabla u - v^{(k+1)}\|_2^2. \end{aligned}$$

We solve sub-problem (sp2) through deconvolution and efficient estimation via separable filters and warm-start initialization for fast GPU implementation, and sub-problem (sp1) through non-convex proximal form.

2. **Structure-texture separation:** design of an efficient algorithm using non-convex terms on both the data-fitting and the prior. The resulting problem is solved via a combination of Half-Quadratic (HQ) and Maximization-Minimization (MM) methods. We extract challenging texture layers outperforming existing techniques while maintaining a low computational cost. Using spectral sparsity in the framework of low-rank estimation, we propose to use robust Principal Component Analysis (RPCA) to perform robust separation on multi-channel images such as glare and artifacts removal of flash/no-flash photographs. As in this case, the matrix to decompose has much less columns than lines, we propose to use a QR decomposition trick instead of a direct singular value decomposition (SVD) which makes the decomposition faster.
3. **Robust integration:** in many applications, we need to reconstruct an image from corrupted gradient fields. The corruption can take the form of outliers only when the vector field is the result of transformed gradient fields (low-level vision), or mixed outliers and noise when the field is estimated from corrupted measurements (surface reconstruction, gradient camera, Magnetic Resonance Imaging (MRI) compressed sensing, etc.). We use non-convexity and multi-sparsity to build efficient integrability enforcement algorithms. We present two algorithms : 1) a local algorithm that uses sparsity in the gradient field as a prior together with a sparse data-fitting term, 2) a non-local algorithm that uses sparsity in the spectral domain of non-local patches as a prior together with a sparse data-fitting term. Both methods make use of a multi-sparse version of the Half-Quadratic solver. The proposed methods were the first in the literature to propose a sparse regularization to improve integration. Results produced with these methods significantly outperform previous works that use no regularization or simple l_1 minimization. Exact or near-exact recovery of surfaces is possible with the proposed methods from highly corrupted gradient fields with outliers.
4. **Learning image denoising:** deep convolutional networks that consist in extracting features by repeated convolutions with high-pass filters and pooling/downsampling operators have shown to give near-human recognition rates. Training the filters of a multi-layer network is costly and requires powerful machines. However, visualizing the first layers of the filters shows that they resemble wavelet filters, leading to sparse representations in each layer. We propose to use the concept of scale invariance of multifractals to extract invariant features on each sparse representation. We build a bi-Lipschitz invariant descriptor based on the distribution of the singularities of the sparsified images in each layer. Combining the descriptors of each layer in one feature vector leads to a compact representation of a texture image that is invariant to various transformations. Using this descriptor that is efficient to calculate with learning techniques such as classifiers combination and artificially adding training data, we build a powerful texture recognition system that outperforms previous works on 3 challenging datasets. In fact, this system leads to quite close recognition rates compared to latest advanced deep nets while not requiring any filters training.

I4S Project-Team (section vide)

INOCS Project-Team

4. Application Domains

4.1. Energy

In energy, the team mainly focuses on pricing models for demand side management. Demand side management methods are traditionally used to control electricity demand which became quite irregular recently and resulted in inefficiency in supply. We have explored the relationship between energy suppliers and customers who are connected to a smart grid. The smart grid technology allows customers to keep track of hourly prices and shift their demand accordingly, and allows the provider to observe the actual demand response to its pricing strategy. We tackle pricing problems in energy according to the bilevel optimization approaches. Some research works in this domain are supported by bilateral grants with EDF.

4.2. Transportation and Logistics

In transportation and logistics, the team addresses mainly integrated problems, which require taking into account simultaneously different types of decision. Examples are location and routing, inventory management and routing or staff scheduling and warehouse operations management. Such problems occur from the supply chain design level to the logistic facility level.

4.3. Telecommunications

In telecommunications, the team mainly focuses on network design problems and on routing problems. Such problems are optimization problems with complex structure, since the optimization of capacity installation and traffic flow routing have to be addressed simultaneously.

MATERIALS Project-Team (section vide)

MATHRISK Project-Team

4. Application Domains

4.1. Financial Mathematics, Insurance

The domains of application are quantitative finance and insurance with emphasis on risk modeling and control. In particular, Mathrisk focuses on dependence modeling, systemic risk, market microstructure modeling and risk measures.

MCTAO Project-Team

4. Application Domains

4.1. Aerospace Engineering

Participants: Bernard Bonnard, Jean-Baptiste Caillau, Thierry d'Argent, Lamberto Dell'Elce, Jean-Baptiste Pomet, Jérémy Rouot.

Space engineering is very demanding in terms of safe and high-performance control laws. It is therefore prone to fruitful industrial collaborations. McTAO now has an established expertise in space and celestial mechanics. Our collaborations with industry are mostly on orbit transfer problems with low-thrust propulsion. It can be orbit transfer to put a commercial satellite on station, in which case the dynamics are a Newtonian force field plus perturbations and the small control. There is also, currently, a renewed interest in low-thrust missions such as Lisa Pathfinder (ESA mission towards a Lagrange point of the Sun-Earth system) or BepiColombo (joint ESA-JAXA mission towards Mercury). Such missions look more like a controlled multibody system. In all cases the problem involves long orbit transfers, typically with many revolutions around the primary celestial body. When minimizing time, averaging techniques provide a good approximation. Another important criterion in practice is fuel consumption minimization (crucial because only a finite amount of fuel is onboard a satellite for all its "life"), which amounts to L^1 -minimization. Both topics are studied by the team. We have a steady relationships with CNES and Thales Alenia Space (Cannes), that have financed or co-financed 3 PhDs and 2 post-docs in the Sophia location of the team in the decade and are a source of inspiration even at the methodological level. Team members also have close connections with Airbus-Safran (Les Mureaux) on launchers. Some of the authoritative papers in the field were written by team members, with an emphasis on the geometric analysis and on algorithms (coupling of shooting and continuation methods). There are also connections with peers more on the applied side, like D. Scheeres (Colorado Center for Astrodynamics Research at Boulder), the group of F. Bernelli (Politecnico Milano), and colleagues from U. Barcelona (A. Farrès, A. Jorba).

4.2. Magnetic resonance imaging (MRI)

Participants: Bernard Bonnard, Alice Nolot [Professeur Classes Préparatoires Troyes], Jérémy Rouot, Joseph Gergaud, Olivier Cots [ENSEEIH, Toulouse], Stephen Glaser [TU München, Germany], Dominique Sugny [Univ. de Bourgogne].

The starting point of our interest in optimal control for quantum systems was a collaboration with physicist from ICB, University of Burgundy (Dominique Sugny), motivated by an ANR project where we worked on the control of molecular orientation in a dissipative environment using a laser field, and developed optimal control tools, combined with numerical simulations, to analyze the problem for Qubits. This was related to quantum computing rather than MRI. Using this expertise and under the impulse of Prof. S. Glaser and his group (Chemistry, TU München), we investigated Nuclear Magnetic resonance (NMR) for medical imaging (MRI), where the model is the Bloch equation describing the evolution of the Magnetization vector controlled by a magnetic field, but in fine is a specific Qubit model without decoherence. We worked on, and brought strong contributions to, the contrast problem: typically, given two chemical substances that have an importance in medicine, like oxygenated and de-oxygenated blood, find the (time-dependent) magnetic field that will produce the highest difference in brightness between these two species on the image resulting from Nuclear Magnetic Resonance. This has immediate and important industrial applications in medical imaging. Our contacts are with the above mentioned physics academic labs, who are themselves in contact with major companies. The team has produced and is producing important work on this problem. One may find a good overview in [50], a reference book has been published on the topic [54], a very complete numerical study comparing different optimization techniques was performed in [49]. We conduct this project in parallel with S. Glaser team, which validated experimentally the pertinence of the methods, the main achievement being the in vivo experiments

realized at the Creatis team of Insa Lyon showing the interest to use optimal control methods implemented in modern softwares in MRI in order to produce a better image in a shorter time. A goal is to arrive to a cartography of the optimal contrast with respect to the relaxation parameters using LMI techniques and numerical simulations with the Hamapth and Bocop code; note that the theoretical study is connected to the problem of understanding the behavior of the extremal solutions of a controlled pair of Bloch equations, and this is an ambitious task. Also, one of the difficulties to go from the obtained results, checkable on experiments, to practical control laws for production is to deal with magnetic field space inhomogeneities.

4.3. Swimming at low-Reynolds number

Participants: Bernard Bonnard, Yacine El Alaoui-Faris, Laetitia Giraldi, Clément Moreau, Alice Nolot, Jean-Baptiste Pomet, Jérémy Rouot.

Following the historical reference for low Reynolds number locomotion [72], the study of the swimming strategies of micro-organisms is attracting increasing attention in the recent literature. This is both because of the intrinsic biological interest, and for the possible implications these studies may have on the design of bio-inspired artificial replicas reproducing the functionalities of biological systems. In the case of micro-swimmers, the surrounding fluid is dominated by the viscosity effects of the water and becomes reversible. In this regime, it turns out that the infinite dimensional dynamics of the fluid do not have to be retained as state variables, so that the dynamics of a micro-swimmer can be expressed by ordinary differential equations if its shape has a finite number of degrees of freedom. Assuming this finite dimension, and if the control is the rate of deformation, one obtains a control system that is linear (affine without drift) with respect to the controls, *i.e.* the optimal control problem with a quadratic cost defines a sub-Riemannian structure (see section 3.2). This is the case where the shape is “fully actuated”, *i.e.* if all the variables describing the shape are angles, there is an actuator on each of these angles. For artificial micro-swimmers, this is usually unrealistic, hence (artificial) magneto-elastic micro-swimmers, that are magnetized in order to be deformed by an external magnetic field. In this case, the control functions are the external magnetic field. In both cases, questions are controllability (straightforward in the fully actuated case), optimal control, possibly path planning. We collaborate with teams that have physical experiments for both.

- In collaboration with D. Takagi and M. Chyba (Univ of Hawaii), this approach is currently at the experimental level for copepod-like swimmer at the university of Hawaii: on the one hand, this zooplankton and its locomotion can be observed, and a robot micro swimmer mimicking a copepod has been constructed, but in fact large enough for direct actuation to be possible, and the low Reynolds number is achieved by using a more viscous fluid. This gives possibilities, through an inverse optimization problem, to determine what cost can be optimised by these crustaceans, see [40], [78], and to validate models on the robot.
- For magneto-elastic micro-robots, Y. El-Alaoui’s PhD is co-advised with Stéphane Régnier from the robotics lab ISIR, Univ. Paris 6. Magneto-elastic micro-robots and their magnetic actuation are actually built at ISIR and the aim of the collaboration is to validate models and improve the existing control laws both in performance and in energy; of course, the micro scale does make things difficult.

The questions about optimality of periodic controls raised in section 3.2 are related to these applications for periodic deformations, or strokes, play an important role in locomotion.

4.4. Stability of high frequency amplifiers

Participants: Sébastien Fueyo, Gilles Lebeau, Jean-Baptiste Pomet, Laurent Baratchart [FACTAS project-team].

Nonlinear hyper-frequency amplifiers are ubiquitous in cell phone relays and many other devices. They must be as compact as possible, yielding a more complicated design. Computer Assisted Design tools are extensively used; for a given amplifier design, they provide frequency responses but fail to provide information of the stability of the response for each frequency. This stability is crucial for an unstable response will not be observed in practice; the actual device should not be built before stability is asserted. Predicting stability/instability from “simulations” in the Computer Assisted Design tool is of utmost importance (simulation between quotation marks because these simulations are in fact computations in the frequency domain). Potential transfer is important.

Some techniques do exist, see [77], based on creating some virtual perturbations and treating them as the input of a (linearized) control system to be “simulated” using the same tools. In an ongoing collaboration between McTAO and the project-team FACTAS, we work on the mathematical ground of these methods and in particular of the relation between stability and the property of the identified time-varying infinite dimensional systems. See recent developments in Section 7.14.

4.5. Optimal control of microbial cells

Participants: Jean-Baptiste Caillaud, Walid Djema [BIOCORE project-team], Laetitia Giraldi, Jean-Luc Gouzé [BIOCORE project-team], Sofya Maslovskaya, Jean-Baptiste Pomet, Agustín Yabo.

The growth of microorganisms is fundamentally an optimization problem which consists in dynamically allocating resources to cellular functions so as to maximize growth rate or another fitness criterion. Simple ordinary differential equation models, called self-replicators, have been used to formulate this problem in the framework of optimal and feedback control theory, allowing observations in microbial physiology to be explained. The resulting control problems are very challenging due to the nonlinearity of the models, parameter uncertainty, the coexistence of different time-scales, a dynamically changing environment, and various other physical and chemical constraints. In the framework of the ANR Maximic (PI Hidde de Jong, Inria Grenoble Rhône-Alpes) we aim at developing novel theoretical approaches for addressing these challenges in order to (i) study natural resource allocation strategies in microorganisms and (ii) propose new synthetic control strategies for biotechnological applications. In order to address (i), we develop extended self-replicator models accounting for the cost of regulation and energy metabolism in bacterial cells. We study these models by a combination of analytical and numerical approaches to derive optimal control solutions and a control synthesis, dealing with the bang-bang-singular structure of the solutions. Moreover, we define quasi-optimal feedback control strategies inspired by known regulatory mechanisms in the cell. To test whether bacteria follow the predicted optimal strategies, we quantify dynamic resource allocation in the bacterium *Escherichia coli* by monitoring, by means of time-lapse fluorescent microscopy, the expression of selected genes in single cells growing in a microfluidics device. In order to address (ii), we build self-replicator models that include a pathway for the production of a metabolite of interest. We also add a mechanism to turn off microbial growth by means of an external input signal, at the profit of the production of the metabolite. We formulate the maximization of the amount of metabolite produced as an optimal control problem, and derive optimal solutions and a control synthesis, as well as quasi-optimal feedback strategies satisfying chemical and physical design constraints. The proposed synthetic control strategies are being tested experimentally by growing *E. coli* strains capable of producing glycerol from glucose in a mini-bioreactor system. We aim at quantifying the amount of glucose consumed and glycerol produced, in the case of a predefined input signal (open-loop control) and the adaptive regulation of the input signal based on on-line measurements of the growth rate and the expression of fluorescent reporters of selected genes (closed-loop control). Currently, one PhD (A. Yabo) and one postdoc (S. Maslovskaya) are involved in these tasks and jointly supervised by colleagues from McTAO and Biocore teams at Sophia. Preliminary results concern the definition on extended (higher dimensional) models for the bacteria dynamics, check of second order optimality conditions on the resulting optimal control problem, and study of the turnpike phenomenon for these optimization problems.

MEMPHIS Project-Team

4. Application Domains

4.1. Energy conversion

4.1.1. Fluid-structure interaction

We apply the methods developed in our team to the domain of wind engineering and sea-wave converters. In Figure 1, we show results of a numerical model for a sea-wave energy converter. We here rely on a monolithic model to describe the interaction between the rigid floater, air and water; material properties such as densities, viscosities and rigidity vary across the domain. The appropriate boundary conditions are imposed at interfaces that arbitrarily cross the grid using adapted schemes built thanks to geometrical information computed via level set functions [46]. The background method for fluid-structure interface is the volume penalization method [27] where the level set functions is used to improve the degree of accuracy of the method [3] and also to follow the object. The underlined mathematical model is unsteady, and three dimensional; numerical simulations based on a grid with $\mathcal{O}(10^8)$ degrees of freedom are executed in parallel using 512 CPUs .

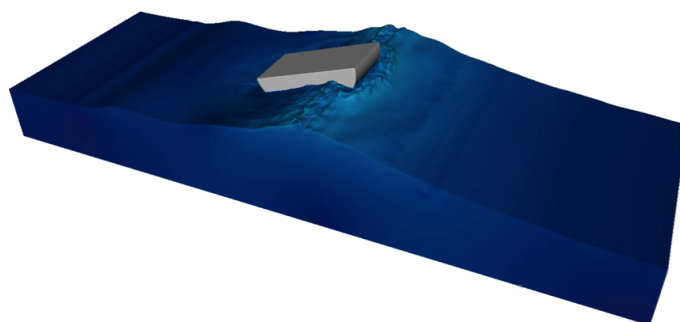


Figure 1. numerical modeling of a sea-wave converter by a monolithic model and Cartesian meshes.

In the context of the Aerogust (Aeroelastic gust modelling) European project, together with Valorem, we investigated the behavior of wind turbine blades under gust loading. The aim of the project was to optimize the design of wind turbine blades to maximize the power extracted. A meteorological mast (Figure 2 (a)) has been installed in March 2017 in Brittany to measure wind on-site: data provided by the mast have been exploited to initialize the mathematical model. Due to the large cost of the full-order mathematical model, we relied on a simplified model [38] to optimize the global twist. Then, we validated the optimal configuration using the full-order Cartesian model based on the NaSCar solver. Figure 2 (b) shows the flow around the optimized optimized wind turbine rotor.

4.1.2. Schemes for turbulent flow simulations using Octrees

We have initially developed and tested a 3D first-order Octree code for unsteady incompressible Navier-Stokes equations for full windmill simulations with an LES model and wall laws. We have validated this code on Occigen for complex flows at increasing Reynolds numbers. This step implied identifying stable and feasible schemes compatible with the parallel linear Octree structure. The validation has been conducted with respect to the results of a fully Cartesian code (NaSCAR) that we run on Turing (with significantly more degrees of freedom) and with respect to experimental results.



Figure 2. Aerogust project. Left: met mast after its installation. Right: flow around the optimized wind turbine rotor (as predicted by NaSCar).

Subsequently, we have developed a second-order Octree scheme that has been validated on Occigen for a sphere at a moderate Reynolds number ($Re = 500$), see Table 1 . Then, for a cylinder at ($Re = 140000$) (Figures 3 (a) and 3 (b)), close to real applications, we have preliminary validation results for the second-order scheme with respect to experimental drag coefficient (Table 2). Additional resources will be asked on Occigen to complete the study.

Table 1. Flow past a sphere at $Re = 500$. Results in the literature are spread between $C_D = 0.48$ and $C_D = 0.52$.

Mesh	Δx_{\min}	number of cells	C_D (1 st -order scheme)	C_D (2 nd -order scheme)
1	0.094	$0.72 \cdot 10^5$	N.A.	0.526
2	0.047	$4.9 \cdot 10^5$	0.595	0.522
3	0.023	$4.7 \cdot 10^6$	0.546	0.492
4	0.012	$37.6 \cdot 10^6$	0.555	0.496

Table 2. Flow past a sphere at $Re = 14000$.

Case	C_D
Octree, 1 st -order scheme	1.007
Octree, 2 nd -order scheme	1.157
Cartesian	1.188
Experimental estimate [34]	1.237

4.2. Vascular flows

A new research direction pursued by the team is the mathematical modelling of vascular blood flows in arteries. Together with the start-up Nurea (<http://nurea-soft.com/>) and the surgeon Eric Ducasse, we aim at developing reliable and automatic procedures for aneurysm segmentation and for the prediction of aneurysm rupture

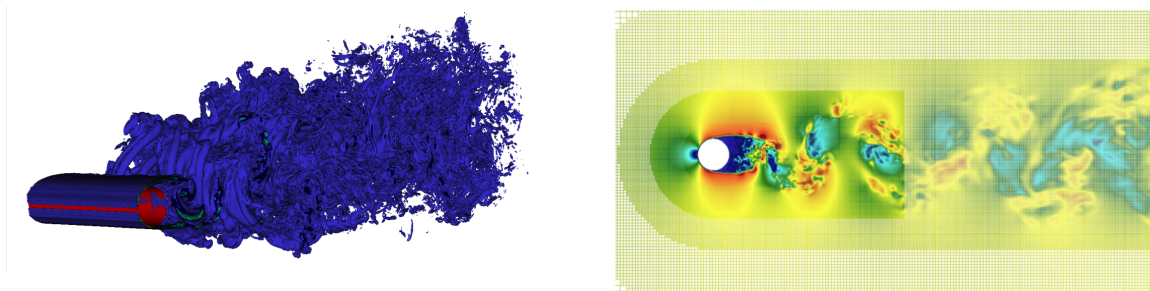


Figure 3. flow past a cylinder at $Re = 140000$. Left: vorticity contour lines. Right: streamwise velocity section and grid for the second-order Octree scheme.

risk. Our approach exploits two sources of information: (i) numerical simulations of blood flows in complex geometries, based on an octree discretization, and (ii) computed tomography angiography (CTA) data. Figure 4 shows the force distribution on the walls of the abdominal aorta in presence of an aneurysm; results are obtained using a parallelized hierarchical Cartesian scheme based on octrees.

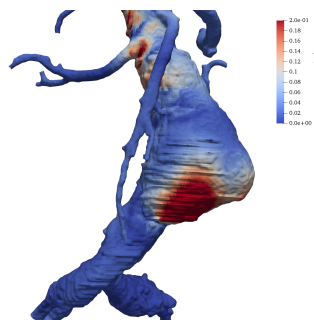


Figure 4. force distribution on the walls of the abdominal aorta in presence of an aneurysm.

4.3. Eulerian non-linear elasticity models

Mathematical and numerical modelling of continuum systems undergoing extreme regimes is challenging due to the presence of large deformations and displacements of the solid part, and due to the strongly non-linear behaviour of the fluid part. At the same time, proper experiments of impact phenomena are particularly dangerous and require expensive facilities, which make them largely impractical. For this reason, there is a growing interest in the development of predictive models for impact phenomena.

In MEMPHIS, we rely on a fully Eulerian approach based on conservation laws, where the different materials are characterized by their specific constitutive laws, to address these tasks. This approach was introduced in [37] and subsequently pursued and extended in [43], [36], [29], [33]. In Figure 5, we show the results of the numerical simulation of the impact of a copper projectile immersed in air over a copper shield. Results are obtained using a fully parallel monolithic Cartesian method, based on a 4000^2 fixed Cartesian grid. Simulations are performed on a cluster of 512 processors, and benefits from the isomorphism between grid partitioning and processor topology.

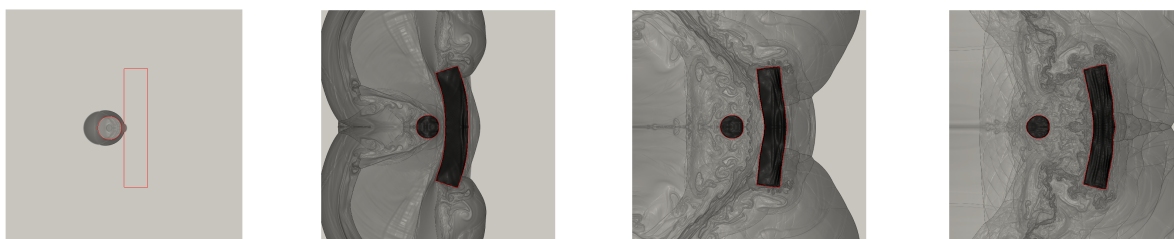


Figure 5. impact and rebound of a copper projectile on a copper plate. Interface and schlieren at $50\mu s$, $199\mu s$, $398\mu s$ and $710\mu s$.

MEPHYSTO Team (section vide)

MINGUS Project-Team

4. Application Domains

4.1. Application domains

The MINGUS project aims at applying the new numerical methods on realistic problems arising for instance in physics of nanotechnology and physics of plasmas. Therefore, in addition to efforts devoted to the design and the analysis of numerical methods, the inherent large size of the problems at hand requires advanced mathematical and computational methods which are hard to implement. Another application is concerned with population dynamics for which the main goal is to understand how the spatial propagation phenomena affect the demography of a population (plankton, parasite fungi, ...). Our activity is mostly at an early stage in the process of transfer to industry. However, all the models we use are physically relevant and all have applications in many areas (ITER, Bose-Einstein condensate, wave turbulence, optical tomography, transport phenomena, population dynamics, ...). As a consequence, our research aims at reaching theoretical physicists or computational scientists in various fields who have strong links with industrial applications. In order to tackle as realistic physical problems as possible, a fundamental aspect will consist in working on the realization of numerical methods and algorithms which are able to make an efficient use of a large number of processors. Then, it is essential for the numerical methods developed in the MINGUS project to be thought through this prism. We will benefit from the strong expertise of P. Navaro in scientific computing and more precisely on the Selalib software library (see description below). Below, we detail our main applications: first, the modeling and numerical approximation of magnetized plasmas is our major application and will require important efforts in terms of software developments to scale-up our multiscale methods; second, the transport of charged particles in nanostructures has very interesting applications (like graphene material), for which our contributions will mainly focus on dedicated problems; lastly, applications on population dynamics will be dedicated to mathematical modeling and some numerical validations.

4.1.1. Plasmas problems

The Selalib (SEmi-LAgrangian LIBrary) software library⁰ is a modular library for kinetic and gyrokinetic simulations of plasmas in fusion energy devices. Selalib is a collection of fortran modules aimed at facilitating the development of kinetic simulations, particularly in the study of turbulence in fusion plasmas. Selalib offers basic capabilities and modules to help parallelization (both MPI and OpenMP), as well as pre-packaged simulations. Its main objective is to develop a documented library implementing several numerical methods for the numerical approximation of kinetic models. Another objective of the library is to provide physicists with easy-to-use gyrokinetic solvers. It has been originally developed by E. Sonnendrücker and his collaborators in the past CALVI Inria project, and has played an important role in the activities of the IPL FRATRES. P. Navaro is one of the main software engineer of this library and as such he played an important daily role in its development and its portability on supercomputers. Though Selalib has reached a certain maturity some additional works are needed to make available by the community. There are currently discussions for a possible evolution of Selalib, namely the writing of a new release which will be available for free download. Obviously, the team will be involved in this process. At the scientific level, Selalib is of great interest for us since it provides a powerful tool with which we can test, validate and compare our new methods and algorithms (users level). Besides numerical algorithms the library provides low-level utilities, input-output modules as well as parallelization strategies dedicated to kinetic problems. Moreover, a collection of simulations for typical test cases (of increasing difficulties) with various discretization schemes supplements the library. This library turns out to be the ideal complement of our activities and it will help us to scale-up our numerical methods to high-dimensional kinetic problems. During the last years, several experiments have been successfully performed in this direction (especially with PhD students) and it is important for us that this approach remains throughout. Then, we intend to integrate several of the numerical methods developed by the team within the Selalib library,

⁰SELALIB, <http://selalib.gforge.inria.fr>

with the strong help of P. Navaro (contributors level). This work has important advantages: (i) it will improve our research codes (in terms of efficiency but also of software maintenance point of view); (ii) it will help us to promote our research by making our methods available to the research community.

4.1.2. Quantum problems

Nowadays, a great challenge consists in the downscaling at the nanometer scale of electronic components in order to improve speed and efficiency of semiconductor materials. In this task, modeling and numerical simulations play an important role in the determination of the limit size of the nanotransistors. At the nanoscale, quantum effects have to be considered and the Schrödinger equation is prominent equation in this context. In the so-called semiclassical regime or when the transport is strongly confined, the solution endows space-time highly oscillations which are very difficult to capture numerically. An important application is the modeling of charged particles transport in graphene. Graphene is a sheet of carbone made of a single layer of molecule, organized in a bidimensional honeycomb crystal. The transport of charged particles in this structure is usually performed by Dirac equation (which is the relativistic counterpart of the Schrödinger equation). Due to the unusual properties of graphene -at room temperature, electrons moving in graphene behave as massless relativistic particles- physicists and compagnies are nowadays actively studying this material. Here, predicting how the material properties are affected by the uncertainties in the hexagonal lattice structure or in external potentials, is a major issue.

4.1.3. Population dynamics

The main goal is to characterize how spatial propagation phenomena (diffusion, transport, advection, . . .) affect the time evolution of the demography of a population. In collaboration with Y. Lagadeuc (ECOBIO, Rennes), this question has been studied for plankton. In this context, mathematical models have been proposed and it has been shown that the spatial dynamic (in this context, due to the marine current) which is fast compared to demographic scales, can strongly modify the demographic evolution of the plankton. In collaboration with Ecole d'Agronomie de Rennes, a mathematical study on the demography of a parasite fungi of plants has been performed. In this context, the demography is specific: the fungi can proliferate through sexual reproduction or through parthenogenesis. This two ways of reproduction give rise mathematically to quadratic and linear growth rates with respect to the population variable. The demography is then coupled with transport (transport of fungi spore by wind). Here, the goal is characterize the propagation of the fungi population by finding travelling waves solutions which are well adapted to describe the evolution of invasive fronts. Moreover, this approach enables to recover with a good agreement realistic examples (infection of ash or banana tree) for which experimental data are available. In these contexts, mathematical models are a powerful tool for biologists since measurements are very complicated to obtain and laboratory experiments hardly reproduce reality. The models derived are multiscale due to the nature of the underlying phenomena and the next step is to provide efficient numerical schemes.

MISTIS Project-Team

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Veronica Munoz Ramirez, Florence Forbes, Stephane Girard, Hongliang Lu, Fabien Boux, Benoit Kugler, Alexandre Constantin.

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 and functional MRI data, in collaboration with the Grenoble Institute of Neuroscience (GIN). We also consider other statistical 2D fields coming from other domains such as remote sensing, in collaboration with the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) and the Centre National d'Etudes Spatiales (CNES). In this context, we worked on hyperspectral and/or multitemporal images. In the context of the "pole de compétitivité" project I-VP, we worked on images of PC Boards.

4.2. Biology, Environment and Medicine

Participants: Alexis Arnaud, Florence Forbes, Stephane Girard, Jean-Baptiste Durand, Julyan Arbel, Brice Olivier, Karina Ashurbekova, Fabien Boux, Veronica Munoz Ramirez, Fei Zheng.

A third domain of applications concerns biology and medicine. We considered the use of mixture models to identify biomarkers. We also investigated statistical tools for the analysis of fluorescence signals in molecular biology. Applications in neurosciences are also considered. In the environmental domain, we considered the modelling of high-impact weather events and the use of hyperspectral data as a new tool for quantitative ecology.

MODAL Project-Team

4. Application Domains

4.1. Economic world

The Modal team applies its research to the economic world through CIFRE Phd supervision such as CACF (credit scoring), A-Volute (expert in 3D sound), Meilleur Taux (insurance comparator), Worldline. It also has several contracts with companies such as COLAS, Nokia-Apsys/Airbus.

4.2. Biology

The second main application domain of the team is the biology. Members of the team are involved in the supervision and scientific animation of the bilille platform, the bioinformatics and bioanalysis platform and OncoLille project of Lille.

MOKAPLAN Project-Team (section vide)

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 method due to Yee [61]. 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 [8].

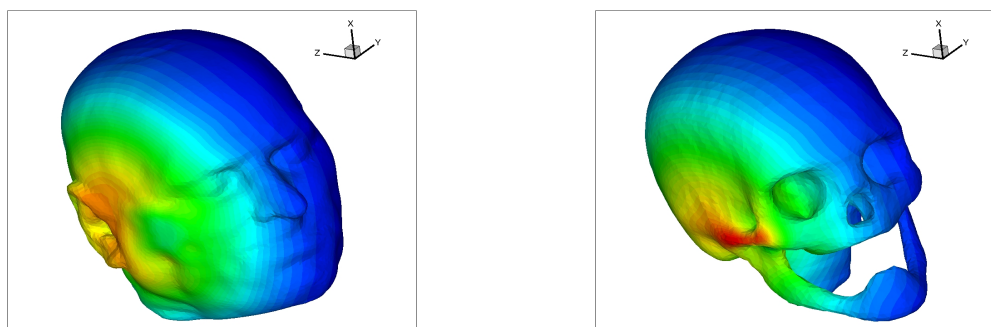


Figure 1. Exposure of head tissues to an electromagnetic wave emitted by a localized source. Top figures: surface triangulations of the skin and the skull. Bottom figures: contour lines of the amplitude of the electric field.

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 plasmomic 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 non-linear

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 [53]- [62].

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 as the Drude and Drude-Lorentz models. Here again, the FDTD method is a widely used approach for solving the resulting system of PDEs [58]. 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 [56]- [57]- [51]. 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.

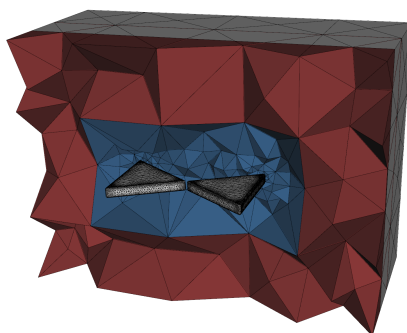


Figure 2. Simulation of the field enhancement at the tip of a gold bowtie nanoantenna (PhD thesis of Jonathan Viquerat).

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 for 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.

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 [59] and based on the first order velocity-stress hyperbolic system of elastic waves equations, which is an extension of the scheme derived by Yee [61] 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 collaborated with CETE Méditerranée which is a regional technical and engineering centre whose activities are concerned with seismic hazard assessment studies, and IFSTTAR (<https://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.

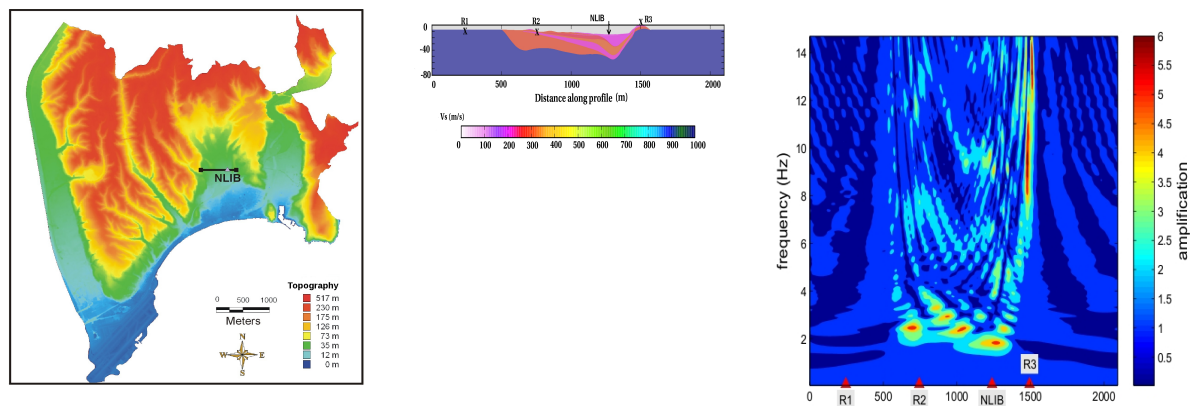


Figure 3. Propagation of a plane wave in a heterogeneous model of Nice area (provided by CETE Méditerranée).

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 is considered 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 Team (section vide)

NECS 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.
- Inertial navigation and 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.
- Collaborative indoor and outdoor navigation of pedestrians.

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

4.2. Intelligent transportation systems

Throughout the world, roadways are notorious for their congestion, from dense urban network to large freeway systems. This situation tends to get worse over time due to the continuous increase of transportation demand whereas public investments are decreasing and space is lacking to build new infrastructures. The most obvious impact of traffic congestion for citizens is the increase of travel times and fuel consumption. Another critical effect is that infrastructures are not operated at their capacity during congestion, implying that fewer vehicles are served than the amount they were designed for. Using macroscopic fluid-like models, the NECS team has initiated new researches to develop innovative traffic management policies able to improve the infrastructure operations. The research activity is on two main challenges: (1) modeling and forecasting, so as to provide accurate information to users, e.g., travel times; and (2) control, via ramp-metering and/or variable speed limits. The Grenoble Traffic Lab (see <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/>)

4.3. Inertial navigation

The team is exploring techniques and approaches from estimation, filtering and machine learning, in order to use inertial sensor units in pedestrian navigation, attitude estimation, transportation modes and human activities recognition. These units are composed of accelerometers, magnetometers and gyroscopes, sensors that we find usually in smartphones, tablets and smartwatches. This area of research in the team will evolve towards multimodal navigation, cooperative and collaborative navigation in indoor and outdoor environments.

POEMS Project-Team (section vide)

QUANTIC Project-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.

RANDOPT Project-Team

4. Application Domains

4.1. Application Domains

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have a great need to apply black-box algorithms on a daily basis. We do not target a specific application domain and are interested in possible black-box applications stemming from various origins. This is for us intrinsic to the nature of the methods we develop that are general purpose algorithms. Hence our strategy with respect to applications can be seen as opportunistic and our main selection criteria when approached by colleagues who want to develop a collaboration around an application is whether we judge the application interesting: that is the application brings new challenges and/or gives us the opportunity to work on topics we already intended to work on.

The concrete applications related to industrial collaborations we are currently dealing with are:

- With Thales for the theses of Konstantinos Varelas and Paul Dufossé (DGA-CIFRE theses) related to the design of radars (shape optimization of the wave form). Those theses investigate the development of large-scale variants of CMA-ES and constrained-handling for CMA-ES.
- With Storengy, a subsidiary of Engie specialized in gas storage for the thesis of Cheikh Touré. Different multiobjective applications are considered in this context but the primary motivation of Storengy is to get at their disposal a better multiobjective variant of CMA-ES which is the main objective of the developments within the thesis.
- With PSA in the context of the OpenLab and the thesis of Marie-Ange Dahito for the design of part of a car body.
- With Onera in the context of the thesis of Alann Cheral related to the optimization of the choice of hyperspectral bandwidth.

RAPSODI Project-Team

4. Application Domains

4.1. Porous media flows

Porous media flows are of great interest in many contexts, like, e.g., oil engineering, water resource management, nuclear waste repository management, or carbon dioxide sequestration. We refer to [73], [72] for an extensive discussion on porous media flow models.

From a mathematical point of view, the transport of complex fluids in porous media often leads to possibly degenerate parabolic conservation laws. The porous rocks can be highly heterogeneous and anisotropic. Moreover, the grids on which one intends to solve numerically the problems are prescribed by the geological data, and might be non-conformal with cells of various shapes. Therefore, the schemes used for simulating such complex flows must be particularly robust.

4.2. Corrosion and concrete carbonation

The team is interested in the theoretical and numerical analysis of mathematical models describing degradation of materials as concrete carbonation and corrosion. The study of such models is an important environmental and industrial issue. Atmospheric carbonation degrades reinforced concretes and limits the lifetime of civil engineering structures. Corrosion phenomena issues occur for instance in the reliability of nuclear power plants and the nuclear waste repository. The study of the long time evolution of these phenomena is of course fundamental in order to predict the lifetime of the structures.

From a mathematical point of view, the modeling of concrete carbonation (see [67]) as the modeling of corrosion in an underground repository (DPCM model developed by Bataillon *et al.* [71]) lead to systems of PDEs posed on moving domains. The coupling between convection-diffusion-reaction equations and moving boundary equations leads to challenging mathematical questions.

4.3. Complex fluid flows

The team is interested in some numerical methods for the simulation of systems of PDEs describing complex flows, like for instance, mixture flows, granular gases, rarefied gases, or quantum fluids.

Variable-density, low-Mach flows have been widely studied in the recent literature because of their applicability in various phenomena such as flows in high-temperature gas reactors, meteorological flows, flows with convective and/or conductive heat transfer or combustion processes. In such cases, the resolution of the full compressible Navier–Stokes system is not adapted, because of the sound waves speed. The Boussinesq incompressible model is not a better alternative for such low-speed phenomena, because the compressibility effects can not be totally cancelled due to large variations of temperature and density. Consequently, some models have been formally derived, leading to the filtering of the acoustic waves by the use of some formal asymptotic expansions and two families of methods have been developed in the literature in order to compute these flows. We are interested in particular in the so-called pressure-based methods which are more robust than density-based solvers, although their range of validity is in general more limited.

Kinetic theory of molecular gases models a gas as a system of elastically colliding spheres, conserving mechanical energy during impact. Once initialized, it takes a molecular gas not more than few collisions per particle to relax to its equilibrium state, characterized by a Maxwellian velocity distribution and a certain homogeneous density (in the absence of external forces). A granular gas is a system of dissipatively colliding, macroscopic particles (grains). This slight change in the microscopic dynamics (converting energy into heat) causes drastic changes in the behavior of the gas: granular gases are open systems, which exhibit self-organized spatio-temporal cluster formations, and have no equilibrium distribution. They can be used to model silos, avalanches, pollen or planetary rings.

The quantum models can be used to describe superfluids, quantum semiconductors, weakly interacting Bose gases or quantum trajectories of Bohmian mechanics. They have attracted considerable attention in the last decades, due in particular to the development of the nanotechnology applications. To describe quantum phenomena, there exists a large variety of models. In particular there exist three different levels of description: microscopic, mesoscopic and macroscopic. The quantum Navier–Stokes equations deal with a macroscopic description in which the quantum effects are taken into account through a third order term called the quantum Bohm potential. This Bohm potential arises from the fluid dynamical formulation of the single-state Schrödinger equation. The non-locality of quantum mechanics is approximated by the fact that the equations of state do not only depend on the particle density but also on its gradient. These equations were employed to model field emissions from metals and steady-state tunneling in metal- insulator- metal structures and to simulate ultra-small semiconductor devices.

4.4. Stratigraphy

The knowledge of the geology is a prerequisite before simulating flows within the subsoil. Numerical simulations of the geological history thanks to stratigraphy numerical codes allow to complete the knowledge of the geology where experimental data are lacking. Stratigraphic models consist in a description of the erosion and sedimentation phenomena at geological scales.

The characteristic time scales for the sediments are much larger than the characteristic time scales for the water in the river. However, the (time-averaged) water flux plays a crucial role in the evolution of the stratigraphy. Therefore, defining appropriate models that take the coupling between the rivers and the sediments into account is fundamental and challenging. Once the models are at hand, efficient numerical methods must be developed.

4.5. Low-frequency electromagnetism

Numerical simulation is nowadays an essential tool in order to design electromagnetic systems, by estimating the electromagnetic fields generated in a wide variety of devices. An important challenge for many applications is to quantify the intensity of the electric field induced in a conductor by a current generated in its neighborhood. In the low-frequency regime, we can for example quote the study of the impact on the human body of a high-tension line or, for higher frequencies, the one of a smartphone. But the ability to simulate accurately some electromagnetic fields is also very useful for non-destructive control, in the context of the maintenance of nuclear power stations for example. The development of efficient numerical tools, among which *a posteriori* error estimators, is consequently necessary to reach a high precision of calculation in order to provide estimations as reliable as possible.

REALOPT Project-Team

4. Application Domains

4.1. 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 [54]. 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-electronics system installation cost. We developed and compared several decomposition approaches [79], [78], [77] 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 [76]. 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 [59], [60].

The Dial-a-Ride Problem is a variant of the pickup and delivery problem with time windows, where the user inconvenience must be taken into account. In [69], ride time and customer waiting time are modeled through both constraints and an associated penalty in the objective function. We develop a column generation approach, dynamically generating feasible vehicle routes. Handling ride time constraints explicitly in the pricing problem solver requires specific developments. Our dynamic programming approach for pricing problem makes use of a heuristic dominance rule and a heuristic enumeration procedure, which in turns implies that our overall branch-and-price procedure is a heuristic. However, in practice our heuristic solutions are experimentally very close to exact solutions and our approach is numerically competitive in terms of computation times.

In [67], [66], we consider the problem of covering an urban area with sectors under additional constraints. We adapt the aggregation method to our column generation algorithm and focus on the problem of disaggregating the dual solution returned by the aggregated master problem.

We studied several time dependent formulations for the unit demand vehicle routing problem [44], [43]. 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 [75] 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 algorithms based on the column generation approach. These algorithms have been tested on a set of real-life instances coming from a real Russian freight transportation company. Our algorithms have been faster on these instances than the current solution approach being used by the company.

4.2. 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 [74]. 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 [33] 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 [62]: we faced a multiple-class integer knapsack problem with setups [61] (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 [56], [55], [58], [57]. 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 [38]. 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 that are used as raw material. There are extra sequencing constraints regarding the production that make the problem even more complex.

We have also organized a European challenge on packing with society Renault. This challenge was about loading trucks under practical constraints.

4.3. 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 branch-and-price algorithm where periodic plans are generated for vehicles by solving a multiple choice knapsack subproblem, and the global planning of customer visits is coordinated by the master program [63]. 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 [64].

Together with our partner company GAPSO from the associate team SAMBA, we worked on the equipment routing task scheduling problem [68] arising during port operations. In this problem, a set of tasks needs to be performed using equipments of different types with the objective to maximize the weighted sum of performed tasks.

We participated to the project on an airborne radar scheduling. For this problem, we developed fast heuristics [53] and exact algorithms [35]. A substantial research has been done on machine scheduling problems. A new compact MIP formulation was proposed for a large class of these problems [34]. An exact decomposition algorithm was developed for the NP-hard maximizing the weighted number of late jobs problem on a single machine [70]. A dominant class of schedules for malleable parallel jobs was discovered in the NP-hard problem to minimize the total weighted completion time [72]. We proved that a special case of the scheduling problem at cross docking terminals to minimize the storage cost is polynomially solvable [73], [71].

Another application area in which we have successfully developed MIP approaches is in the area of tactical production and supply chain planning. In [32], we proposed a simple heuristic for challenging multi-echelon problems that makes effective use of a standard MIP solver. [31] 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 [39] provide demonstrably stronger formulations for some problem classes than any previously proposed. We are now working on planning phytosanitary treatments in vineries.

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 [49], [48], 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 [40]. We considered train timetabling problems and their re-optimization after a perturbation in the network [51], [50]. 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 [45], [46], [47]. 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 [52]. 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 [41].

4.4. Resource Allocation for High Performance and Cloud Computing

In the context of numerical simulations on high performance machines, optimizing data locality and resource usage is very important for faster execution times and lower energy consumption. This optimization can be seen as a special case of scheduling problem on parallel resource, with several challenges. First, instances are

typically large: a large matrix factorization (with 50×50 blocks) involves about $30 \cdot 10^3$ tasks. Then, HPC platforms consist of heterogeneous and unrelated resources, what is known to make scheduling problems hard to approximate. Finally, due to co-scheduling effects and shared communication resources, it is not realistic to accurately model the exact duration of tasks. All these observations make it impossible to rely on static optimal solutions, and HPC applications have gone from simple generic static allocations to runtime dynamic scheduling strategies that make their decisions based on the current state of the platform (the location of input data), the expected transfer and running times for the tasks, and some affinity and priority information that have possibly been computed offline. In this context, we are strongly involved in the design of scheduling strategies for the StarPU runtime, with two goals: proving that it is possible to design approximation algorithms whose complexity is extremely small (typically sub-linear in the number of ready tasks), and show that they can be used in practice with good performance results. We are pursuing collaborations both with teams developing the StarPU system (Storm) by designing algorithms for the generic scheduling problems [37], and with teams developing linear algebra algorithms over the runtime (Hiepac), by proposing specialized algorithms for specific cases. For example, in the case of linear algebra applications on heterogeneous platforms, we have considered the combinatorial optimization problem associated to matrix multiplication, that is amenable to partitioning the unit square into zones of prescribed areas while minimizing the overall size of the boundaries. We have improved the best known approximation ratio to 1.15 in [36] and we have shown that the resulting distribution schemes can indeed be used to design efficient implementations using StarPU in [42].

SEQUEL Project-Team

4. Application Domains

4.1. Sequential decision making under uncertainty and prediction

The spectrum of applications of our research is very wide: it ranges from the core of our research, that is sequential decision making under uncertainty, to the application of components used to solve this decision making problem.

To be more specific, we work on computational advertising and recommendation systems; these problems are considered as a sequential matching problem in which resources available in a limited amount have to be matched to meet some users' expectations. The sequential approach we advocate paves the way to better tackle the cold-start problem, and non stationary environments. More generally, these approaches are applied to the optimization of budgeted resources under uncertainty, in a time-varying environment, including constraints on computational times (typically, a decision has to be made in less than 1 ms in a recommendation system). An other field of application of our research is related to education which we consider as a sequential matching problem between a student, and educational contents.

The algorithms to solve these tasks heavily rely on tools from machine learning, statistics, and optimization. Henceforth, we also apply our work to more classical supervised learning, and prediction tasks, as well as unsupervised learning tasks. The whole range of methods is used, from decision forests, to kernel methods, to deep learning. For instance, we have recently used deep learning on images. We also have a line of work related to software development studying how machine learning can improve the quality of software being developed. More generally, we apply our research to data science.

SIERRA Project-Team

4. Application Domains

4.1. Applications for Machine Learning

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

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

SIMSMART Project-Team

4. Application Domains

4.1. Domain 1 – Computational Physics

The development of large-scale computing facilities has enabled simulations of systems at the *atomistic scale* on a daily basis. The aim of these simulations is to bridge the time and space scales between the macroscopic properties of matter and the stochastic atomistic description. Typically, such simulations are based on the ordinary differential equations of classical mechanics supplemented with a random perturbation modeling temperature, or collisions between particles.

Let us give a few examples. In bio-chemistry, such simulations are key to predict the influence of a ligand on the behavior of a protein, with applications to drug design. The computer can thus be used as a *numerical microscope* in order to access data that would be very difficult and costly to obtain experimentally. In that case, a rare event (Objective 1) is given by a macroscopic system change such as a conformation change of the protein. In nuclear safety, such simulations are key to predict the transport of neutrons in nuclear plants, with application to assessing aging of concrete. In that case, a rare event is given by a high energy neutron impacting concrete containment structures.

A typical model used in molecular dynamics simulation of open systems at given temperature is a stochastic differential equation of Langevin type. The large time behavior of such systems is typically characterized by a hopping dynamics between 'metastable' configurations, usually defined by local minima of a potential energy. In order to bridge the time and space scales between the atomistic level and the macroscopic level, specific algorithms enforcing the realization of rare events have been developed. For instance, splitting particle methods (Objective 1) have become popular within the computational physics community only within the last few years, partially as a consequence of interactions between physicists and Inria mathematicians in ASPI (parent of SIMSMART) and MATHERIALS project-teams.

4.2. Domain 2 – Meteorology

The traditional trend in data assimilation in geophysical sciences (climate, meteorology) is to use as prior information some very complex deterministic models formulated in terms of fluid dynamics and reflecting as much as possible the underlying physical phenomenon (see *e.g.* <https://www.metoffice.gov.uk/research/modelling-systems/unified-model/>). Weather/climate forecasting can then be recast in terms of a Bayesian filtering problem (see Objective 2) using weather observations collected *in situ*.

The main issue is therefore to perform such Bayesian estimations with very expensive infinite dimensional prior models, and observations in large dimension. The use of some linear assumption in prior models (Kalman filtering) to filter non-linear hydrodynamical phenomena is the state-of-the-art approach, and a current field of research, but is plagued with intractable instabilities.

This context motivates two research trends: (i) the introduction of non-parametric, model-free prior dynamics constructed from a large amount of past, recorded real weather data; and (ii) the development of appropriate non-linear filtering approaches (Objective 2 and Objective 3).

SIMSMART will also test its new methods on multi-source data collected in North-Atlantic paying particular attention to coastal areas (*e.g.* within the inter-Labex SEACS).

SPHINX Project-Team

4. Application Domains

4.1. Robotic swimmers

Some companies aim at building biomimetic robots that can swim in an aquarium, as toys but also for medical purposes. An objective of Sphinx is to model and to analyze several models of these robotic swimmers. For the moment, we focus in the motion of a nanorobot. In that case, the size of the swimmers leads to neglect the inertia forces and to only consider the viscosity effects. Such nanorobots could be used for medical purposes to introduce some medicine or perform small surgical operations. In order to get a better understanding of such robotic swimmers, we have obtained control results via shape changes and we have developed simulation tools (see [71], [70], [101], [98]). Among all the important issues, we aim to consider the following ones:

1. Solve the control problem by limiting the set of admissible deformations.
2. Find the “best” location of the actuators, in the sense of being the closest to the exact optimal control.

The main tools for this investigation are the 3D codes that we have developed for simulation of fish in a viscous incompressible fluid (SUSHI3D) or in a inviscid incompressible fluid (SOLEIL).

4.2. Aeronautics

We will develop robust and efficient solvers for problems arising in aeronautics (or aerospace) like electromagnetic compatibility and acoustic problems related to noise reduction in an aircraft. Our interest for these issues is motivated by our close contacts with companies like Airbus or “Thales Systèmes Aéroportés”. We will propose new applications needed by these partners and assist them in integrating these new scientific developments in their home-made solvers. In particular, in collaboration with C. Geuzaine (Université de Liège), we are building a freely available parallel solver based on Domain Decomposition Methods that can handle complex engineering simulations, in terms of geometry, discretization methods as well as physics problems, see <http://onelab.info/wiki/GetDDM>.

TAU Project-Team

4. Application Domains

4.1. Computational Social Sciences

Participants: Philippe Caillou, Isabelle Guyon, Michèle Sebag, Paola Tubaro

Collaboration: Jean-Pierre Nadal (EHESS); Marco Cuturi, Bruno Crépon (ENSAE); Thierry Weil (Mines); Jean-Luc Bazet (RITM)

Computational Social Sciences (CSS) studies social and economic phenomena, ranging from technological innovation to politics, from media to social networks, from human resources to education, from inequalities to health. It combines perspectives from different scientific disciplines, building upon the tradition of computer simulation and modeling of complex social systems [102] on the one hand, and data science on the other hand, fueled by the capacity to collect and analyze massive amounts of digital data.

The emerging field of CSS raises formidable challenges along three dimensions. Firstly, the definition of the research questions, the formulation of hypotheses and the validation of the results require a tight pluridisciplinary interaction and dialogue between researchers from different backgrounds. Secondly, the development of CSS is a touchstone for ethical AI. On the one hand, CSS gains ground in major, data-rich private companies; on the other hand, public researchers around the world are engaging in an effort to use it for the benefit of society as a whole [124]. The key technical difficulties related to data and model biases, and to self-fulfilling prophecies have been discussed in section 3.1. Thirdly, CSS does not only regard scientists: it is essential that the civil society participate in the science of society [152].

TAO was involved in CSS for the last five years, and its activities have been strengthened thanks to P. Tubaro's and I. Guyon's expertises respectively in sociology and economics, and in causal modeling. Details are given in Section 7.3.

4.2. Energy Management

Participants: Isabelle Guyon, Marc Schoenauer, Michèle Sebag

PhD: Victor Berger, Benjamin Donnot, Balthazar Donon, Herilalaina Rakotoarison

Collaboration: Antoine Marot, Patrick Panciatici (RTE), Vincent Renault (Artelys)

Energy Management has been an application domain of choice for TAO since the end 2000s, with main partners SME Artelys (METIS Ilab Inria; ADEME project POST; on-going ADEME project NEXT), RTE (See.4C European challenge; two CIFRE PhDs), and, since Oct. 2019, IFPEN. The goals concern i) optimal planning over several spatio-temporal scales, from investments on continental Europe/North Africa grid at the decade scale (POST), to daily planning of local or regional power networks (NEXT); ii) monitoring and control of the French grid enforcing the prevention of power breaks (RTE); iii) improvement of house-made numerical methods using data-intense learning in all aspects of IFPEN activities (as described in Section 3.2).

Optimal planning over long periods of time amounts to optimal sequential decision under high uncertainties, ranging from stochastic uncertainties (weather, market prices, demand prediction) handled based on massive data, to non-stochastic uncertainties (e.g., political decisions about the nuclear policy) handled through defining and selecting a tractable number of scenarios. Note that non-anticipativity constraints forbid the use of dynamic programming-related methods; this led to propose the *Direct Value Search* method [77] at the end of the POST project.

The daily maintenance of power grids requires the building of approximate predictive models on the top of any given network topology. Deep Networks are natural candidates for such modelling, considering the size of the French grid (~ 10000 nodes), but the representation of the topology is a challenge when, e.g. the RTE goal is to quickly ensure the "n-1" security constraint (the network should remain safe even if any of the 10000 nodes fails). Existing simulators are too slow to be used in real time, and the size of actual grids makes it intractable to train surrogate models for all possible (n-1) topologies (see Section 7.4 for more details).

Furthermore, predictive models of local grids are based on the estimated consumption of end-customers: Linky meters only provide coarse grain information due to privacy issues, and very few samples of fine-grained consumption are available (from volunteer customers). A first task is to transfer knowledge from small data to the whole domain of application. A second task is to directly predict the peaks of consumption based on the user cluster profiles and their representativity (see Section 7.4.2).

4.3. Data-driven Numerical Modeling

Participants: Alessandro Bucci, Guillaume Charpiat, Cécile Germain, Isabelle Guyon, Flora Jay, Marc Schoenauer, Michèle Sebag

PhD and Post-doc: Victor Estrade, Loris Felardos, Adrian Pol, Théophile Sanchez, Wenzhuo Liu

Collaboration: D. Rousseau (LAL), M. Pierini (CERN)

As said (section 3.2), in domains where both first principle-based models and equations, and empirical or simulated data are available, their combined usage can support more accurate modelling and prediction, and when appropriate, optimization, control and design. This section describes such applications, with the goal of improving the time-to-design chain through fast interactions between the simulation, optimization, control and design stages. The expected advances regard: i) the quality of the models or simulators (through data assimilation, e.g. coupling first principles and data, or repairing/extending closed-form models); ii) the exploitation of data derived from different distributions and/or related phenomena; and, most interestingly, iii) the task of optimal design and the assessment of the resulting designs.

The proposed approaches are based on generative and adversarial modelling [121], [106], extending both the generator and the discriminator modules to take advantage of the domain knowledge.

A first challenge regards the design of the model space, and the architecture used to enforce the known domain properties (symmetries, invariance operators, temporal structures). When appropriate, data from different distributions (e.g. simulated vs real-world data) will be reconciled, for instance taking inspiration from real-valued non-volume preserving transformations [85] in order to preserve the natural interpretation.

Another challenge regards the validation of the models and solutions of the optimal design problems. The more flexible the models, the more intensive the validation must be, as reminded by Leon Bottou. Along this way, generative models will be used to support the design of "what if" scenarios, to enhance anomaly detection and monitoring via refined likelihood criteria.

In the application case of dynamical systems such as fluid mechanics, the goal of incorporating machine learning into classical simulators is to speed up the simulations. Many possible tracks are possible for this; for instance one can search to provide better initialization heuristics to solvers (which make sure that physical constraints are satisfied, and which are responsible of most of the computational complexity of simulations) at each time step; one can also aim at predicting directly the state at $t + 100$, for instance, or at learning a representation space where the dynamics are linear (Koopman - von Neumann). The topic is very active in the deep learning community. To guarantee the quality of the predictions, concepts such as Liapunov coefficients (which express the speed at which simulated trajectories diverge from the true ones) can provide a suitable theoretical framework.

TOSCA 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 involve 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 behavior; 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 solar, 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

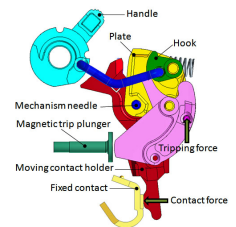
Diffusion hitting times are of great interest in finance (a typical example is the study of barrier options) and also in Geophysics and Neurosciences. On the one hand, analytic expressions for hitting time densities are well known and studied only in some very particular situations (essentially in Brownian 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.

TRIPPO Project-Team

4. Application Domains

4.1. Domain 1

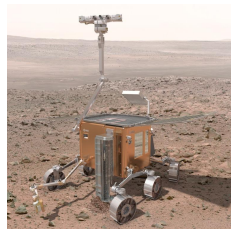
Nonsmooth dynamical systems arise in a lot of application fields. *We briefly expose here some applications that have been treated in the BIPOP team and that we will continue in the TRIPOP team, as a validation for the research axes and also in terms of transfer.* In mechanics, the main instances of nonsmooth dynamical systems are multibody systems with Signorini's unilateral contact, set-valued (Coulomb-like) friction and impacts, or in continuum mechanics, ideal plasticity, fracture or damage. Some illustrations are given in Figure 5 (a-f). Other instances of nonsmooth dynamical systems can also be found in electrical circuits with ideal components (see Figure 5 (g)) and in control theory, mainly with sliding mode control and variable structure systems (see Figure 5 (h)). More generally, every time a piecewise, possibly set-valued, model of systems is invoked, we end up with a nonsmooth system. This is the case, for instance, for hybrid systems in nonlinear control or for piecewise linear modeling of gene regulatory networks in mathematical biology (see Figure 5 (i)). Another common example of nonsmooth dynamics is also found when the vector field of a dynamical system is defined as a solution of an optimization problem under constraints, or a variational inequality. Examples of this kind are found in the optimal control theory, in dynamic Nash equilibrium or in the theory of dynamic flows over networks.



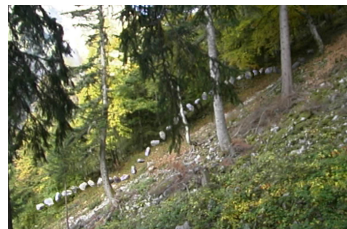
Circuit breakers mechanisms [38]



Granular flows

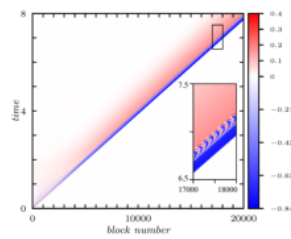
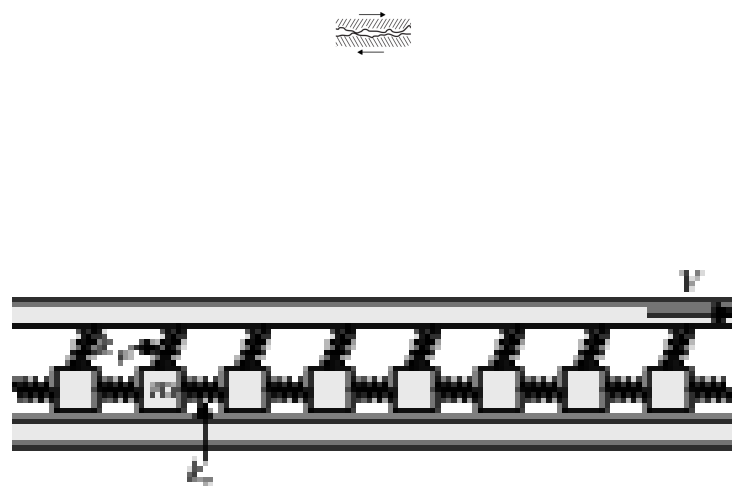


Robots (ESA ExoMars Rover [32])

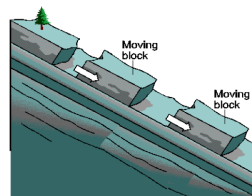


Rockfall [44], [43], [58]

Figure 3. Application fields of nonsmooth dynamics (mechanics)

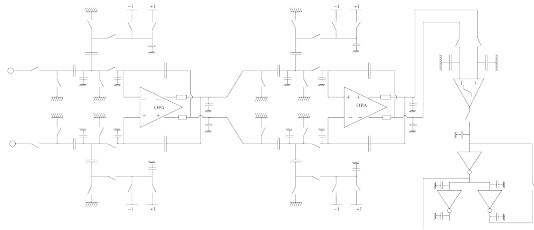


Frictional interface and solitary waves in the Burridge-Knopoff model [80]

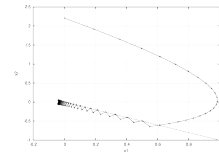
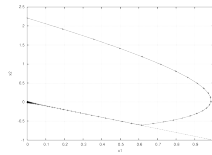
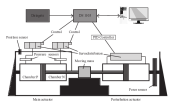


Sliding blocks

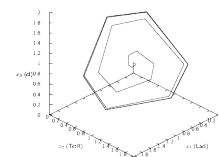
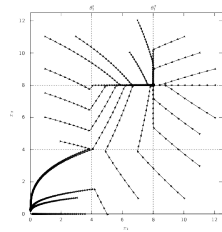
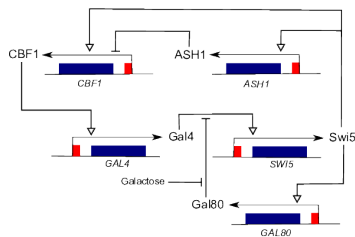
Figure 4. Application fields of nonsmooth dynamics (waves)



Switched electrical circuits (delta-sigma converter) [26]



Sliding mode control [29], [31], [69], [70], [79]



Gene regulatory networks [35]

Figure 5. Application fields of nonsmooth dynamics

TROPICAL Project-Team

4. Application Domains

4.1. Discrete event systems (manufacturing systems, networks)

One important class of applications of max-plus algebra comes from discrete event dynamical systems [56]. In particular, modelling timed systems subject to synchronization and concurrency phenomena leads to studying dynamical systems that are non-smooth, but which have remarkable structural properties (nonexpansiveness in certain metrics, monotonicity) or combinatorial properties. Algebraic methods allow one to obtain analytical expressions for performance measures (throughput, waiting time, etc). A recent application, to emergency call centers, can be found in [46].

4.2. Optimal control and games

Optimal control and game theory have numerous well established applications fields: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In most of these applications, one needs either to derive analytical or qualitative properties of solutions, or design exact or approximation algorithms adapted to large scale problems.

4.3. Operations Research

We develop, or have developed, several aspects of operations research, including the application of stochastic control to optimal pricing, optimal measurement in networks [112]. Applications of tropical methods arise in particular from discrete optimization [62], [63], scheduling problems with and-or constraints [103], or product mix auctions [120].

4.4. Computing program and dynamical systems invariants

A number of programs and systems verification questions, in which safety considerations are involved, reduce to computing invariant subsets of dynamical systems. This approach appears in various guises in computer science, for instance in static analysis of program by abstract interpretation, along the lines of P. and R. Cousot [69], but also in control (eg, computing safety regions by solving Isaacs PDEs). These invariant sets are often sought in some tractable effective class: ellipsoids, polyhedra, parametric classes of polyhedra with a controlled complexity (the so called “templates” introduced by Sankaranarayanan, Sipma and Manna [113]), shadows of sets represented by linear matrix inequalities, disjunctive constraints represented by tropical polyhedra [48], etc. The computation of invariants boils down to solving large scale fixed point problems. The latter are of the same nature as the ones encountered in the theory of zero-sum games, and so, the techniques developed in the previous research directions (especially methods of monotonicity, nonexpansiveness, discretization of PDEs, etc) apply to the present setting, see e.g. [76], [81] for the application of policy iteration type algorithms, or for the application for fixed point problems over the space of quadratic forms [7]. The problem of computation of invariants is indeed a key issue needing the methods of several fields: convex and nonconvex programming, semidefinite programming and symbolic computation (to handle semialgebraic invariants), nonlinear fixed point theory, approximation theory, tropical methods (to handle disjunctions), and formal proof (to certify numerical invariants or inequalities).

VALSE Project-Team

4. Application Domains

4.1. Application Domains

An objective of the team is the application of the developed control and estimation algorithms for different scenarios in IoT or CPSs. The participation in various potential applications allows Valse team to better understand the features of CPSs and their required performances, and to formulate properly the control and estimation problems that have to be solved. Here is a list of ongoing and potential applications addressed in the team:

- smart bivalve-based biosensor for water quality monitoring (ANR project **WaQMoS**, the developed sensor is shown in Fig. 2): in living beings, the presence of persistent external perturbations, may be difficult to measure, and important model uncertainties render the application of conventional techniques complicated; another issue for estimation is the consensus seeking between animals for a contamination detection [68];

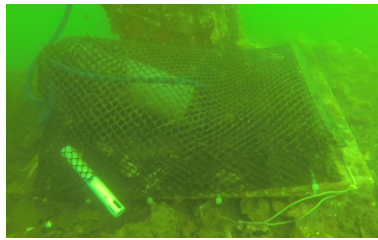


Figure 2. The valvometer used in ANR project **WaQMoS**

- control and estimation for flying vehicles, *e.g.* quadrotors or blimps given in Fig. 3 (PhD Centrale Lille): nonlinearity of the model and its uncertainty coupled with important aerodynamic perturbations have to be compensated by fast (finite- or fixed-time) and robust control and estimation algorithms;



Figure 3. Blimp and quadrotor robots

- human behavior modeling and identification with posterior design of algorithms for human-computer interaction (ANR project **TurboTouch**): robust finite-time differentiators demonstrate good estimation capabilities needed for prediction in this application [88], [69];
- human physiological characteristics estimation (like emotion detection, galvanic skin response filtering, fatigue evaluation in collaborations with **Neotrope** and **Ellicie Healthy**): intelligent robust filtering and finite-time distributed estimation are key features in these scenarios;
- path planning for autonomous vehicles taking into account the behavior of humans (PhD CIFRE with SEQUEL team and Renault): application of interval and finite-time adaptive estimation and prediction techniques allows for treating the uncertainty of the environment by reducing computational complexity of reinforcement learning [79]⁰;
- flow control (in the framework of ContrATech subprogram of **CPER ELSAT**, see also [78]): the case of control and estimation of a distributed-parameter system with very fast and uncertain dynamics, where finite-time solutions developed by Valse are necessary (an example of results is given in Fig. 4);

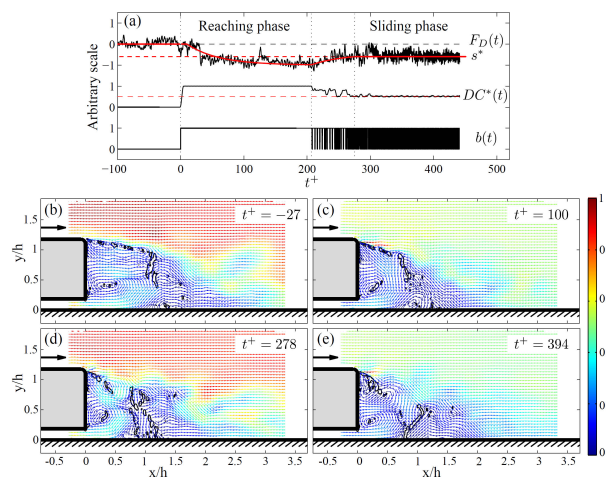


Figure 4. Particle Image Velocimetry on flow control for an Ahmed body (LAMIH wind tunnel)

- control of synthetic microbial communities (in the framework of **IPL COSY**, the experimental platform is shown in Fig. 5): here again, the problem is an important uncertainty of the model, which can be handled by robust sliding mode control algorithms, or by applying adaptive finite-time estimation and identification tools;

It is worth to highlight a widespread distribution of various scientific domains in the list of applications for the team given above. Such an *interdisciplinarity* for Valse is unsurprising, since the control theory is a science of systems whose interest today is, by nature, to interface with other disciplines and their fields of application. This is also well aligned with the domain of CPSs, which by its origin requires multidisciplinary competences.

⁰The examples of interval prediction algorithm application can be consulted [here](#).



Figure 5. Chemostat platform at Inria, Grenoble

ABS Project-Team (section vide)

AIRSEA Project-Team

4. Application Domains

4.1. The Ocean-Atmosphere System

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) or the prediction of **floods**.

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 AIRSEA 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.

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, bio/geo-chemistry of oceans, resilience of society w.r.t. hazardous flows, urban pollutions, ...

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 state, which is the balance between the gravity field and the Coriolis force, can significantly improve the forecasts. In addition, 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 expensive and generally not efficient in the long term. The objective is to determine a configuration almost stable for the facilities. In addition, it is also important to determine the impact of major events like emptying dam which 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 damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worst damages. For example, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last require 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 this requires fluid/structure interactions be well understood. Moreover, underground flows, in particular in sewers, 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 no longer the case close to the epicenter and in the coastal zone where the bathymetry leads to vertical accretions and produces substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

While 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 kinetic and potential parts. The buoys use the potential energy whereas the water-turbines 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.

4.6. Urban environment

The urban environment is essentially studied for air and noise pollutions. Air pollution levels and noise pollution levels vary a lot from one street to next. The simulations are therefore carried out at street resolution and take into account the city geometry. The associated numerical models are subject to large uncertainties. Their input parameters, e.g. pollution emissions from road traffic, are also uncertain. Quantifying

the simulation uncertainties is challenging because of the high computational costs of the numerical models. An appealing approach in this context is the use of metamodels, from which ensembles of simulations can be generated for uncertainty quantification.

The simulation uncertainties can be reduced by the assimilation of fixed and mobile sensors. High-quality fixed monitoring sensors are deployed in cities, and an increasing number of mobile sensors are added to the observational networks. Even smartphones can be used as noise sensors and dramatically increase the spatial coverage of the observations. The processing and assimilation of the observations raises many questions regarding the quality of the measurements and the design of the network of sensors.

4.7. SmartCity

There is a growing interest for environmental problems at city scale, where a large part of the population is concentrated and where major pollutions can occur. Numerical simulation is well established to study the urban environment, *e.g.* for road traffic modelling. As part of the smartcity movement, an increasing number of sensors collect measurements, at traditional fixed observation stations, but also on mobile devices, like smartphones. They must properly be taken into account given their number but also their potential low quality.

Practical applications include air pollution and noise pollution. These directly relate to road traffic. Data assimilation and uncertainty propagation are key topics in these applications.

ARAMIS Project-Team

4. Application Domains

4.1. Introduction

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

- better understanding the pathophysiology of brain disorders;
- designing systems to support clinical decisions such as diagnosis, prognosis and design of clinical trials;
- developing brain computer interfaces for clinical applications.

4.2. Understanding brain disorders

Computational and statistical approaches have the potential to help understand the pathophysiology of brain disorders. We first aim to contribute to better understand the relationships between pathological processes, anatomical and functional alterations, and symptoms. Moreover, within a single disease, there is an important variability between patients. The models that we develop have the potential to identify more homogeneous disease subtypes, that would constitute more adequate targets for new treatments. Finally, we aim to establish the chronology of the different types of alterations. We focus these activities on neurodegenerative diseases: dementia (Alzheimer's disease, fronto-temporal dementia), Parkinson's disease, multiple sclerosis.

4.3. Supporting clinical decisions

We aim to design computational tools to support clinical decisions, including diagnosis, prognosis and the design of clinical trials. The differential diagnosis of neurodegenerative diseases can be difficult. Our tools have the potential to help clinicians by providing automated classification that can integrate multiple types of data (clinical/cognitive tests, imaging, biomarkers). Predicting the evolution of disease in individual patients is even more difficult. We aim to develop approaches that can predict which alterations and symptoms will occur and when. Finally, new approaches are needed to select participants in clinical trials. Indeed, it is widely recognized that, to have a chance to be successful, treatments should be administered at a very early stage.

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 basis 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.

ATHENA Project-Team

4. Application Domains

4.1. Applications of diffusion MRI

Clinical domain: Diagnosis of neurological disorder

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 structural imaging modality that will be considered to recover the CNS connectivity.

4.2. Applications of M/EEG

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.
- Collaboration with the *Institut de Neurosciences des Systèmes* on these topics <http://ins.univ-amu.fr/>

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 develop a research collaboration with the eemagine/ANT-Neuro company. We collaborate with Nice University Hospital on the usage of BCI-based communication for ALS⁰ patients.

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

BEAGLE Project-Team

4. Application Domains

4.1. Functional and Evolutionary Biology

We do not distinguish our research and its application domains. Our shared idea is that the research is oriented by a scientific question, which is a multidisciplinary one, most often of biological nature. We do not develop methodologies independently from this question and then look for applications. Instead we collectively work with other disciplines to solve a question, with our competencies.

In consequence the application domains are already listed in the description of our projects and goals. They concern functional and evolutionary biology, related to critical social questions as human and plant health.

4.2. Social and Environmental Responsibility (Implication domains)

These last years we have maintained a frequent team discussion on the social and environmental responsibility of researchers. It has become more frequent this year, with the announcements of serious environmental issues by many governmental or non governmental organizations.

We are engaged in many actions regarding this responsibility. A constant ethics questioning, directing our research projects according to our values, teaching and popularization of ethical values. In particular we are engaged in several research projects on health and environment, and one of us has been a member of the institutional workgroup on environmental issues at Inria.

Regarding the functioning of research activities, we attempted a measure of the environmental footprint of our activities, regardless of their aims. It has the shape of a carbon footprint analysis, gathering the carbon footprint of travels, computer usage, computer equipment. We are aware of the incompleteness of this analysis, as well by not including many activities (nutrition, homeplace-workplace trips), and not taking other environmental issues than carbon emissions.

However it is a starting point, that we presented to many colleagues who were interested in reproducing the computation, so we give the headlines here. We used the unitary costs given by a website we constructed: <https://ferme.yeswiki.net/Empreinte>. The data was collected on an average taken over 3 years, 2016-2018 (we cannot yet at this stage make the analysis for 2019). Travels by members of the team or invited researchers emitted 39.86tCO₂. Computing hours on local clusters emitted 17.46tCO₂. The acquisition of computers accounts for the emission of 4.53tCO₂. The total is 61.85tCO₂.

Based on this lower bound of our CO₂ emissions, the total per person is around 3 tons per person. Probably this number is highly underevaluated since it accounts for only part of our professional activities.

Whether it is too high or acceptable, and the establishment of a carbon budget is a difficult question. If we refer to the goals of the COP21 conclusions, we should emit less than 2 tons of CO₂ per person in 2050 to reach carbon neutrality. This includes professional and personal life, and all the services we benefit from. We did not arrive yet at a consensus on the objective we should reach to consider we have a sustainable activity, but we in majority recognize that we are anyway far above what would be a consensus objective. We are still in a discussion to engage in a reduction.

BIGS Project-Team

4. Application Domains

4.1. Tumor growth-oncology

On this topic, we want to propose branching processes to model appearance of mutations in tumor through new collaborations with clinicians. The observed process is the "circulating DNA" (ctDNA). The final purpose is to use ctDNA as a early biomarker of the resistance to an immunotherapy treatment. It is the aim of the ITMO project. Another topic is the identification of dynamic network of expression. In the ongoing work on low-grade gliomas, a local database of 400 patients will be soon available to construct models. We plan to extend it through national and international collaborations (Montpellier CHU, Montreal CRHUM). Our aim is to build a decision-aid tool for personalised medicine. In the same context, there is a topic of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

4.2. Genomic data and micro-organisms population

Despite of his 'G' in the name of BIGS, Genetics is not central in the applications of the team. However, we want to contribute to a better understanding of the correlations between genes trough their expression data and of the genetic bases of drug response and disease. We have contributed to methods detecting proteomics and transcriptomics variables linked with the outcome of a treatment.

4.3. Epidemiology and e-health

We have many works to do in our ongoing projects in the context of personalized medicine with CHU Nancy. They deal with biomarkers research, prognostic value of quantitative variables and events, scoring, and adverse events. We also want to develop our expertise in rupture detection in a project with APHP (Assistance Publique Hôpitaux de Paris) for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram, body temperature, spirometry parameters, sleep duration, etc. Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

4.4. Dynamics of telomeres

Telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Trough a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently obtained data on the distribution of the length of telomeres from blood cells. With members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

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) [3].

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 research concentrates on the protection of cultures of photosynthetic organisms against their pests or their competitors. The cultures 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 harmful pathogens and invading species. We opt for protecting the culture through the use of biocontrol in a broad sense.

In crop production, biocontrol is indeed a very promising alternative to reduce pesticide use: it helps protecting the environment, as well as the health of consumers and producers; it limits the development of resistance (in comparison to chemicals). The use of biocontrol agents, which are, generically, natural enemies (predators, parasitoids or pathogens) of crop pests [89], 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. Resistant crops are also used instead of pesticides to control pests and pathogens, but the latter eventually more or less rapidly overcome the resistance, so these crops need to be replaced by new resistant crops. As resistant genes are a potentially limited resource, a challenge is to ensure the durability of crop resistance. 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 and of crop resistance.

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 [92] must be extended from a theoretical point of view and validated experimentally.

BIOVISION Project-Team

4. Application Domains

4.1. Applications of virtual/augmented reality for low-vision

- **Rehabilitation:** Serious games are games designed for a primary purpose which is not pure entertainment. In our context, we think about serious games as a way to help low-vision patients in performing rehabilitation exercises. Virtual/augmented reality technology is a promising platform to develop such rehabilitation exercises targeted to specific pathologies. For example, with Age Macular Degeneration (AMD), our objective is to propose solutions allowing rehabilitation of visuo-perceptual-motor functions to optimally use residual portions of the peripheral retina and obtain efficient eccentric viewing.
- **Vision aid-systems:** A variety of aids for low-vision people are already on the market using different kinds of virtual/augmented reality platforms (dedicated or large public ones). They offer different functionalities (magnification, image enhancement, text to speech, face and object recognition). Our goal is to design new solutions allowing autonomous interaction in mixed reality environments, and take advantage of the improvement of functions obtained via rehabilitation protocols.
- **Cognitive research:** Virtual/augmented reality technology represents a new opportunity to conduct cognitive and behavioural research using virtual environments where all parameters can be psychophysically controlled. Our objective is to re-assess common theories by allowing patients to freely explore their environment in more ecological conditions.

4.2. Applications of vision modeling studies

- **Neuroscience research.** Making in-silico experiments is a way to reduce the experimental costs, to test hypotheses and design models, and to test algorithms. Our goal is to develop a large-scale simulations platform of impaired retinas, called Macular, allowing to mimic specific degeneracies or pharmacologically induced impairments, as well as to emulate electric stimulation by prostheses. In addition, the platform provides a realistic entry to models or simulators of the thalamus or the visual cortex, in contrast to the entries usually considered in modelling studies.
- **Education.** Macular is also targeted as a useful tool for educational purposes, illustrating for students how the retina works and responds to visual stimuli.

CAMIN Project-Team (section vide)

CAPSID Project-Team

4. Application Domains

4.1. Biomedical Knowledge Discovery

Participants: Marie-Dominique Devignes [contact person], Malika Smaïl-Tabbone [contact person], Sabeur Aridhi, David Ritchie, Gabin Personeni, Seyed Ziaeddin Alborzi, Kevin Dalleau, Bishnu Sarker, Emmanuel Bresso, Claire Lacomblez, Floriane Odje, Athénaïs Vaginay.

Our main application for Axis 1 : "New Approaches for Knowledge Discovery in Structural Databases", concerns biomedical knowledge discovery. We intend to develop KDD approaches on preclinical (experimental) or clinical datasets integrated with knowledge graphs with a focus on discovering which PPIs or molecular machines play an essential role in the onset of a disease and/or for personalized medicine.

As a first step we have been involved since 2015 in the ANR RHU "FIGHT-HF" (Fight Heart Failure) project, which is coordinated by the CIC-P (Centre d'Investigation Clinique Plurithématique) at the CHRU Nancy and INSERM U1116. In this project, the molecular mechanisms that underly heart failure (HF) are re-visited at the cellular and tissue levels in order to adapt treatments to patients' needs in a more personalized way. The Capsid team is in charge of a workpackage dedicated to network science. A platform has been constructed with the help of a company called Edgeleap (Utrecht, NL) in which biological molecular data and ontologies, available from public sources, are represented in a single integrated complex network also known as knowledge graph. We are developing querying and analysis facilities to help biologists and clinicians interpreting their cohort results in the light of existing interactions and knowledge. We are also currently analyzing pre-clinical data produced at the INSERM unit on the comparison of aging process in obese versus lean rats. Using our expertise in receptor-ligand docking, we are investigating possible cross-talks between mineralocorticoid and other nuclear receptors.

Another application is carried out in the context of a UL-funded interdisciplinary project in collaboration with the CRAN laboratory. It concerns the study of the role of estrogen receptors in the development of glioblastoma tumors. The available data is high-dimensional but involves rather small numbers of samples. The challenge is to identify relevant sets of genes which are differentially expressed in various phenotyped groups (w.r.t. gender, age, tumor grade). The objectives are to infer pathways involving these genes and to propose candidate models of tumor development which will be experimentally tested thanks to an ex-vivo experimental system available at the CRAN.

Finally, simulating biological networks will be important to understand biological systems and test new hypotheses. One major challenge is the identification of perturbations responsible for the transformation of a healthy system to a pathological one and the discovery of therapeutic targets to reverse this transformation. Control theory, which consists in finding interventions on a system in order to prevent it to go in undesirable states or to force it to converge towards a desired state, is of great interest for this challenge. It can be formulated as "How to force a broken system (pathological) to act as it should do (normal state)?" Many formalisms are used to model biological processes, such as Differential Equations (DE), Boolean Networks (BN), cellular automata. In her PhD thesis, Athénaïs Vaginay investigates ways to find a BN fitting both the knowledge about topology and state transitions "inferred" from experimental data. This step is known as "boolean function synthesis". Our aim is to design automated methods for building biological networks and define operators to intervene on them[29]. Our approaches will be driven by knowledge and keep close connection with experimental data.

4.2. Prokaryotic Type IV Secretion Systems

Participants: Marie-Dominique Devignes [contact person], Isaure Chauvot de Beauchêne [contact person], Bernard Maigret, David Ritchie, Philippe Noel, Antoine Moniot, Dominique Mias-Lucquin.

Concerning Axis 2 : "Integrative Multi-Component Assembly and Modeling", our first application domain is related to prokaryotic type IV secretion systems.

Prokaryotic type IV secretion systems constitute a fascinating example of a family of nanomachines capable of translocating DNA and protein molecules through the cell membrane from one cell to another [36]. The complete system involves at least 12 proteins. The structure of the core channel involving three of these proteins has recently been determined by cryo-EM experiments for Gram-negative bacteria [47], [63]. However, the detailed nature of the interactions between the other components and the core channel remains to be found. Therefore, these secretion systems represent a family of complex biological systems that call for integrated modeling approaches to fully understand their machinery.

In the framework of the Lorraine Université d'Excellence (LUE-FEDER) "CITRAM" project we are pursuing our collaboration with Nathalie Leblond of the Genome Dynamics and Microbial Adaptation (DynAMic) laboratory (UMR 1128, Université de Lorraine, INRA) on the mechanism of horizontal transfer by integrative conjugative elements (ICEs) and integrative mobilisable elements (IMEs) in prokaryotic genomes. These elements use Type IV secretion systems for transferring DNA horizontally from one cell to another. We have discovered more than 200 new ICEs/IMEs by systematic exploration of 72 *Streptococcus* genomes and characterized a new class of relaxases [21]. We have modeled the dimer of this relaxase protein by homology with a known structure. For this, we have created a new pipeline to model symmetrical dimers of multi-domains proteins. As one activity of the relaxase is to cut the DNA for its transfer, we are also currently studying the DNA-protein interactions that are involved in this very first step of horizontal transfer (see next section).

4.3. Protein - Nucleic Acids Interactions

Participants: Isaure Chauvot de Beauchêne [contact person], David Ritchie, Dominique Mias-Lucquin, Antoine Moniot, Honey Jain, Anna Kravchenko, Hrishikesh Dhondge, Malika Smaïl-Tabbone, Marie-Dominique Devignes.

The second application domain of Axis 2 concerns protein-nucleic acids interactions. We need to assess and optimize our new algorithms on concrete protein-nucleic acids complexes in close collaboration with external partners coming from the experimental field of structural biology. To facilitate such collaborations, we will have to create automated and re-usable protein-nucleic acid docking pipelines.

This is the case for our PEPS collaboration "InterANRIL" with the IMoPA lab (CNRS-Université de Lorraine). We are currently working with biologists to apply our fragment-based docking approach to model complexes of the long non-coding RNA (lncRNA) ANRIL with proteins and DNA. In order to extend this approach to partially structured RNA molecules, we have built an automated pipeline to create (i) libraries of RNA fragments with arbitrary characteristics such as secondary structure, and (ii) testing benchmarks for applying these libraries to docking assays.

In the framework of our LUE-FEDER CITRAM project (see above), we adapted this approach and this pipeline to single-strand DNA docking in order to model the complex formed by a bacterial relaxase and its target DNA.

In the future, we will tackle a defined group of RNA-binding proteins containing RNA-Recognition Motif (RRM) domains. We will study existing and predicted complexes between various types of RRM and various RNA sequences with computational methods in order to calculate CG force-field energy and to help design new synthetic proteins with targeted RNA specificity. This is the goal of the ITN RNAct project and it will require the construction of a dedicated database equipped with querying and analysis facilities, including machine-learning approaches, as well as many interactions within the ITN RNAct consortium.

CARMEN Project-Team

4. Application Domains

4.1. Scientific context: the LIRYC

The University Hospital of Bordeaux (*CHU de Bordeaux*) is equipped with a specialized cardiology hospital, the *Hôpital Cardiologique du Haut-Lévêque*, where the group of Professor Michel Haïssaguerre has established itself as a global leader in the field of cardiac electrophysiology [63], [61], [48]. Their discoveries in the area of atrial fibrillation and sudden cardiac death syndromes are widely acclaimed, and the group is a national and international referral center for treatment of cardiac arrhythmia. Thus the group also sees large numbers of patients with rare cardiac diseases.

In 2011 the group has won the competition for a 40 million euro *Investissements d'Avenir* grant for the establishment of IHU Liryc, an institute that combines clinical, experimental, and numerical research in the area of cardiac arrhythmia (<http://ihu-liryc.fr>). The institute works in all areas of modern cardiac electrophysiology: atrial arrhythmias, sudden death due to ventricular fibrillation, heart failure related to ventricular dyssynchrony, and metabolic disorders. It is recognized as one of the most important centers worldwide in this area.

The Carmen team was founded to partner with IHU Liryc. We bring applied mathematics and scientific computing closer to experimental and clinical cardiac electrophysiology. In collaboration with experimental and clinical researchers at Liry we work to enhance fundamental knowledge of the normal and abnormal cardiac electrical activity and of the patterns of the electrocardiogram, and we 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 IHU Liryc. It help to write new concepts concerning the multiscale organisation of the cardiac action potentials that will serve our understanding in many electrical pathologies. For example, we model the structural heterogeneities at the cellular scale [50], and at an intermediate scale between the cellular and tissue scales.

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 veins. We will model the heterogeneities specific to the atria, like fibrosis or fatty infiltration [69] [60]. These heterogeneities ara thought to play a major role in the development of atrial fibrillation.

At the ventricular level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles, which is supposed to play a major role in sudden cardiac death, and (2) modeling the heteogeneities related to the complex organization and disorganization of the myocytes and fibroblasts, which is important in the study of infarct scars for instance.

4.3. Clinical electrophysiology

Treatment of cardiac arrhythmia is possible by pharmacological means, by implantation of pacemakers and defibrillators, and by curative ablation of diseased tissue by local heating or freezing. In particular the ablative therapies create challenges that can be addressed by numerical means. Cardiologists would like to know, preferably by noninvasive means, where an arrhythmia originates and by what mechanism it is sustained.

We address this issue in the first place using inverse models, which attempt to estimate the cardiac activity from a (high-density) electrocardiogram. A new project aims at performing this estimation on-site in the catheterization laboratory and presenting the results, together with the cardiac anatomy, on the screen that the cardiologist uses to monitor the catheter positions [66].

An important prerequisite for this kind of interventions and for inverse modeling is the creation of anatomical models from imaging data. The Carmen team contributes to better and more efficient segmentation and meshing through the IDAM project.

CASTOR Project-Team (section vide)

COFFEE Project-Team

4. Application Domains

4.1. Multiphase porous media flows and multi-physics coupling

Our research focuses on the numerical modeling of multiphase porous media flows accounting for complex geology and for nonlinear and multi-physics couplings. It is applied to various problems in the field of energy such as the simulation of geothermal systems in collaboration with BRGM, of nuclear waste repositories in collaboration with Andra, and of oil and gas recovery in collaboration with Total. Our research directions include the development of advanced numerical schemes adapted to polyhedral meshes and highly heterogeneous media in order to represent more accurately complex geologies. A special focus is made on the modeling of multiphase flows in network of faults or fractures represented as interfaces of co-dimension one coupled to the surrounding matrix. We also investigate nonlinear solvers adapted to the nonlinear couplings between gravity, capillary and viscous forces in highly heterogeneous porous media. In the same line, we study new domain decomposition algorithms to couple non-isothermal compositional liquid gas flows in a porous medium with free gas flows occurring at the interface between the ventilation gallery and the nuclear waste repository or between a geothermal reservoir and the atmosphere. We have begun exploring the coupling between the multiphase flow in the porous matrix and the solid mechanics involved in opening fractures.

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. Fungal network growth

Members of the team have started an original research program devoted to fungal network growth. We started working on this subject through a collaboration with biologists and physicists at LIED (Université Paris Diderot) and probabilists in CMAP (Ecole Polytechnique) and Université Paris Sud, involving Rémi Catellier and Yves D'Angelo in LJAD in Nice. The motivation is to understand branching networks as an efficient space exploration strategy, with fungus *Podospora Anserina* being the biological model considered. This research is submitted as an ANR-project and has been supported by various local fundings.

COMMEDIA Project-Team

4. Application Domains

4.1. Cardiovascular hemodynamics

The heart is a double pump whose purpose is to deliver blood to the tissue and organs of the body. This function is made possible through the opening and closing of the heart valves. Cardiac diseases generally manifest by affecting the pumping function of the heart. Numerical simulations of cardiac hemodynamics, in normal and pathological conditions, are recognized as a tool of paramount importance for improving the understanding, diagnosis and treatment of cardiac pathologies, and also for the development of implantable devices (see, e.g., [63], [46]). As an example, we can mention the case of cardiac mitral valve regurgitation, one of the most common heart valve diseases. For this pathology, clinical data are known to be insufficient for determining the optimal timing for surgery, the best surgical strategy and the long-term outcome of a surgical repair. Contrary to imaging techniques, numerical simulations provide local information, such as pressure and stresses, which are of fundamental importance for the prediction of the mechanical behavior of native valves and of implantable devices.

4.2. Respiratory flows

Respiration involves the transport of air through the airways from the mouth to the alveoli of the lungs. These units where diffusion of oxygen and carbon dioxide take place, are surrounded by a viscoelastic medium (the parenchyma) consisting of blood vessels and collagen fibers. Air flows due to the displacement of the diaphragm, which drives the pulmonary parenchyma. Accidental inhalations of foreign bodies or pathologies such as asthma, emphysema and fibrosis might prevent the lung of fulfilling its function. Therapies mostly use aerosols (set of small particles, solid or liquid), which must reach the specific areas of the lung targeted for treatment. Understanding the airflow mechanisms within the respiratory network is a fundamental ingredient for predicting the particles motion and their deposition (see, e.g., [44]). Moreover, understanding of the gas diffusion in the lung is also of major importance since the main function of this organ is to deliver oxygen to the blood.

4.3. Safety pharmacology

The problem of safety pharmacology can be summarized as follows: given a molecule which is a candidate to become a drug, is its use dangerous due to side effects? Among all the different problems to be addressed, one of the most relevant questions in pharmacology is cardio-toxicity (see [58]). More precisely, the objective is to determine whether or not a molecule alters in a significant way the normal functioning of the cardiac cells. To answer these questions, the CiPA initiative promotes the introduction of novel techniques and their standardisation (see [50]). One of the proposed tests of the CiPA panel is to measure the the electrical activity using Micro-Electrodes Array: these are microchips that record the electrical activity of an ensemble of cells. The task is to infer the impact of a drug on the ionic currents of each cell based on the electrical signal measured (electrograms) and, in perspective, to be able to assess whether a molecule can induce arrhythmia (see [49]).

DRACULA Project-Team (section vide)

DYLISS Project-Team

4. Application Domains

4.1. Application fields in biology

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with laboratories of biology in order to consolidate their studies by a smart use of our tools and prototypes and to generate new biological hypotheses to be tested experimentally.

Marine Biology: seaweed enzymes and metabolism & sea-urchin cell-cycle. Our main field of study is **marine biology**, as it is a transversal field covering challenges in integrative biology, dynamical systems and sequence analysis. Our methods based on combinatorial optimization for the reconstruction of genome-scale metabolic networks and on classification of enzyme families based on local and partial alignments allowed the seaweed metabolism *E. Siliculosus* to be deciphered [62], [53]. The study of the *HAD* superfamily of proteins thanks to partial local alignments, produced by *Protomata* tools, allows sub-families to be deciphered and classified, and the metabolic map reconstructed with *Meneco* enabled the reannotation of 56 genes within the *E. siliculosus* genome. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpus* [55] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. Similarly, our analyses suggest that the bacterium *Ca. P. ectocarpus* is able to provide both β -alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway [63].

Micro-biology: elucidating the functioning of extremophile consortiums of bacteria. In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (co-funded by the Center of Mathematical Modeling, the Center of Regulation Genomics and Inria-Chile). In order to elucidate the main characteristics of these bacteria, our integrative methods were developed to identify the main groups of regulators for their specific response in their living environment. The integrative biology tools *Meneco*, *Lombarde* and *Shogen* have been designed in this context. In particular, genome-scale metabolic network been recently reconstructed and studied with the *Meneco* and *Shogen* approaches, especially on bacteria involved in biomining processes [48] and in Salmon pathogenicity [52].

Agriculture and environmental sciences: upstream controllers of cow, pork and pea-aphid metabolism and regulation. In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [47] and on semantic-based analysis of metabolic networks [45] was very valuable for interpreting differences of gene expression in pork meat [60] and figure out the main gene-regulators of the response of porks to several diets [59]. In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants. In terms of biological output of the network studies on the pea aphid microRNAs, we have identified one new microRNA (apmir-3019, not present in any known species other than the pea aphid) who has more than 900 putative mRNA targets.

Health: deciphering pathways involved in the TGF- β signalling network. TGF- β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both beneficial anti-tumor effect [49] and pro-tumor effect [61]. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. For that purpose we developed *Cadbiom* and identified gene networks associated with innate immune response to viral infection that combine TGF- β and interleukine signaling pathways [43], [51].

EMPENN Project-Team

4. Application Domains

4.1. Population imaging

One major objective of neuroimaging researchers and clinicians is to be able to stratify brain imaging data in order to derive new and more specific population models. In practice this requires to set up large-scale experiments that, due to the lack of resources and capabilities to recruit locally subjects who meet specific inclusion criteria, motivates the need for sharing the load.

But, building and using multi-site large-scale resources poses specific challenges to deal with the huge quantity of data produced and their diversity. Empenn will focus on two challenges in particular:

- Provide computational environments for the computation and use of imaging biomarkers in the targeted brain diseases, a solution to be used by radiologists and neurologists/psychiatrists for the clinical follow-up of a large patient population.
- Modeling analytic variability of image processing pipelines to better understand and predict the behaviour of imaging biomarker detection solutions and improve reproducibility and productivity in clinical neuroimaging research.

4.2. Detection and learning

We intend to make significant contributions with major impacts in learning coupling models between functional recordings during neurofeedback procedures. These advances will provide a breakthrough in brain-computer interfaces for rehabilitation protocols. Our aim is to:

- Provide a computational environment that combines data-driven (machine learning) and Bayesian solutions to improve the detection of abnormal patterns in images through decision or evidence theory data fusion strategies. The major initial application will be for multiple sclerosis. Over the longer term, we also expect to adapt these methods to address a wider range of neurological diseases (epilepsy, stroke, tumors, etc.) in neonate and adult brains.
- Develop solutions for combining brain state measurements from multimodal sensors or sequences (e.g. fMRI, ASL, EEG, NIRS, etc.) with applications in the spatiotemporal reconstruction of brain activity from MRI-EEG or the combined detection of the endogenous hemodynamic and resting state network of the brain from ASL and NIRS. Over the longer term, the advent of new hybrid brain imaging sensors (e.g. PET-MRI) will require these methods to be extended to a larger spectrum of information combining structural, morphological, metabolic, electrophysiological and cellular/molecular information (e.g. through the use of specific ligands/nanocarriers).

4.3. Quantitative imaging

The Empenn research group focuses on the development of several quantitative techniques in magnetic resonance imaging of the brain. These methods allow for a characterization of both the function and the structure of the brain with high precision. Arterial spin labelling (ASL) is a contrast agent-free imaging technique which labels arterial blood water as an endogenous tracer for perfusion and can measure resting-state cerebral blood flow. We are interested in estimating multiparametric hemodynamics using ASL, such as combined cerebral blood flow and arterial transit times, and derive statistical descriptors to represent significant differences between groups. In addition to quantitative perfusion parameters, our contributions on tissue compartment imaging aim at delineating neural circuits and characterize their microstructure properties, using both diffusion MRI and relaxometry. In diffusion MRI, arbitrary gradient waveforms were shown to exhibit higher sensitivity to microstructure parameters than standard pulsed gradients. We work on the optimization of sampling protocols in this domain, with the objective to propose sequences compatible with in vivo acquisition.

Complementary to diffusion MRI, we develop methods for the reconstruction of myelin-bound, extra-axonal and cerebrospinal fluid water using multi-compartment modelling of the T2-relaxometry signal. We combine these techniques with tractography to identify trajectories of pathologies associated to the evolution of these microstructural parameters along specific fiber bundles in the brain white matter.

4.4. Behavior

Advances in the field of in vivo imaging offer new opportunities for addressing the management of resistant affective disorders and their consequences (suicide risk and socio-professional impact), and the management of spatial cognition disorders after stroke and their consequences (postural perturbations and the loss of autonomy). Our objective, and the main challenge in this context, will be to introduce medical image computing methods to the multidisciplinary field of behavioral disorders (cognitive disorders, particularly spatial and postural control disorders or anterograde memory impairment, mood disorders, notably resistant depression, schizophrenic disorders, pervasive developmental disorders, attention disorders, etc.) in order to gain a better understanding of the pathology and devise innovative therapeutic approaches.

We also expect to become a major player in the future and make important contributions with significant impacts, primarily in drug-resistant depression in young and old populations. In particular, we expect to provide new image-related metrics combining perfusion, metabolism and microstructural information regarding the brain in order to better characterize pathologies, provide prospective evolution values and potentially provide new brain stimulation targets that could be used in neurofeedback rehabilitation protocols or other types of brain stimulation procedure.

We aim to provide new imaging markers of mental diseases, especially in the context of mood disorders. The new biomarkers will be derived from the metabolic (ASL and later ASL+PET) point of view as well as from the microstructural point of view (multicompartment diffusion MRI and relaxometry). Similarly, we expect to exhibit imaging biomarker regularities combining metabolic and structural information. Over the longer term, we expect these biomarkers to be the target of neurofeedback rehabilitation procedures. Also, over the longer term, we expect to supplement the MRI markers with molecular marker ones coming from new PET tracers, especially those associated with serotonin intake, at one time point or during a rehabilitation protocol under hybrid PET-EEG-MRI neurofeedback procedures.

4.5. Neuroinflammation

Some of the major ongoing research issues in the neuroimaging of neuro-inflammatory diseases concern the definition of new biomarkers for tracking the development of the pathology using high-dimensional data (e.g. nD+t MRI). This includes the use of white matter-specific imaging, such as magnetization transfer MRI, relaxometry and diffusion-weighted imaging (DW-MRI). Our objective is (1) to develop information-processing tools to tag the spatiotemporal evolutions of MS patterns at the brain parenchyma and spinal cord levels from their different signatures (inflammatory cells visible with USPIO or Gd contrast agents on MRI, persistent black holes, eloquent regional atrophy and microstructure signatures); and (2) to test these new tools on new imaging cohorts. In this respect, we for instance conduct studies on brain and spinal cord imaging, continuing on from the PHRC multicentric EMISEP project (PI: G. Edan), as it is very likely that lesions in the spine will directly affect the ambulatory ability of the patient (and thereby the clinical scores). In order to extend this experiment to a larger MS population, based on our expertise from the OFSEP cohort, we also plan to improve the MS therapeutic decision process through the MUSIC project (Multiple Sclerosis Imaging Check out, a public/private project). Our goal is to develop and assess a standardized monitoring tool that provides a robust, long-term computerized MRI follow-up that will become the gold standard in clinical practice for therapeutic decisions in MS treatment. As part of this project, Empenn will share its expertise in data management systems (Shanoir and FLI-IAM) and automatic processing tools (through the medInria and Anima software repositories) to extract quantitative indices from the images.

4.6. Recovery

Mental and neurological disorders are the leading cause of years lived with a disability. Treatment-resistant depression affects approximately 2% of the European population. Meanwhile, in the case of brain disorders, almost 1.5 million Europeans (15 million people worldwide) suffer a stroke event each year. Current recovery methods for brain disorders and traumatic brain injuries remain limited, preventing many from achieving full recuperation. We propose addressing the issue of brain recovery by introducing new advances from recent breakthroughs in computational medical imaging, data processing and human-machine interfaces, and demonstrating how these new concepts can be used, in particular for the treatment of stroke and major depressive disorders.

We ambition to combine advanced instrumental devices (Hybrid EEG, NIRS and MRI platforms), with new hybrid brain computer interface paradigms and new computational models to provide neurofeedback-based therapeutic and neuro-rehabilitation paradigms in some of the major mental and neurological disorders of the developmental and the aging brain.

Neurofeedback involves using a brain-computer interface that provides an individual with real-time biofeedback about his or her brain activity in the form of sensory feedback. It enables individuals to learn to better control their brain activity, which can be measured in real time using various non-invasive sensors as described above. Although EEG is currently the only modality used by clinical practitioners in that context, it lacks specificity due to its low spatial resolution. Dynamic research into fMRI-neurofeedback has held promise for treating depression, chronic pain and stroke, since it offers the prospect of real-time imagery of the activity in deep brain structures with high spatial resolution. However, the low temporal resolution and high cost of fMRI-Neurofeedback has hampered the development of many applications. We believe that the future belongs to hybrid responses that combine multimodal sensors and intend to demonstrate this in the Empenn project.

EPIONE Project-Team (section vide)

ERABLE Project-Team

4. Application Domains

4.1. Biology and Health

The main areas of application of ERABLE are: (1) biology understood in its more general sense, with a special focus on symbiosis and on intracellular interactions, and (2) health with a special emphasis for now on infectious diseases, rare diseases, and cancer.

FLUMINANCE Project-Team (section vide)

GENSCALE Project-Team

4. Application Domains

4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

4.2. Health

Genetic and cancer disease diagnostic: Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Bioinformatics analysis can be based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. One can also scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with genetic or cancer diseases.

Neurodegenerative disorders: The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new genetic biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

4.3. Agronomy

Insect genomics: Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities [6].

Improving plant breeding: Such projects aim at identifying favorable alleles at loci contributing to phenotypic variation, characterizing polymorphism at the functional level and providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

4.4. Environment

Food quality control: One way to check food contaminated with bacteria is to extract DNA from a product and identify the different strains it contains. This can now be done quickly with low-cost sequencing technologies such as the MinION sequencer from Oxford Nanopore Technologies.

Ocean biodiversity: The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and its role, for example, in the CO₂ sequestration.

IBIS Project-Team (section vide)

LEMON Project-Team (section vide)

LIFEWARE Project-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. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Furthermore, one important application of our research is the development of a **modeling software** for computational systems biology.

4.2. Modeling software for systems biology and synthetic 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. Coupled models of the cell cycle and the 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 protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of high-level functions in non-living vesicles for medical applications, such as biosensors for medical diagnosis⁰. Our approach is based on purely 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 Sys2Diag 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, in contrast to living cells in synthetic biology, is particularly appealing for security considerations 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.

⁰Alexis Courbet, Patrick Amar, François Fages, Eric Renard, Franck Molina. Computer-aided biochemical programming of synthetic microreactors as diagnostic devices. *Molecular Systems Biology*, 14(4), 2018.

4.5. Functional characterization of the resistance of bacterial populations to antimicrobial treatments

Antibiotic resistance is becoming a problem of central importance at a global level. Two mechanisms are at the origin of non-susceptibility to antimicrobial treatments. The first one comes from adaptation of bacterial cells to antibacterial treatments, notably through the modification of efflux pumps or the expression of enzymes that degrade the antibiotics. Cells are individually resistant. The second one, typically found in resistances to β -lactams, a broad class of antibiotics, originates from the release in the environment of the antibiotic degrading enzymes by the dead cells. This leads to population effects by which cells become collectively resilient.

The functional characterization of these different effects is important for the best use of antibiotics (antibiotic stewardship). In collaboration with Lingchong You (Duke University) and with Philippe Glaser (Institut Pasteur), we develop experimental platforms, models, and optimal model calibration methods that gives precise estimations of individual resistance and collective resilience of bacterial populations to antibiotic treatments.

M3DISIM Project-Team (section vide)

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. Imaging complex media with ultrasonic waves

The acoustic behavior of heterogeneous or composite materials attracts considerable excitement. Indeed, their acoustic response may be extremely different from the single constituents responses. In particular, dispersions of resonators in a matrix are the object of large research efforts, both experimentally and theoretically. However it is still a challenge to dispose of numerical tools with sufficient abilities to deal with the simulation and imaging of such materials behavior. Indeed, not only acoustic simulations are very time-consuming, but they have to be performed on realistic enough solution domains, i.e. domains which capture well enough the structural features of the considered materials.

This collaboration with I2M, University of Bordeaux aims at addressing this type of challenges by developing numerical and experimental tools in order to understand the propagation of ultrasonic waves in complex media, image these media, and in the future, help design composite materials for industrial purposes.

4.3. Helioseismology

This collaboration with the Max Planck Institute for Solar System, Göttingen, Germany, 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. 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 better understand Sun spots.

MAMBA Project-Team

4. Application Domains

4.1. Introduction

The team has two main application-driven research axes. Applicative axis 1 focuses on cancer, an application on which almost all team members work, with various approaches. A main focus of the team is to study cancer as a Darwinian evolutionary phenomenon in phenotype-structured cell populations. Optimal control methods take into account the two main pitfalls of clinical cancer therapeutics, namely unwanted toxic side effects in healthy cell populations and drug resistance in cancer cell populations. Other studies concern telomere shortening, and multi-scale models. Applicative axis 2 is devoted to growth, evolution and regeneration in populations and tissues. It involves protein aggregation and fragmentation models for neurodegenerative diseases (prion, Alzheimer), organ modeling, mainly of the liver, its damages induced by toxic molecules, and its regeneration after toxic insult. Newcomers in this applicative field are epidemiological modeling of propagation of insect vector-borne diseases by reaction-diffusion equations and of their optimal control, bacterial growth and wound healing.

4.2. Applicative axis 1: Focus on cancer

Personnel

Luis Almeida, Jean Clairambault, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil.

Project-team positioning

The MAMBA team designs and analyses mathematical models of tumor growth and therapy, at the cell population level, using agent-based or partial differential equations, with special interest in methodologies for therapeutic optimization using combined anticancer drug treatments. Rather than, or not only, modeling the effect of drugs on molecular targets, we represent these effects by their functional consequences on the fate of healthy and cancer cell populations: proliferation (velocity of the cell division cycle, decreasing it, e.g., by antagonizing growth factor receptors), apoptosis, cell death or senescence. Our goal in doing this is to circumvent the two main issues of anticancer therapy in the clinic, namely unwanted toxic side effects in populations of healthy cells and emergence of drug-induced drug resistance in cancer cell populations. This point of view leads us to take into account phenomena of transient and reversible resistance, observed in many cancer cell populations, by designing and analyzing models of cell populations structured in continuous phenotypes, relevant for the description of the behavior of cell populations exposed to drugs: either degree of resistance to a given drug, or potential of resistance to drug-induced stress, proliferation potential, and plasticity. Such modeling options naturally lead us to take into account in a continuous way (i.e., by continuous-valued phenotype or relevant gene expression) the wide phenotypic heterogeneity of cancer cell populations. They also lead us to adopt the point of view of adaptive dynamics according to which characteristic traits of cell populations evolve with tumor environmental pressure (drugs, cytokines or metabolic conditions, mechanical stress and spatial conditions), in particular from drug sensitivity to resistance. This position is original on the international scene of teams dealing with drug resistance in cancer.

Scientific achievements

Modeling Acute Myeloid Leukemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations

In collaboration with Catherine Bonnet (Inria DISCO, Saclay) and François Delhommeau (St Antoine hospital in Paris), together with DISCO PhD students José Luis Avila Alonso and Walid Djema, this theme has led to common published proceedings of conferences: IFAC, ACC, CDC, MTNS [66], [67], [68], [77], [95], [65]. These works study the stability of the haematopoietic system and its possible restabilization by combinations of anticancer drugs with functional targets on cell populations: proliferation, apoptosis, differentiation.

Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control

We tackle the problem to represent and inhibit - using optimal control algorithms, in collaboration with Emmanuel Trélat, proposed Inria team CAGE - drug-induced drug resistance in cancer cell populations. This theme, presently at the core of our works on cancer modeling with a evolutionary perspective on tumor heterogeneity, is documented in a series of articles [90], [92], [124], [125], [127]. Taking into account the two main pitfalls of cancer therapy, unwanted side effects on healthy cells and evolution towards resistance in cancer cells, it has attracted to our team the interest of several teams of biologists, with whom we have undertaken common collaborative works, funded by laureate answers to national calls (see ITMO Cancer HTE call).

This theme is also at the origin of methodological developments (see Research axis 1). In collaboration with Shensi Shen from Institut Gustave Roussy and Francois Vallette from Université de Nantes, we aim to develop simple non-spatial models to understand the mechanisms of drug resistance acquisition -and loss- in melanoma and glioblastoma. The models are systematically compared with in vitro and in vivo data generated by our collaborators and treated via image processing techniques developed in the team.

Senescence modeling by telomere shortening

In many animals, aging tissues accumulate senescent cells, a process which is beneficial to protect from cancer in the young organism. In collaboration with Teresa Teixeira and Zhou Xu from IBCP, we proposed a mathematical model based on the molecular mechanisms of telomere replication and shortening and fitted it on individual lineages of senescent *Saccharomyces cerevisiae* cells, in order to decipher the causes of heterogeneity in replicative senescence [79].

Biomechanically mediated growth control of cancer cells

Model simulations indicate that the response of growing cell populations on mechanical stress follows a simple universal functional relationship and is predictable over different cell lines and growth conditions despite the response curves look largely different. We developed a hybrid model strategy in which cells were represented by coarse-grained individual units calibrated in a high resolution cell model and parameterized each model cell by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics. Our model simulation results suggest that the growth response of cell population upon externally applied mechanical stress is the same, as it can be quantitatively predicted using the same growth progression function [123].

Bio-mechanical models of tissue growth

The degenerate Cahn-Hilliard equation is a standard model to describe living tissues. It takes into account cell populations undergoing short-range attraction and long-range repulsion effects. In this framework, we consider the usual Cahn-Hilliard equation with a singular single-well potential and degenerate mobility. These degeneracy and singularity induce numerous difficulties, in particular for its numerical simulation. To overcome these issues, we propose in [hal-02274417] a relaxation system formed of two second order equations which can be solved with standard packages. This system is endowed with an energy and an entropy structure compatible with the limiting equation. Here, we study the theoretical properties of this system; global existence and convergence of the relaxed system to the degenerate Cahn-Hilliard equation. We also study the long-time asymptotics which interest relies on the numerous possible steady states with given mass.

Free boundary multiphase models of tumor growth

Multiphase mechanical models are now commonly used to describe living tissues including tumour growth. The specific model we study here consists of two equations of mixed parabolic and hyperbolic type which extend the standard compressible porous media equation, including cross-reaction terms. We study the incompressible limit, when the pressure becomes stiff, which generates a free boundary problem. We establish the complementarity relation and also a segregation result. Several major mathematical difficulties arise in the two species case which are addressed in [43]. Firstly, the system structure makes comparison principles fail. Secondly, segregation and internal layers limit the regularity available on some quantities to BV. Thirdly, the Aronson-Bénilan estimates cannot be established in our context. We are lead, as it is classical, to add correction terms. This procedure requires technical manipulations based on BV estimates only valid in one space dimension. Another novelty is to establish an L^1 version in place of the standard upper bound.

Philosophy of cancer

The quite natural idea that cancer is a disease of the control of coherent multicellularity, expressed when cohesion of tissues and coherence of (unknown, except maybe for the case of a centralised circadian clock) synchronising signals fail to ensure it, by a regression towards unicellularity, stopping in this “reverse evolution path” at a coarse, incoherent multicellularity state⁰ continues to be developed and popularised by Jean Clairambault in seminars and workshops, and published in review articles [13], [45]. This view, and the investigation of the immune system in the design of such coherence of all multicellular organisms⁰ is naturally inscribed in a *philosophy of cancer* perspective, and from a mathematical viewpoint, to multicellularity genes - and links between them and unicellularity genes - seen as a *hyperstructure*⁰ above structures consisting of the genes of unicellularity, i.e., those that make a single cell a coherent living system, such hyperstructure being failed in cancer; this view is presently under development with colleagues from universities of the Paris region, together with Nils Baas at NTNU, Trondheim, Norway). This perspective, that makes use of category theory as a structuring point of view to apprehend multicellularity and cancer, is also meant to endow us with an innovative methodology to apply *topological data analysis (TDA)* to investigate cancer genome data.

Modelling of TMZ induced drug resistance

Temozolomide (TMZ) is a standard chemotherapy treatment in patients with glioblastoma. Resistance to this drug is correlated to the presence of a specific enzyme, which activity in cancer cells creates a drug-induced cell death resistant phenotype. Understanding the transition of cancer cells to a resistant phenotype is still a topic of research where multiple hypothesis have been studied: From an adaptive process to an inherent resistance to treatment. Moreover it has been recently shown that TMZ treatment has an influence on the spatial structuration of cancer cell aggregates. In the frame of the HTE project MoGIIImaging and through the recent hiring of a post-doctoral candidate (Gissell Estrada Rodriguez), we are currently developing a mathematical framework to study and analyse the evolution of a population of glioblastoma cells that are exposed to TMZ. Based on the experimental data generated by our partner team led by F. Valette (Inserm Nantes), we propose a Keller-Segel type model for the formation of spheroid as a result of a chemoattractant produced by cancer cells, and study the influence of a chemotherapeutic agent on the structuration of the cancer cell aggregates. By confronting the model results to experimental data, different modelling choices are currently explored to identify which key mechanisms could be responsible for the apparition of drug resistance in glioblastoma.

Collaborations

- AML modelling: **Catherine Bonnet**, DISCO Inria team, Saclay, and **François Delhommeau**, INSERM St Antoine (also collaborator in the INSERM HTE laureate project EcoAML, see below).
- INSERM HTE laureate project MoGIIImaging, headed by E. Moyal (Toulouse): **François Vallette**, CRCNA and INSERM Nantes
- INSERM HTE laureate project EcoAML, headed by **François Delhommeau**, INSERM St Antoine: François Delhommeau, Thierry Jaffredo (IBPS), Delphine Salort (LCQB-IBPS)
- Adaptive dynamics to model drug resistance and optimal control to circumvent it:

⁰Metazoa 1.0, as theorised by PCW Davies and CH Lineweaver in their article “Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors”, Physical Biology 2011, that popularised the so-called atavistic hypothesis of cancer

⁰this latter point partly, however nicely, developed in Thomas Pradeu’s book “The limits of the self”, OUP 2012

⁰See on this point, e.g., Nils Baas: “On the philosophy of higher structures”, Int. J. General Systems 2019

Alexandre Escargueil, Michèle Sabbah (1 PhD thesis in common), St Antoine Hospital, Paris

Emmanuel Trélat (1 PhD thesis in common) at Inria team CAGE and Laboratoire Jacques-Louis Lions at Sorbonne Université.

Frédéric Thomas at CREEC, Montpellier.

Tommaso Lorenzi (Univ. of St Andrews).

- Telomere shortening: **Teresa Teixeira** and **Zhou Xu** (IBCP, Paris), **Philippe Robert** (Inria RAP).
- TRAIL treatment: **Gregory Batt**, Inria Saclay and Inst. Pasteur (France)
- Biomechanical control of cancer cells: **Pierre Nassoy**, Bioimaging and Optofluidics Group, LP2N – UMR 5298, IOGS, CNRS & University of Bordeaux

4.3. Applicative axis 2: Growth, evolution and regeneration in populations and tissues

Luis Almeida, Pierre-Alexandre Bliman, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil, Philippe Robert

Project-team positioning

The applications in this category span very different subjects from amyloid diseases, dengue fever, wound healing, liver regeneration and toxicity, up to bacterial growth and development of organisms. As the applications, the methods span a wide range. Those concerning identification of models and parameters with regard to data have partially been outlined in axis 3. Focus in this axis is on the model contribution to the biologically and/or medically relevant insights and aspects.

Liver-related modelling is partially performed within the Inria team MIMESIS (Strasbourg) with the focus on real-time, patient-specific biomechanical liver models to guide surgery and surgeons. Internationally, spatial temporal liver related models are developed in Fraunhofer MEVIS (Bremen), by T. Ricken (TU Dortmund), and P. Segers group (Leuven).

Different from these, Mamba has a strong focus on spatial-temporal modeling on the histological scale, integration of molecular processes in each individual cell, and single-cell (agent) based models [102]. Works by Schliess [149], [107] have been highlighted in editorials.

Mathematical modeling of protein aggregation is a relatively recent domain, only a few other groups have emerged yet; among them we can cite the Inria team Dracula, with whom we are in close contact, and e.g., the work by Jean-Michel Coron (Sorbonne Université) and Monique Chyba (Hawaii, USA) in control, and Suzanne Sindi (USA) for the modeling of the yeast prion. We have interactions with all these groups and organized a workshop in June 2017, gathering both the biophysics and applied mathematics communities.

Scientific achievements

Amyloid disease

Application to protein aggregation in amyloid diseases is a long-standing interest of Mamba, dating back to 2010 [85], and developed through the collaboration with Human Rezaei's team at Inria. More recently, with Wei-Feng Xue in Canterbury, we investigated the intrinsic variability among identical experiments of nucleation [98], [106], Sarah Eugène's Ph.D subject (co-supervised by Philippe Robert) [105].

In collaboration with Tom Banks first [70], [69] and then Philippe Moireau, we developed quantitative comparisons between model and data. Through data assimilation and statistical methods [63], we proposed new models and mechanisms.

Biological control of arboviruses

Sterile Insect Technique (SIT) [104] is a biological control method relying on massive releases of sterile male insects into the wild. The latter compete with wild males to mate with the females, and induce no offspring to the latter, thus reducing the next generation's population. This can result in a progressive reduction, or even disparition, of the target population.

A related technique is based on the infection by *Wolbachia* [111]. This symbiotic bacterium is maternally transmitted from infected females to their offspring, but induces *cytoplasmic incompatibility* [151], [80]: mating between infected males and uninfected females gives no offspring. Releases of *Wolbachia* infected males alone is thus comparable to classical SIT.

On the other hand, releasing both infected males and females in sufficient quantity may result in infection of the wild population. This gives rise to an interesting new control principle, as *Wolbachia* has been shown to severely reduce the insect vectorial ability to transmit dengue, zika or chikungunya, indirectly by lifespan and fertility reduction, and directly by reducing the ability of the viruses to proliferate within the organism [130].

We proposed new insights on the practical and theoretical issues raised by the implementation of the previous methods. Concerning the SIT, we obtained control synthesis results through impulsive periodic release of controlled amplitude [10], and through optimal control approach [42]. Concerning *Wolbachia* technique, we investigated general control principles [39] capable of spreading the infection.

Wound healing 1: epithelial tissues

We studied cell motion in 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, *in vivo* wound healing and morphogenetic movements involving closure of holes in epithelia have been the object of many studies. In our works ⁰ we considered wound healing and epithelial gap closure in both *in vivo* (in particular *Drosophila* pupa) and *in vitro* (MDCK cell and human keratinocytes). We found some similarities in the geometry dependence of the wound closure strategies between these two situations, indicating the existence of conserved mechanisms that should be widespread across living beings. We are concentrating on the study of actin cable formation.

Wound healing 2: adipose tissues

After injury, if regeneration can be observed in hydra, planaria and some vertebrates, regeneration is rare in mammals and particularly in humans. In this research axis, we investigated the mechanisms by which biological tissues recover after injury. We explored this question on adipose tissue, using the mathematical framework recently developed in [144]. Our assumption is that simple mechanical cues between the Extra-Cellular Matrix (ECM) and differentiated cells can explain adipose tissue morphogenesis and that regeneration requires after injury the same mechanisms. We validated this hypothesis by means of a two-dimensional Individual Based Model (IBM) of interacting adipocytes and ECM fiber elements [145]. The model successfully generated regeneration or scar formation as functions of few key parameters, and seemed to indicate that the fate of injury outcome could be mainly due to ECM rigidity.

Following these encouraging results, the team is currently taking a step further in the model validation and confrontation to experimental data. The first direction concerns the development of a 3D framework to validate the mechanisms observed in 2D. In collaboration with S. Merino-Aceituno from the University of Vienna, efforts are being made in the development of a complete synthetic tissue model coupling the ECM and cell modelling with a vascularization model. A PhD project has been started to implement the coupled models and reduce the CPU time with the long-term goal to develop a usable software which would serve to investigate the role of different mechanisms in tissue development (not restricted to adipose tissues). Finally, further developments in collaboration with Imperial College London aim at pursuing the derivation of macroscopic PDE models from the agent-based formalisms.

Mathematical modelling of axolotl regeneration

Tissue response after injury/amputation induces one or two alternatives: scar formation versus regeneration (complete recovery of tissue shape and functions). In most mammals, regeneration is considered largely impaired for the benefit of a fibrotic scar after injury automatically associated with dysfunctions, but complete regeneration has been largely described and investigated in animal models such as zebra fish, salamander, or axolotl. Despite several processes regulating regeneration have been identified at different scales -from diffusing molecules and cellular gene expression patterns up to tissue mechanics-, how these mechanisms individually or collectively play a role in the regulation of regenerative processes remains poorly understood.

⁰[ravasio:hal-01245750](https://arxiv.org/abs/ravasio:hal-01245750), [vedula:hal-01298859](https://arxiv.org/abs/vedula:hal-01298859)

In order to give insights into the mechanisms of tissue regeneration, Valeria Caliaro started an Inria PhD project in October 2019, in collaboration with Osvaldo Chara, internationally recognized group leader of SysBio in Argentina. This project focuses on the role of cell proliferation in space and time along the two first phases of regeneration after injury: (i) initiation of a regeneration response, (ii) tissue patterning during regenerate growth. The first part of the project aims at building an agent-based model featuring few key mechanisms regulating cell proliferation after injury. The model construction is based on recent works where the authors developed a mathematical model given by ordinary differential equations (ODEs) [2] and a mathematical framework in 1D [3] showing that acceleration of the cell cycle is the major driver of regenerative spinal cord outgrowth in axolotls. Building on both mathematical models and introducing heuristic rules which rely on Prof O. Chara expertise, we propose a 2D-ABM using methodologies borrowed from socio-dynamics and collective behavior studies (based on many interacting agent systems). While the focus is made on proliferation-based mechanisms, other mechanisms responsible for collective behavior such as volume exclusion, diffusion or aggregation will be tested and compared with experimental data. The resulting model will provide a synthetic tissue model which will serve to investigate regeneration in cellular systems, focusing on cell proliferation properties. The second part of the PhD will be devoted to the derivation of continuous models from the agent-based formalism. This will provide a large scale ‘synthetic tissue’ model to explore the role of large scale effects in general tissue models. Model validation and calibration will be ensured by quantitative comparison with biological data already present in the literature and generated by the SysBio group of O. Chara, particularly the representative images of regenerative spinal cords after tail amputation. By varying the model parameters and observing the resulting alteration of the spinal cord size and architecture as a consequence of these variations, it will be possible to provide ‘in silico’ setting experiments to guide and plan future in vivo or ex vivo experiments. Altogether, the project is expected to provide a mechanistic understanding of the cellular mechanisms driving spinal cord regeneration, and to identify how spatial structuration can influence cell differentiation and growth.

[1] Rodrigo Albors A, Tazaki A, Rost F, Nowoshilow S, Chara O & Tanaka EM. 2015. Planar cell polarity-mediated induction of neural stem cell expansion during axolotl spinal cord regeneration. *Elife*. 4:e10230. [2] Rost F, Rodrigo Albors A, Mazurov V, Brusch L, Deutsch A, Tanaka EM & Chara O. 2016. Accelerated cell divisions drive the outgrowth of the regenerating spinal cord in axolotls. *Elife*. 5. pii: e20357. [3] Cura Costa E, Rodrigo Albors A, Tanaka EM & Chara O. Spatiotemporal distribution of cell proliferation rules spinal cord regeneration in the axolotl. MS in preparation.

Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids

Model simulations indicate that the response of growing cell populations on mechanical stress follows the same functional relationship and is predictable over different cell lines and growth conditions despite experimental response curves look largely different. We developed a hybrid model strategy in which cells are represented by coarse-grained individual units calibrated with a high resolution cell model and parameterized by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics from the growth kinetics in absence of external stress. Our model simulation results suggest a generic, even quantitatively same, growth response of cell populations upon externally applied mechanical stress, as it can be quantitatively predicted using the same growth progression function ([52])

Modeling of morphogen diffusion in *Drosophila* oogenesis

In collaboration with a team of developmental biologists of Rutgers University (Camden, New Jersey), we have built a model for the diffusion of the Gurken morphogen during *Drosophila* oogenesis, taking into account a wide variety of biological mechanisms such as diffusion of the morphogen, reactions of components of the EGFR signaling pathway, movement of the source of morphogen, shift of the overlying follicle cells and growth of the egg chamber. This model, together with a complete numerical code developed in Matlab,

provides a tool to understand how each mechanism influences the signal distribution. The overall aim of the project is to use this tool to guide future experiments, and to understand what mechanisms contribute to the different distributions of signal among species.

Bacterial population growth

We exploited all the methods developed to estimate the division rate of a population (see axis 3) to address a seminal question of biology: is it a size-sensing or a timing mechanism which triggers bacterial growth? In [148], we showed that a sizer model is robust and fits the data well. Several studies from other groups came at the same time, showing a renewed interest on a question dated back to Jacques Monod's PhD thesis (1941). Of special interest is the "adder" model, for which we are currently developing new estimation methods [46].

A quantitative high resolution computational mechanics cell model for growing and regenerating tissues

Mathematical models are increasingly designed to guide experiments in biology, biotechnology, as well as to assist in medical decision making. They are in particular important to understand emergent collective cell behavior. For this purpose, the models, despite still abstractions of reality, need to be quantitative in all aspects relevant for the question of interest. Considered as a showcase example the regeneration of liver after drug-induced depletion of hepatocytes, in which the surviving and dividing hepatocytes must squeeze in between the blood vessels of a network to refill the emerged lesions. Here, the cells' response to mechanical stress might significantly impact the regeneration process. We present a 3D high-resolution cell-based model integrating information from measurements in order to obtain a refined and quantitative understanding of the impact of cell-biomechanical effects on the closure of drug-induced lesions in liver. Our model represents each cell individually and is constructed by a discrete, physically scalable network of viscoelastic elements, capable of mimicking realistic cell deformation and supplying information at subcellular scales. The cells have the capability to migrate, grow, and divide, and the nature and parameters of their mechanical elements can be inferred from comparisons with optical stretcher experiments. Due to triangulation of the cell surface, interactions of cells with arbitrarily shaped (triangulated) structures such as blood vessels can be captured naturally. Comparing our simulations with those of so-called center-based models, in which cells have a largely rigid shape and forces are exerted between cell centers, we find that the migration forces a cell needs to exert on its environment to close a tissue lesion, is much smaller than predicted by center-based models. To stress generality of the approach, the liver simulations were complemented by monolayer and multicellular spheroid growth simulations. In summary, our model can give quantitative insight in many tissue organization processes, permits hypothesis testing in silico, and guide experiments in situations in which cell mechanics is considered important [123].

Liver regeneration and disease: towards a full virtual liver model at histological scale

In our work towards a full virtual liver model at histological level, a number of steps were performed. The models under points (1)-(4) focus on either a single or a few liver lobules. A liver lobule is the smallest repetitive functional and anatomical building block of liver, while (5) addresses a much larger organisational building block of the liver, a liver lobe that consists of thousands to hundreds of thousands of lobules depending on the species. A second strand (6), (7) addresses image analysis, which in most cases forms the entrance to modeling as it provides the data necessary to generate model hypotheses and to parameterize a model.

(1) Cell types: In a former work by Hoehme et. al. ([113]) a model of liver regeneration after drug-induced damage was established considering hepatocytes and blood vessels. This model has now been expanded to include all relevant cell types, including hepatocytes, blood vessels, hepatic stellate cells, Kupffer cells, invading macrophages and other immune cells. Thereby it is now possible to study perturbations in the temporal scenario of damage and regeneration after signaling events or cells types are knocked down individually or collectively. This model is currently compared to respective perturbation experiments.

(2) Liver disease: Degenerative liver diseases such as liver fibrosis and cirrhosis develop out of a disturbed balance of degenerative and regenerative processes. The model under (1) has thereby been extended by the formation of extracellular matrix, mimicked as fiber networks, to capture the disease process leading to liver fibrosis. In that process characteristic streets form that modify the mechanics, perfusion behavior and detoxification capacity of the liver.

(3) Consequence of liver fibrosis: Whole-slide scans from fibrotic liver in a mouse model has been analysed at different time points after emergence of the disease with regard to the degree of excess matrix to mimic the possible consequences of fibrotic inclusions on perfusion and function of liver within a multiscale model that considers ammonia detoxification in each individual hepatocyte as well as blood flow and transport processes in the liver lobule.

(4) Bile flux: Bile flux has been for decades believed to be controlled by convection at the level of liver lobules as well as at the level of the entire organ. By a methodology based on correlative imaging for quantitative intravital flux analysis no directed advection was detectable in bile canaliculi at the resolution limit. Instead, after active transport across hepatocyte membranes bile salts within the liver lobules are transported in the canaliculi by a diffusion-dominated process. Only in the interlobular ducts i.e., at super-lobular level, diffusion is augmented by advection. In silico simulations of bile transport in real 3D bile network microarchitectures can quantitatively explain the data assuming diffusive transport as sole mechanism.

(5) Liver regeneration after partial hepatectomy (partial organ removal): Partial hepatectomy is an adequate therapy in case of diseases or events that destructed only part of the liver. A typical case is a primary tumor or a metastasis affecting only a single liver lobe. Within an biophysical agent-based model capturing many aspects of the cell mechanics we studied regrowth of liver after partial organ removal in mouse calibrated with multivariate experimental data. Our model predicts characteristic proliferation pattern that change from small animals (as mouse) to large animals (as pig).

(6) Bile duct ligation: Bile duct ligation (BDL) is an experimental procedure that mimics obstructive cholestatic disease. One of the early consequences of BDL in rodents is the appearance of so-called bile infarcts that correspond to Charcot-Gombault necrosis in human cholestasis. The mechanisms causing bile infarcts and their pathophysiological relevance are unclear. Therefore, intravital two photon-based imaging of BDL mice was performed with fluorescent bile salts (BS) and non-BS organic anion analogues. Key findings were followed up by matrix-assisted laser desorption ionization imaging, clinical chemistry, immunostaining, and gene expression analyses. Our group performed analysis of intravital imaging. The key finding is that bile microinfarcts occur in the acute phase after BDL in a limited number of dispersed hepatocytes followed by larger infarcts involving neighboring hepatocytes, and they allow leakage of bile from the BS-overloaded biliary tract into blood, thereby protecting the liver from BS toxicity; in the chronic phase after BDL, reduced sinusoidal BS uptake is a dominant protective factor, and the kidney contributes to the elimination of BS until cholemic nephropathy sets in [108].

(7) Periportalisation during liver fibrosis formation: Within a liver lobule, the function of hepatocytes is zoned i.e., certain functions are only executed by either hepatocytes close to the center (pericentral region) or hepatocytes in the periphery of the lobule (periportal region). Little is known about how liver fibrosis influences lobular zonation. To address this question, three mouse models of liver fibrosis were used, CCl₄ administration repeated for 2, 6 and 12 months to induce pericentral damage, as well as bile duct ligation (21 days) and a particular *mdr2*-mouse model to study periportal fibrosis. Analyses were performed by RNA-sequencing, immunostaining of zoned proteins and image analysis. Image analysis was performed by our group. The key result was that liver fibrosis leads to strong alterations of lobular zonation, where the pericentral region adopts periportal features. Beside adverse consequences, periportalization supports adaptation to repeated doses of hepatotoxic compounds [17].

Toxicity extrapolation from in vitro to in vivo

In vivo toxicity prediction from in vitro data is a major objective in toxicology as it permits bypassing animal experiments, and as the predictive power of animal experiments for human is limited. Objective was the prediction of paracetamol (acetaminophen)-induced hepatotoxicity from in vitro experiments. For this purpose, numerous iterations between in vitro experiments, in vivo experiments and simulations were performed for mouse. Using a recent thesis (Géraldine Cellière's PhD thesis [88]) as a start point, two candidate mechanisms could be identified both explaining the in vivo data after calibration of the in silico model with in vitro toxicity data.

Relating imaging on microscopic scales with imaging on macroscopic scales: From Diffusion-Weighted MRI Calibrated With Histological Data: an Example From Lung Cancer

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment, reflecting Brownian movement of water molecules in tissues. Since densely packed cells restrict molecule mobility, tumor tissues produce usually higher signal (less attenuated signal) on isotropic maps compared with normal tissues. However, no general quantitative relation between DWI data and the cell density has been established. In order to link low-resolution clinical cross-sectional data with high resolution histological information, we developed an image processing and analysis chain, which was used to study the correlation between the diffusion coefficient (D value) estimated from DWI and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor. Color deconvolution followed by cell nuclei segmentation was performed on digitized histological images to determine local and cell-type specific 2d (two-dimensional) densities. From these, the 3d cell density was inferred by a model-based sampling technique, which is necessary for the calculation of local and global 3d tumor cell count. Next, DWI sequence information was overlaid with high resolution CT data and the resected histology using prominent anatomical hallmarks for co-registration of histology tissue blocks and non-invasive imaging modalities' data. The integration of cell numbers information and DWI data derived from different tumor areas revealed a clear negative correlation between cell density and D value. Importantly, spatial tumor cell density can be calculated based on DWI data. In summary, our results demonstrate that tumor cell count and heterogeneity can be predicted from DWI data, which may open new opportunities for personalized diagnosis and therapy optimization [157]. The work of that paper has been further advanced to adapt the procedures for clinical use (in preparation).

Collaborations

- Biological control of arboviruses: **Nicolas Vauchelet** (Université Paris 13); **Grégoire Nadin** (LJLL, Sorbonne Université); **Yannick Privat** (Université de Strasbourg); **D. Villela**, **C. Struchiner** (Fiocruz, Brazil); **Jorge Zubelli** (IMPA, Brazil); **Alain Rapaport** (INRA-Montpellier), **Y. Dumont** (CIRAD-Montpellier); **Ch. Schaerer**, **P. Pérez-Estigarribia** (UNA, Paraguay), **O. Vasilieva** (Universidad del Valle, Cali, Colombia), **D. Cardona-Salgado** (Universidad Autónoma de Occidente, Cali, Colombia).
- Protein aggregation in amyloid diseases: **Human Rezaei**'s team at Inra Jouy-en-Josas (France) and **W-F Xue**'s team in at university of Kent (Great Britain); **Tom Banks** at the North Carolina State University (USA) and **Philippe Moireau** (M3DISIM)
- Bacterial growth and division: **Lydia Robert**, Sorbonne Université (France)
- Liver research & toxicology: **JG. Hengstler** group (IfADo, Dortmund, Germany); **R. Gebhardt** (Univ. Leipzig); **U. Klingmueller** (DKFZ, Heidelberg); **Irène Vignon-Clementel** (Inria, COMME-DIA)
- Growth in capsules and biomechanics: **Pierre Nassoy**, Institut dOptique Graduate School, Talence, France; **Josef Kaes**, Peter Debye Institute for Soft Matter Physics, Physics, Univ. Leipzig, Germany.
- Wound healing: **Patrizia Bagnerini** (Genova, Numerical methods), **Benoît Ladoux** (Institut Jacques Monod et Mechanobiology Institute Singapore, Biophysics) and **Antonio Jacinto** (CEDOC, Lisbon, Biology and Medicine). (Adipose tissue regeneration) team of **L. Casteilla** (StromaLab, Toulouse). (Axolotl regeneration) team of O. Chara, SysBio group, Argentina.
- Diffusion of morphogen: Center for Computational and Integrative Biology, Rutgers University (Camden, New Jersey), joint work with Professor Nir Yakoby's Drosophila Laboratory
- Linking micro and macro-image information: Oliver Sedlacek, Univ. and DKFZ Heidelberg, Kai Breuhahn, Univ. Heidelberg.

MATHNEURO Project-Team (section vide)

MIMESIS Team

4. Application Domains

4.1. Surgical training

Virtual training helps medical students to get familiar with surgical procedures before manipulation of real patients. The development of simulation used for medical training usually requires important computational power, since realistic behaviors are key to deliver a high-fidelity experience to the trainee. Further, the quality of interaction with the simulator (usually via visual and haptic rendering) is also of paramount importance. All these constraints make the development of training systems time-consuming, thus limiting the deployment of virtual simulators in standard medical curriculum.

4.2. Pre-operative planning

Beyond training, clinicians ask for innovative tools that can assist them in the pre-operative planning of an intervention. Using the patient information acquired before the operation, physics-based simulations allow to simulate the effect of therapy with no risk to the patient. The clinicians can thus virtually assess different strategies and select the optimal procedure. Compared to a training simulation, a planning system requires a high accuracy to ensure reliability. Constrained by the time elapsed between the preoperative acquisition and the intervention, the computation must also be efficient.

4.3. Intra-operative guidance

Besides the surgery training and planning, another major need from clinicians is surgical guidance. While the clinician is performing the operation, a guidance system provides enriched visual feedback. This is especially useful with the emergence of minimally invasive surgery (MIS) where the visual information is often strongly limited. It can be used for example to avoid critical areas such as vessels or to highlight the position of a tumor during its resection. In the MIS technique, the clinician does not interact with organs directly as in the open surgery, but manipulates instruments inserted through trocars placed in small incisions in the wall of the abdominal cavity. The surgeon can observe these instruments on a display showing a video stream captured by an endoscopic camera inserted through the navel. The main advantage of the method resides in reducing pain and time recovery, in addition to reducing bleeding and risks of infection. However, from a surgical standpoint, the procedure is quite complex since the field of view is considerably reduced and the direct manipulation of organs is not possible.

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 and related domains like for example educational science.

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 extracortical regions. The variety of these regions, together with large mnesic 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.

MONC Project-Team

4. Application Domains

4.1. Tumor growth monitoring and therapeutic evaluation

Each type of cancer is different and requires an adequate model. More specifically, we are currently working on the following diseases:

- Glioma (brain tumors) of various grades,
- Metastases to the lung, liver and brain from various organs,
- Soft-tissue sarcoma,
- Kidney cancer and its metastases,
- non small cell lung carcinoma.

In this context our application domains are

- Image-driven patient-specific simulations of tumor growth and treatments,
- Parameter estimation and data assimilation of medical images.
- Machine and deep learning methods for delineating the lesions and stratifying patients according to their responses to treatment or risks of relapse.

4.2. Biophysical therapies

- Modeling of electrochemotherapy on biological and clinical scales.
- Evaluation of radiotherapy and radiofrequency ablation.

4.3. In-vitro and animals experimentations in oncology

- 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.

MORPHEME Project-Team (section vide)

MOSAIC Project-Team (section vide)

NEUROSYS Project-Team

4. Application Domains

4.1. Medical applications

Our research directions are motivated by applications with a high healthcare or social impact. They are developed in collaboration with medical partners, neuroscientists and psychologists. Almost all of our applications can be seen as neural interfaces which require the analysis and modeling of sensorimotor rhythms.

4.1.1. *Per-operative awareness during general anesthesia*

Collaborators: Univ. Hospital of Nancy-Brabois/dept. Anesthesia & Resuscitation

During general anesthesia, brain oscillations change according to the anesthetic drug concentration. Nowadays, 0.2 to 1.3% of patients regain consciousness during surgery and suffer from post-traumatic disorders. Despite the absence of subject movements due to curare, an electroencephalographic analysis of sensorimotor rhythms can help to detect an intention of movement. Within a clinical protocol, we are working on a brain-computer interface adapted to the detection of intraoperative awareness.

4.1.2. *Recovery after stroke*

Collaborators: Regional Institute of Physical Medicine and Rehabilitation/Center for Physical Medicine and Rehabilitation (Lay St Christophe), Univ. of Lorraine/PERSEUs.

Stroke is the main cause of acquired disability in adults. Neurosys aims at recovering limb control by improving the kinesthetic motor imagery (KMI) generation of post-stroke patients. We propose to design a KMI-based EEG neural interface which integrates complementary modalities of interactions such as tangible and haptic ones to stimulate the sensorimotor loop. This solution would provide a more engaging and compelling stroke rehabilitation training program based on KMI production.

4.1.3. *Modeling Parkinson's disease*

Collaborators: Center for Systems Biomedicine (Luxembourg), Institute of Neurodegenerative Diseases (Bordeaux), Human Performance & Robotics laboratory (California State Univ., Long Beach).

Effective treatment of Parkinson's disease should be based on a realistic model of the disease. We are currently developing a neuronal model based on Hodgkin-Huxley neurons reproducing to a certain extent the pathological synchronization observed in basal ganglia in Parkinsonian rats. Moreover, our mesoscopic models of plastic Central Pattern Generator neural circuitries involved in rhythmic movements will allow us to reproduce incoherent coordination of limbs observed on humans affected by Parkinson's diseases like frozen gait, crouch gait. Our long-term objective is to understand how oscillatory activity in the basal ganglia affects motor control in spinal structures.

4.1.4. *Modeling propagation of epileptic spikes*

Collaborators: Epileptology Unit of the CHRU Nancy (University hospital), CRAN (Research Center in Automation and Signal Processing of Nancy). Effective treatment of patients with refractory epilepsy requires a better understanding of the underlying neuronal mechanisms. In particular, it has been observed that epileptic spikes propagate more easily during stage III sleep (slow wave sleep) than during wakefulness, but the origin of these behaviours still remains misunderstood. At least both, a combination of anatomical structure/connectivity changes and changes in level of neurotransmitters, namely functional connectivity, can cause the propagation. A better knowledge of the functional and structural circuitry could allow a better targeting of structures to be treated, either surgically or pharmacologically, and to better individually adapt the pharmacology to each patient according to their symptomatology.

NUMED Project-Team (section vide)

OPIS Project-Team

4. Application Domains

4.1. Sparse signal processing in chemistry

Participants: Marc Castella, Emilie Chouzenoux, Arthur Marmin, Jean-Christophe Pesquet (Collaboration: Laurent Duval, IFPEN, Rueil Malmaison)

Mass Spectrometry (MS) is a powerful tool used for robust, accurate, and sensitive detection and quantification of molecules of interest. Thanks to its sensibility and selectivity, MS is widely used in proteomics such as anti-doping, metabolomics, medicine or structural biology. In particular, it has applications in clinical research, personalized medicine, diagnosis process and tumours profiling and pharmaceutical quality control. In an MS experiment, the raw signal arising from the molecule ionization in an ion beam is measured as a function of time via Fourier Transform-based measures such as Ion Cyclotron Resonance (FT-ICR) and Orbitrap. A spectral analysis step is then performed to improve the quality of data. The goal is then to determine from this observed pattern distribution the most probable chemical composition of the sample, through the determination of the monoisotopic mass, charge state and abundance of each present molecule. This amounts to solve a large scale signal estimation problem under specific sparsity constraints [35], [55]. Collaboration with Dr. L. Duval, Research Engineer at IFP Energies Nouvelles, France is on-going in this applicative context.

4.2. Image restoration for two-photon microscopy

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Mathieu Chalvidal (Collaboration: Claire Lefort, XLIM, CNRS, Limoges)

Through an ongoing collaboration with physicists from XLIM laboratory (CNRS, Limoges, France), we propose advanced mathematical and computational solutions for multiphoton microscopy (MPM) 3D image restoration. This modality enjoys many benefits such as a decrease in phototoxicity and increase in penetration depth. However, blur and noise issues can be more severe than with standard confocal images. Our objective is to drastically improve the quality of the generated images and their resolution by improving the characterization of the PSF of the system [12] and compensating its effect. We consider the application of the improved MPM imaging tool to the microscopic analysis of muscle ultrastructure and composition, with the aim to help diagnosing muscle disorders including rare and orphan muscle pathologies.

4.3. Representation Learning for Biological Networks

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat (Collaboration: Duong Nguyen, UC San Diego)

Networks (or graphs) are ubiquitous in the domain of biology, as many biological systems can naturally be mapped to graph structures. Characteristic examples include protein-protein interaction and gene regulatory networks. To this extend, machine learning on graphs is an important task with many practical applications in network biology. For example, in the case on protein-protein interaction networks, predicting the function of a protein is a key task that assigns biochemical roles to proteins. The main challenge here is to find appropriate representations of the graph structure, in order to be easily exploited by machine learning models. The traditional approach to the problem was relying on the extraction of “hand-crafted” discriminating features that encode information about the graph, based on user-defined heuristics. Nevertheless, this approach has demonstrated severe limitations, as the learning process heavily depends on the manually extracted features. To this end, feature (or representation) learning techniques can be used to automatically learn to encode the graph structure into low-dimensional feature vectors – which can later be used in learning tasks. Our goal here is to develop a systematic framework for large-scale representation learning on biological graphs. Our approach takes advantage of the clustering structure of these networks, to further enhance the ability of the learned features to capture intrinsic structural properties.

4.4. Breast tomosynthesis

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)

Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing is modeled [50].

4.5. Inference of gene regulatory networks

Participants: Surabhi Jagtap, Fragkiskos Malliaros, Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)

The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in different living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

4.6. Imaging biomarkers and characterization for chronic lung diseases

Participants: Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Marie-Pierre Revel and Nikos Paragios: AP-HP - Hopital Cochin Broca Hotel Dieu; Therapanacea)

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i) severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by deep convolutional neural networks and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

4.7. Imaging radiomics and genes to assess immunotherapy

Participants: Samy Ammari, Enzo Batistella, Emilie Chouzenoux, Théo Estienne, Marvin Lerousseau, Hugues Talbot, Roger Sun, Maria Vakalopoulou (in collaboration with Corinne Balleyguier, Caroline Caramella, Éric Deutsch, Nathalie Lassau, Institut de Cancérologie Gustave Roussy, Nikos Paragios, Therapanacea)

Because responses of cancer patients to immunotherapy can vary considerably, innovative predictors of response to treatment are urgently needed to improve patients outcomes.

We have aimed to develop and independently validate a radiomics-based biomarkers of tumour-infiltrating CD8 cells in patients included in phase 1 trials of anti-programmed cell death protein (PD)-1 or anti-programmed cell death ligand 1 (PD-L1) mono-therapy. We also aimed to evaluate the association between the biomarker, tumour immune phenotype and clinical outcomes of these patients.

Concurrently, we have evaluated various ways of estimating patient response to treatment based on well-established radiomics such as estimated tumour count and volumes. Among published metrics, we have select those that shown good predictive power and proposed a new one, which is particularly effective for patient with a poor response [63].

Furthermore, we have developed and validated a novel imaging-based decision-making algorithm for use by the clinician that helps differentiate pituitary metastasis from autoimmune hypophysitis in patients undergoing immune checkpoint blockade therapy [21].

These works are in collaboration with the Institut de Cancérologie Gustave Roussy Paris.

4.8. Development of a heart ventricle vessel generation model for perfusion analysis

Participant: Hugues Talbot (collaboration with L. Najman ESIEE Paris, I. Vignon-Clementel, REO Team leader, Inria, Charles Taylor, Heartflow Inc.)

Cardio-vascular diseases are the leading cause of mortality in the world. Understanding these diseases is a current, challenging and essential research project. The leading cause of heart malfunction are stenoses causing ischemia in the coronary vessels. Current CT and MRI technology can assess coronary diseases but are typically invasive, requiring catheterization and relatively toxic contrast agents injection. In collaboration with the REO team headed by Irène Vignon-Clementel, and Heartflow, a US based company, we have in the past worked to use image-based exams only, limiting the use of contrast agents and in many cases eliminating catheterisation. Heartflow is current the market leader in non-invasive coronary exams.

Unfortunately, current imaging technology is unable to assess the full length of coronary vessels. CT is limited to a resolution of about 1mm, whereas coronary vessels can be much smaller, down to about 10 micrometers in diameter. Blood perfusion throughout the heart muscle can provide insight regarding coronary health in areas that CT or MRI cannot assess. Perfusion imaging with PET or a Gamma camera, the current gold standard, is an invasive technology requiring the use of radioactive tracers.

We have investigated patient-specific vessel generation models together with porous model simulations in order to propose a forward model of perfusion imaging, based on the known patient data, computer flow dynamic simulations as well as experimental data consistent with known vessel and heart muscle physiology. The objective of this work is to both provide a useful, complex forward model of perfusion image generation, and to solve the inverse problem of locating and assessing coronary diseases given a perfusion exam, even though the affected vessels may be too small to be imaged directly.

In 2019, we have produced a functional myocardial perfusion model consisting of the CT-derived segmented coronary vessels, a simulated vessel tree consisting of several thousands of terminal vessels, filling the myocardium in a patient-specific way, consistent with physiology data, physics-based and empirically-observed vessel growth rules, and a porous medium. We have produced a CFD code capable of simulating blood flow in all three coupled compartments, which allows us to simulate perfusion realistically.

PARIETAL Project-Team

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. *Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)*

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electrophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain, such as: How is the function of cortical territories constrained by anatomy (connectivity)? How to assess the specificity of brain regions? How can one characterize reliably inter-subject differences?

4.1.2. *Analysis of brain Connectivity*

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. *Modeling of brain processes (MEG)*

While fMRI has been very useful in defining the function of regions at the mm scale, Magnetoencephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. *Current challenges in human neuroimaging (acquisition+analysis)*

Human neuroimaging targets two major goals: *i*) the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii*) the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial⁰ resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.
- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

⁰and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements⁰ and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [72], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [73]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as **NeuroVault** or **OpenfMRI**. A framework for integrating such datasets is however still missing.

⁰The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

PLEIADE Project-Team

4. Application Domains

4.1. Genome and transcriptome annotation, to model function

Sequencing genomes and transcriptomes provides a picture of how a biological system can function, or does function under a given physiological condition. Simultaneous sequencing of a group of related organisms is now a routine procedure in biological laboratories for studying a behavior of interest, and provides a marvelous opportunity for building a comprehensive knowledge base of the relations between genomes. Key elements in mining these relations are: classifying the genes in related organisms and the reactions in their metabolic networks, recognizing the patterns that describe shared features, and highlighting specific differences.

PLEIADE will develop applications in comparative genomics of related organisms, using new mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on computational geometry refine these comparisons. Compact representations can be stored, exchanged, and combined. They will form the basis of new simultaneous genome annotation methods, linked directly to abductive inference methods for building functional models of the organisms and their communities.

Our ambition in biotechnology is to permit the design of synthetic or genetically selected organisms at an abstract level, and guide the modification or assembly of a new genome. Our effort is focused on two main applications: genetic engineering and synthetic biology of oil-producing organisms (biofuels in CAER, palm oils), and improving and selecting starter microorganisms used in winemaking (collaboration with the ISVV and the BioLaffort company).

4.1.1. Oil Palm lipid synthesis

The largest source of vegetable oil⁰ is the fruit mesocarp of the oil palm *Elaeis guineensis*, a remarkable tissue that can accumulate up to 90% oil, the highest level observed in the plant kingdom. The market share of oil palm is expected to increase in order to meet increased demand for vegetable oil, predicted to double by 2030 [18], be it as food or as a source of biofuels in Africa. A significant proportion of palm oil is produced on small estates that do not have access to efficient milling facilities, and run a great risk of spoilage through oil acidification. Improving palm oil quality through genetics and selection will result in economic gains [24] by addressing several targets such as improvement of oil yield, tuning of oil quality through the rate of unsaturated fatty acids or impairment of degradation processes. Furthermore, as genome biodiversity resides mostly in Africa, oil from African oil palms can vary greatly in fatty acid composition according to cultivar genetic differences and to weather conditions, and the precise mechanisms regulating this variability are not yet understood.

A growing body of molecular resources for studying oil palm fruit are making it possible to study and improve the quality and quantity of oil produced by oil palms. In particular, these oils can vary greatly in fatty acid composition, and while the precise mechanisms regulating this variability are not completely understood, establishing a link between oil palm genotype and phenotype appears increasingly feasible. PLEIADE will work with the CNRS/UB UMR 5200 (LBM), a laboratory with an established reputation in studying fatty acid metabolism in *E. guineensis*, to improve understanding of the links between genetic diversity and oil production, and participate in developing applications.

4.1.2. Engineering pico-algae

⁰32% of the world market share [24]

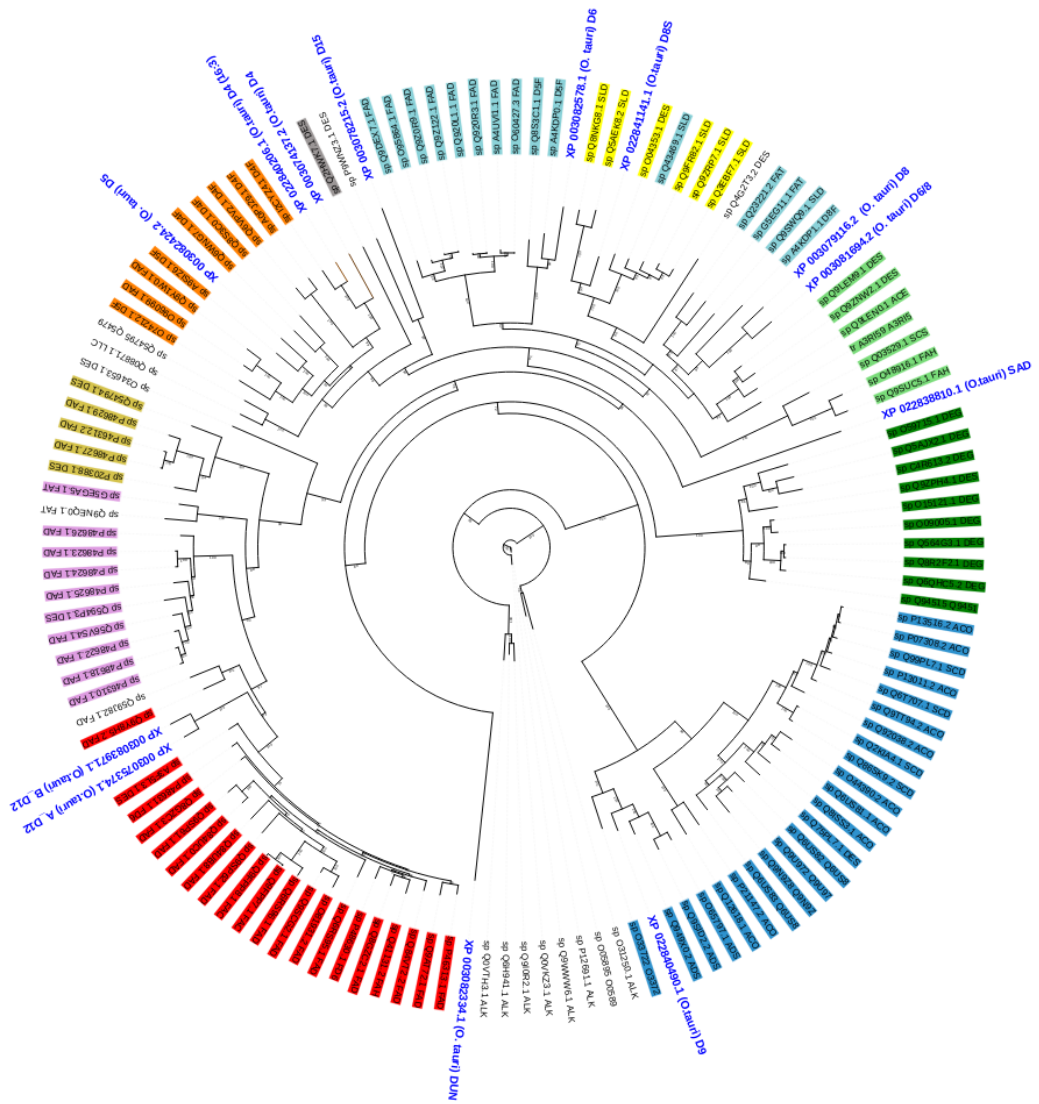


Figure 4. Phylogenetic structure of long-chain polyunsaturated fatty acid desaturase specificity. Highlighted are thirteen desaturases from *Ostreococcus tauri*

Docosahexaenoic acid (DHA) is an essential nutriment for human brain tissue and can only be obtained from marine or riverine fish that live on phytoplankton and zooplankton, since human neurons lack the delta desaturase required for *de novo* synthesis of DHA. [21] Unfortunately, fishing is become less and less a sustainable resource. Since phytoplankton and zooplankton are the ultimate source of DHA consumed, there is considerable interest in obtaining DHA directly rather than through the intermediary of fish. A very promising approach is through the bio-engineering of pico-algae.

In order to produce the long-chain polyunsaturated fatty acids (LC-PUFA) needed for human nutrition, it is necessary to precisely engineer the desaturases that produce them. Desaturases are enzymes responsible for the introduction of double bonds into fatty acids. Desaturases are specific in recognizing their substrates and in placing the double bond in the proper place. The desaturases that produce the LC-PUFA necessary for human nutrition are present only in some species.

Our goal is to design methods to predict the substrate and region specificities for desaturases in algal species, particularly *Ostreococcus tauri*, the smallest photosynthetic eukaryote that can be cultivated. Thirteen desaturases are known in *O. tauri* and can be placed in the phylogeny of the desaturase family (figure 4). The biochemical and structural characterization of these enzymes is as yet very incomplete. This work is ongoing (see section 8.2.2) and requires close collaboration between biologists and computer scientists.

4.2. Molecular based systematics and taxonomy

Defining and recognizing myriads of species in biosphere has taken phenomenal energy over the past centuries and remains a major goal of Natural History. It is an iconic paradigm in pattern recognition (clustering has coevolved with numerical taxonomy many decades ago). Developments in evolution and molecular biology, as well as in data analysis, have over the past decades enabled a profound revolution, where species can be delimited and recognized by data analysis of sequences. We aim at proposing new tools, in the framework of E-science, which make possible (i) better exploration of the diversity in a given clade, and (ii) assignment of a place in these patterns for new, unknown organisms, using information provided by sets of sequences. This will require investment in data analysis, machine learning, and pattern recognition to deal with the volumes of data and their complexity.

One example of this project is about the diversity of trees in Amazonian forest, in collaboration with botanists in French Guiana. Protists (unicellular Eukaryotes) are by far more diverse than plants, and far less known. Molecular exploration of Eukaryotes diversity is nowadays a standard in biodiversity studies. Data are available, through metagenomics, as an avalanche and make molecular diversity enter the domain of Big Data. Hence, an effort will be invested, in collaboration with other Inria teams (GenScale, HiePACS) for porting to HPC algorithms of pattern recognition and machine learning, or distance geometry, for these tools to be available as well in metagenomics. This will be developed first on diatoms (unicellular algae) in collaboration with INRA team at Thonon and University of Uppsala), on pathogens of tomato and grapewine, within an existing network, and on bacterial communities, in collaboration with University of Pau. For the latter, the studies will extend to correlations between molecular diversity and sets of traits and functions in the ecosystem.

4.3. Community ecology and population genetics

Community assembly models how species can assemble or disassemble to build stable or metastable communities. It has grown out of inventories of countable organisms. Using *metagenomics* one can produce molecular based inventories at rates never reached before. Most communities can be understood as pathways of carbon exchange, mostly in the form of sugar, between species. Even a plant cannot exist without carbon exchange with its rhizosphere. Two main routes for carbon exchange have been recognized: predation and parasitism. In predation, interactions—even if sometimes dramatic—may be loose and infrequent, whereas parasitism requires what Claude Combes has called intimate and sustainable interactions [19]. About one decade ago, some works [25] have proposed a comprehensive framework to link the studies of biodiversity with community assembly. This is still incipient research, connecting community ecology and biogeography.

We aim at developing graph-based models of co-occurrence between species from NGS inventories in metagenomics, i.e. recognition of patterns in community assembly, and as a further layer to study links, if any, between diversity at different scales and community assemblies, starting from current, but oversimplified theories, where species assemble from a regional pool either randomly, as in neutral models, or by environmental filtering, as in niche modeling. We propose to study community assembly as a multiscale process between nested pools, both in tree communities in Amazonia, and diatom communities in freshwaters. This will be a step towards community genomics, which adds an ecological flavour to metagenomics.

Convergence between the processes that shape genetic diversity and community diversity—drift, selection, mutation/speciation and migration—has been noted for decades and is now a paradigm, establishing a continuous scale between levels of diversity patterns, beyond classical approaches based on iconic levels like species and populations. We will aim at deciphering diversity pattern along these gradients, connecting population and community genetics. Therefore, some key points must be addressed on reliability of tools.

Next-generation sequencing technologies are now an essential tool in population and community genomics, either for making evolutionary inferences or for developing SNPs for population genotyping analyses. Two problems are highlighted in the literature related to the use of those technologies for population genomics: variable sequence coverage and higher sequencing error in comparison to the Sanger sequencing technology. Methods are developed to develop unbiased estimates of key parameters, especially integrating sequencing errors [23]. An additional problem can be created when sequences are mapped on a reference sequence, either the sequenced species or an heterologous one, since paralogous genes are then considered to be the same physical position, creating a false signal of diversity [20]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [20] or by filtering detected SNPs. Finally, an increasingly popular method (RADseq) is used to develop SNP markers, but it was shown that using RADseq data to estimate diversity directly biases estimates [11]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.

REO 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.* [30], [34], [33] 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.

SERENA Project-Team

4. Application Domains

4.1. Multiphase flows and transport of contaminants in the subsurface

- subsurface depollution after chemical leakage
- nuclear waste disposal in deep underground repositories
- flow in large scale discrete fracture networks
- production of oil and gas

4.2. Industrial risks in energy production

- Stokes and Navier–Stokes flows related to nuclear reactor operation
- seismic wave propagation for detection and protection
- electromagnetism for interfaces between dielectrics and negative metamaterials

4.3. Nonlinear mechanics

- quasi-static and dynamic elastoplastic evolutions with small and large deformations
- quasi-static and dynamic crack propagation
- nonlinear contact and friction conditions
- application to engineering components mainly related to nuclear reactor operation and safety analysis

4.4. Computational quantum chemistry

- guaranteed bounds for ground-state energy (eigenvalues) and ground-state density matrix (eigenvectors) in first-principle molecular simulation
- application to Laplace, Gross–Pitaevskii, Kohn–Sham, and Schrödinger models

SERPICO Project-Team

4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale

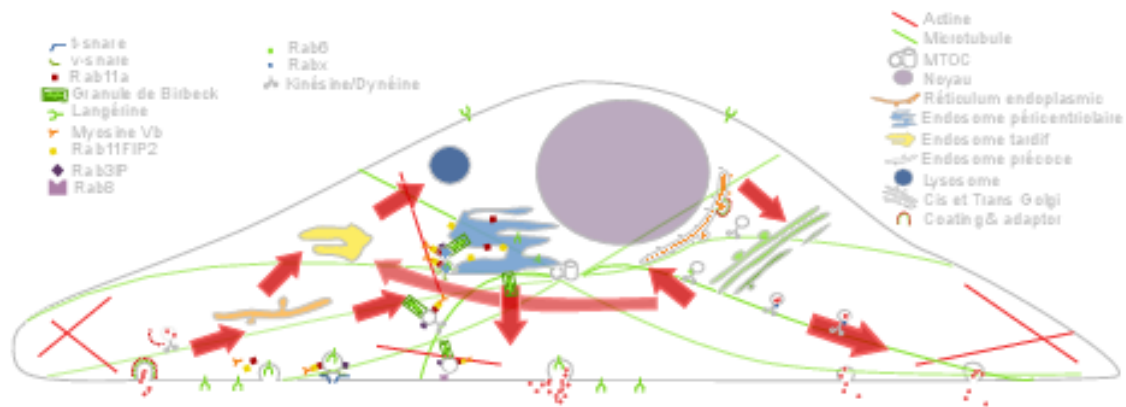


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried out together with the “Space Time imaging of Endomembranes and organelles Dynamics” team at CNRS-UMR 144 Institut Curie contributed to a better understanding of the intracellular compartmentation, particularly in specialized model cells such as melanocytes and Langerhans cells of the epidermis, of the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively and to the understanding on how the dynamics of those structures relate to their physiological functions. These studies have started to highlight: i/ the measurement of 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, an issue that become more and more accessible, thanks to improvement in all domains related to live imaging. 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, how different machineries evolve together and thus coordinate (e.g. see Fig. 1). They help to decipher cell organization and function at different scales through an integrative workflow of methodological and technological developments. New approaches, such as optogenetics may even help controlling cell functions.

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 differentiation and cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, at a time regime compatible with live cell imaging and close to biochemical molecular interactions. The overview of ultrastructural organization will be achieved by complementary electron microscopy methods which have also undergone a revolutionary improvement over the last decade. Image visualization and quantitative analysis are of course essential issues in this context.

4.2. Imaging and analysis of cytoskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells from primary tumor in blood (extravasation) and eventually the colonization of other organs and the formation of secondary tumors.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

- i/ the actin filaments promote cell protrusion, adhesion and retraction;
- ii/ the microtubules are the support of molecule traffic and cell polarization;
- iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix. Another issue is of course to measure and understand how membrane protrusion and retraction keep the cell in homeostasis, which of course relate to membrane traffic.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and how molecular trafficking is coordinated with cytoskeleton changes in these fundamental cellular functions.

SISTM Project-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" [53], "Systems vaccinology" [50], "Systems medicine" [42]. From the data scientist 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. HIV immunotherapies

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 [46]. 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) [41]. 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.

4.3. Translational vaccinology

Vaccines are one of the most efficient tools to prevent and control infectious diseases, and there is a need to increase the number of safe and efficacious vaccines against various pathogens. However, clinical development of vaccines - and of any other investigational product - is a lengthy and costly process. Considering the public health benefits of vaccines, their development needs to be supported and accelerated. During early phase clinical vaccine development (phase I, II trials, translational trials), the number of possible candidate vaccine strategies against a given pathogen that needs to be down-selected in early clinical development is potentially very large. Moreover, during early clinical development there are most often no validated surrogate endpoints to predict the clinical efficacy of a vaccine strategy based on immunogenicity results that could be used as a consensus immunogenicity endpoint and down-selection criterion. This implies considerable uncertainty about the interpretation of immunogenicity results and about the potential value of a vaccine strategy as it transits through early clinical development. Given the complexity of the immune system and the many unknowns in the generation of a protective immune response, early vaccine clinical development nowadays thus takes advantage of high throughput (or "omics") methods allowing to simultaneously assess a

large number of response markers at different levels (“multi-omics”) of the immune system. This has induced a paradigm shift towards early-stage and translational vaccine clinical trials including fewer participants but with thousands of data points collected on every single individual. This is expected to contribute to acceleration of vaccine development thanks to a broader search for immunogenicity signals and a better understanding of the mechanisms induced by each vaccine strategy. However, this remains a difficult research field, both from the immunological as well as from the statistical perspective. Extracting meaningful information from these multi-omics data and transferring it towards an acceleration of vaccine development requires adequate statistical methods, state-of-the art immunological technologies and expertise, and thoughtful interpretation of the results. It thus constitutes research at the interface between disciplines: data science, immunology and vaccinology. Our main current areas of application here are early phase trials of HIV and Ebola vaccine strategies, in which we participate from the initial trial design to the final data analyses.

STEPP Project-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. We 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 is pursuing 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).

4.3.1. 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 has been involved for five years in the ESNET project whose funding came to a close in July of 2017, but the scientific production of the project is still underway. This project bears on the characterization of local Ecosystem Services networks; the project has been 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 is 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 will be useful for the LUCC community at large, independently of the ESNET project needs.

4.3.2. 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

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

⁰Changes in modality choice.

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

retraces the historic path of these theories and models. They 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.

STEEP 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 project CITiES (ANR Modèles Numériques) that ended in 2017 with the PhD defense of Thomas Capelle. This work is being partly pursued in the QAMECS project.

As for LUTI modeling, we have been using the TRANUS model since the creation of our team. 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 Project-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 including 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. Amongst 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. 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.

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 (CPU/GPU). The applications are electro-magnetic simulations for the conception of antennas, 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 modelling 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. A PhD with CEREMA (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) has started in October 2015 (PhD of Pierre Gerhard). The objective is to investigate the model reduction and to implement the resulting acoustic model in our DG solver.
- In September 2017, we started a collaboration with EDF Chatou (PhD of Lucie Quibel) on the modelling of multiphase fluids with complex equations of state. The goal is to simulate the high temperature liquid-vapor flow occurring in a nuclear plant. Among others, we will apply our recent kinetic method for designing efficient implicit schemes for this kind of flows.

XPOP Project-Team

4. Application Domains

4.1. Surface Enhanced Raman Spectroscopy

(joint project with HEGP, AP-HP, and Lip(Sys)2, Université Paris-Saclay)

The objective of this work is to evaluate the feasibility of an evolving technique, surface enhanced Raman spectroscopy (SERS) for the analysis of cytotoxic drug concentration. This technique using silver nanoparticles was applied for quantitative analysis of 5-fluorouracil, one of the most widely used molecules in oncology [8].

In view of the high spectral variability observed between the various repetitions of the experiment, and the observed nonlinear interaction between signal concentration and intensity, nonlinear regression methods that take these variabilities into account have been developed.

4.2. Management of severe trauma

(joint project with the Traumabase group, AP-HP)

Major trauma is defined as any injury that endangers the life or the functional integrity of a person. It has been shown that management of major trauma based on standardized and protocol based care improves prognosis of patients especially for the two main causes of death in major trauma i.e., hemorrhage and traumatic brain injury.

However, evidence shows that patient management even in mature trauma systems often exceeds acceptable time frames, and despite existing guidelines deviations from protocol-based care are often observed. These deviations lead to a high variability in care and are associated with bad outcome such as inadequate hemorrhage control or delayed transfusion. Two main factors explain these observations. First, decision-making in trauma care is particularly demanding, because it requires rapid and complex decisions under time pressure in a very dynamic and multi-player environment characterized by high levels of uncertainty and stress. Second, being a complex and multiplayer process, trauma care is affected by fragmentation. Fragmentation is often the result of loss or deformation of information.

This disruptive influence prevents providers to engage with each other and commit to the care process. In order to respond to this challenge, our program has set the ambitious goal to develop a trauma decision support tool, the TraumaMatrix. The program aims to provide an integrative decision support and information management solution to clinicians for the first 24 hours of major trauma management. This program is divided into three steps.

Based on a detailed and high quality trauma database, Step 1 consists in developing the mathematical tools and models to predict trauma specific outcomes and decisions. This step raises considerable scientific and methodological challenges.

Step 2 will use these methods to apply them to develop in close cooperation with trauma care experts the decision support tool and develop a user friendly and ergonomic interface to be used by clinicians.

Step 3 will further develop the tool and interface and test in real-time its impact on clinician decision making and patient outcome.

4.3. Precision medicine and pharmacogenomics

(joint project with Dassault Systèmes)

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal-regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

4.4. Oncology

(joint project with the Biochemistry lab of Ecole Polytechnique and Institut Curie)

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients.*

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly through a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

4.5. Anesthesiology

(joint project with AP-HP Lariboisière and M3DISIM)

Two hundred million general anaesthetics are performed worldwide every year. Low blood pressure during anaesthesia is common and has been identified as a major factor in morbidity and mortality. These events require great reactivity in order to correct them as quickly as possible and impose constraints of reliability and reactivity to monitoring and treatment.

Recently, studies have demonstrated the usefulness of noradrelanine in preventing and treating intraoperative hypotension. The handling of this drug requires great vigilance with regard to the correct dosage. Currently, these drugs are administered manually by the healthcare staff in bolus and/or continuous infusion. This represents a heavy workload and suffers from a great deal of variability in order to find the right dosage for the desired effect on blood pressure.

The objective of this project is to automate the administration of noradrelanine with a closed-loop system that makes it possible to control the treatment in real time to an instantaneous blood pressure measurement.

4.6. Intracellular processes

(joint project with the InBio and IBIS inria teams and the MSC lab, UMR 7057)

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations [9].

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.
- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are

produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.

- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

4.7. Population pharmacometrics

(joint project with Lixoft)

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage [7], [21], [14].

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

4.8. Mass spectrometry

(joint project with the Molecular Chemistry Laboratory, LCM, of Ecole Polytechnique)

One of the main recent developments in analytical chemistry is the rapid democratization of high-resolution mass spectrometers. These instruments produce extremely complex mass spectra, which can include several hundred thousand ions when analyzing complex samples. The analysis of complex matrices (biological, agri-food, cosmetic, pharmaceutical, environmental, etc.) is precisely one of the major analytical challenges of this new century. Academic and industrial researchers are particularly interested in trying to quickly and effectively establish the chemical consequences of an event on a complex matrix. This may include, for

example, searching for pesticide degradation products and metabolites in fruits and vegetables, photoproducts of active ingredients in a cosmetic emulsion exposed to UV rays or chlorination products of biocides in hospital effluents. The main difficulty of this type of analysis is based on the high spatial and temporal variability of the samples, which is in addition to the experimental uncertainties inherent in any measurement and requires a large number of samples and analyses to be carried out and computerized data processing (up to 16 million per mass spectrum).

A collaboration between XPOP and the Molecular Chemistry Laboratory (LCM) of the Ecole Polytechnique began in 2018. Our objective is to develop new methods for the statistical analysis of mass spectrometry data. These methods are implemented in the SPIX software.

AGORA Project-Team

4. Application Domains

4.1. Smart Cities

One major characteristic of modern societies is that they are prevalently urban. Consequently, the contributions of the Agora team are in particular applied to provide solutions tailored to the emergence of the Internet of Things (IoT) and to Smart Cities applications. A major motivation of the team is the forthcoming explosion of the number of connected devices. In particular, low cost - small data devices are supposed to be densely deployed in our environment, fostering the interest for a convergence of the traditional wireless networking paradigms.

Smart City is a constantly reshaped concept, embracing the future of dense metropolitan areas, with references to efficient and sustainable infrastructure, improving citizens' quality of life and protecting the environment. A consensus on the Smart City philosophy is however that it will be primarily achieved by leveraging a clever integration of Information and Communication Technologies (ICT) in the urban tissue. Indeed, ICTs are enabling an evolution from the current duality between the real world and its digitized counterpart to a continuum in which digital contents and applications are seamlessly interacting with classical infrastructures and services. Smart Cities are often described by the digital services that should be provided which are inherently dependent on dense measurements of the city environment and activities, the collection of these data, their processing into information, and their redistribution. The networking infrastructure plays therefore a critical role in enabling advanced services, in particular the wireless infrastructure supporting density and mobility.

From a wireless networking viewpoint, the digitization of cities can be seen as a paradigm shift extending the Internet of Things (IoT) to a citizen-centric model in order to leverage the massive data collected by pervasive sensors, connected mobiles or fixed devices, and social applications.

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

We focus on methods related to the blend of time reversal techniques and absorbing boundary conditions (ABC) used in a non standard way. 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 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. In [39], we improve this result from a numerical point of view by showing that it can be done numerically without knowing the source. This is done at the expense of not being able to recover the signal in the vicinity of the source. In [40], time dependent wave splitting is performed using ABC and time reversal techniques. We now work on extending these methods to non uniform media.

All our numerical simulations are performed in FreeFem++ which is very flexible. As a byproduct, it enables us to have an end user point of view with respect to FreeFem++ which is very useful for improving it.

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 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 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 on the order of petabyte of data around 2025. 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.

4.5. Molecular simulations

Molecular simulation is one of the most dynamic areas of scientific computing. Its field of application is very broad, ranging from theoretical chemistry and drug design to materials science and nanotechnology. It provides many challenging problems to mathematicians and computer scientists.

In the context of the ERC Synergy Grant EMC2 we address several important limitations of state of the art molecular simulation. In particular, the simulation of very large molecular systems, or smaller systems in which electrons interact strongly with each other, remains out of reach today. In an interdisciplinary collaboration between chemists, mathematicians and computer scientists, we focus on developing a new generation of reliable molecular simulation algorithms and software.

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 brings 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 that we are targeting are presented. In addition to highlighting some application needs, they also constitute some of the use cases that will be used 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 [28] for ocean modelization), code-coupling applications (*e.g.*, the OASIS coupler [33] 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 can 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 volumes of data. For instance, the Vera C. Rubin Observatory (<https://www.vro.org/>) will produce 20 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 (<https://www.euclid-ec.org/>) will generate 1 PB data per year.

AVALON collaborates with the *Institut de Physique des deux Infinis de Lyon* (IP2I) 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 VRO. 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 cooperated 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. AVALON has also started working with the Inria Beagle team (<https://team.inria.fr/beagle/>) on artificial evolution and computational biology as the challenges are around high performance computation and data management.

COAST Project-Team

4. Application Domains

4.1. Crisis Management

Crisis management research investigates all the dimensions regarding the management of unexpected catastrophic events like floods, earthquake or terrorist attacks. All the phases of a crisis, from preparedness to recovery require collaboration between people from many organizations. This provides opportunities to study inter-organizational collaboration at a large scale and to propose and evaluate mechanisms that ensure secure and safe collaboration. The work of Béatrice Linot provides us with a deep understanding of the factors that encourage collaboration and help to maintain trustworthy collaboration between stakeholders. This work is continued by Clélie Amiot who studies the effects of human chat-bot collaboration in this kind of setting.

4.2. Collaborative Editing

Collaborative editing is a common application of optimistic replication in distributed settings. The goal of collaborative editors, irrespective of the kind of document, is to allow a group of users to update a document concurrently while ensuring that they eventually get all the same copy at the end. Our algorithm allows to implement collaborative editor in a peer to peer way. It avoids the need for a central server ensuring a higher level of privacy among collaborators. In this context, it requires to consider the problem of authentication and authorization of participants[4] and of trust between them[7].

COATI Project-Team

4. Application Domains

4.1. Telecommunication Networks

COATI is mostly interested in telecommunications networks but also in the network structure appearing in social, molecular and transportation networks.

We focus on the design and management of heterogeneous physical and logical networks. The project has kept working on the design of backbone networks (optical networks, radio networks, IP networks). However, the fields of Software Defined Networks and Network Function Virtualization are growing in importance in our studies. In all these networks, we study routing algorithms 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 collaborate with experts of some of these domains.

For instance, we collaborate with project-team ABS (Algorithms Biology Structure) from Sophia Antipolis on problems from Structural Biology (co-supervision of a PhD student). In the area of transportation networks, we collaborate with SMEs Benomad and Instant-System on dynamic car-pooling combined with multi-modal transportation systems in the context of ANR project Multimod started in January 2018. Last, we have started a collaboration with GREDEG (Groupe de Recherche en Droit, Economie et Gestion, Univ. Nice Sophia Antipolis) on the analysis of collaboration networks.

CTRL-A Project-Team

4. Application Domains

4.1. Self-adaptive and reconfigurable computing systems in HPC and the IoT

We are attacking the problem of designing well-regulated and efficient self-adaptive computing systems by the development of novel strategies for systems management.

The kind of systems we typically target involve relatively coarse grained computation tasks (e.g. image processing or HPC tasks, components or services), assembled in workflows, application dependency graphs, or composites. At that level, there can be parallel and conditional branches, as well as choices that can be made between alternative branches, corresponding to different ways to perform that part of the application. Such tasks can be achieved following a choice of implementations or versions, such as in service oriented approaches. Each implementation has its own characteristics and requirements, e.g., w.r.t. resources consumed and QoS offered. The systems execution infrastructures present heterogeneity, with different computing processors, a variety of peripheral devices (e.g., I/O, video port, accelerators), and different means of communications. This hardware or middleware level also presents adaptation potential e.g., in varying quantities of resources or sleep and stand-by modes.

The kinds of control problems encountered in these adaptive systems concern the navigation in the configurations space defined by choice points at the levels of applications, tasks, and architecture. Upon events or conditions triggering reconfiguration and adaptation, the controller has to choose a next configuration where, on the one hand, all consistency constraints are satisfied w.r.t. dependencies and resources requirements. On the other hand, it has to apply a policy or strategy deciding between correct configurations which one to chose e.g. by optimizing one or more criteria, or by insuring reachability of some later configuration (goal or fallback). This targeted class of computing systems we consider is mid-sized, in the sense that the combinatorial complexity is large enough for manual solving to be impossible, while remaining within the range where supervisory control techniques are tractable. The pace of control is more sporadic, and slower than the instruction-level computation performance within the coarse-grained tasks.

The objectives of CTRL-A will be achieved and evaluated in both of our main application domains, thereby exhibiting their similarities from the point of view of reconfiguration control. A first application domain is High Performance Computing. In this area, we currently focus especially on the management of Dynamic Partial Reconfiguration in FPGA based hardware, at the level of middleware. Here the particular application we consider is, as in our ANR project HPeC starting end of 2015, video image flow processing for smart cameras implemented on DPR FPGAs themselves embedded in drones.

A second general application domain to confront our models is the Internet of Things (IoT), more specifically self-adaptive middleware platforms for Smart Environments, or Industry 4.0 related topics, like SCADA. We focus on providing coordination components and controllers of software components and services, or rule-based middleware platforms. The adaptation problems concern both the functional aspects of the applications in a smart building, and the middleware support deployment and reconfiguration issues. We are considering perspectives concerning self-protection and security.

DANTE Project-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 modelling 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 modelling approaches crucially depend on our ability to describe the interactions of individuals in the population.

Within PhD work of G. Frusque, we collaborate with Dr. Julien Jung from Hôpital de Neurologie de Bron (HCL) and with Nadine Ravel, DR CNRS (CRNL, INSERM).

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 characterise the topology and temporal behaviour 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 behaviour of very complicated systems.

In this endeavour 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 anonymised 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 (anonymised) 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.

4.3. Social Sciences / Epistemology

Political impacts of the internet and of networks begin to be well known (Cambridge Analytica, Russian trolls, etc.). Hence the public at large begins to be aware of the abuses of the leaders of the internet (privacy by firms and advertising, surveillance by states, fake news by activists, etc.). In the same time, on-line exchanges now give scientific estimations of the political life [60] and political sciences begin to consider the internet as a relevant subject of study. As the internet is a *technology*, philosophy is the best approach to understand what socially happens (or can be made) with the internet. We develop it by two ways:

- political philosophy. Some Dante members are working with Triangle laboratory (social sciences, philosophy and politics; Ens de Lyon and CNRS).
- epistemology, because computer sciences discoveries are related with the evolution of science(s): we discovered that, in the case of the internet, political philosophy can do few if not strongly related to epistemology [61]. Epistemological approach is developed in collaboration with Jean Dhombres (who holds a seminar at Enssib: <http://barthes.enssib.fr/cours/Dhombres2018-2019.html> and with Hcéres (new relations between social and exact sciences, cf. point 9.2.8).

This approach should help computer scientists to understand how their research may depend on foreign initiatives and to create new links between social sciences and Inria.

DATAMOVE Project-Team

4. Application Domains

4.1. Data Aware Batch Scheduling

Large scale high performance computing platforms are becoming increasingly complex. Determining efficient allocation and scheduling strategies that can adapt to technological evolutions is a strategic and difficult challenge. We are interested in scheduling jobs in hierarchical and heterogeneous large scale platforms. On such platforms, application developers typically submit their jobs in centralized waiting queues. The job management system aims at determining a suitable allocation for the jobs, which all compete against each other for the available computing resources. Performances are measured using different classical metrics like maximum completion time or slowdown. Current systems make use of very simple (but fast) algorithms that however rely on simplistic platform and execution models, and thus, have limited performances.

For all target scheduling problems we aim to provide both theoretical analysis and complementary analysis through simulations. Achieving meaningful results will require strong improvements on existing models (on power for example) and the design of new approximation algorithms with various objectives such as stretch, reliability, throughput or energy consumption, while keeping in focus the need for a low-degree polynomial complexity.

4.1.1. Algorithms

The most common batch scheduling policy is to consider the jobs according to the First Come First Served order (FCFS) with backfilling (BF). BF is the most widely used policy due to its easy and robust implementation and known benefits such as high system utilization. It is well-known that this strategy does not optimize any sophisticated function, but it is simple to implement and it guarantees that there is no starvation (i.e. every job will be scheduled at some moment).

More advanced algorithms are seldom used on production platforms due to both the gap between theoretical models and practical systems and speed constraints. When looking at theoretical scheduling problems, the generally accepted goal is to provide polynomial algorithms (in the number of submitted jobs and the number of involved computing units). However, with millions of processing cores where every process and data transfer have to be individually scheduled, polynomial algorithms are prohibitive as soon as the polynomial degree is too large. The model of *parallel tasks* simplifies this problem by bundling many threads and communications into single boxes, either rigid, rectangular or malleable. Especially malleable tasks capture the dynamicity of the execution. Yet these models are ill-adapted to heterogeneous platforms, as the running time depends on more than simply the number of allotted resources, and some of the common underlying assumptions on the speed-up functions (such as monotony or concavity) are most often only partially verified.

In practice, the job execution times depend on their allocation (due to communication interferences and heterogeneity in both computation and communication), while theoretical models of parallel jobs usually consider jobs as black boxes with a fixed (maximum) execution time. Though interesting and powerful, the classical models (namely, synchronous PRAM model, delay, LogP) and their variants (such as hierarchical delay), are not well-suited to large scale parallelism on platforms where the cost of moving data is significant, non uniform and may change over time. Recent studies are still refining such models in order to take into account communication contentions more accurately while remaining tractable enough to provide a useful tool for algorithm design.

Today, all algorithms in use in production systems are oblivious to communications. One of our main goals is to **design a new generation of scheduling algorithms fitting more closely job schedules according to platform topologies.**

4.1.2. Locality Aware Allocations

Recently, we developed modifications of the standard back-filling algorithm taking into account platform topologies. The proposed algorithms take into account locality and contiguity in order to hide communication patterns within parallel tasks. The main result here is to establish good lower bounds and small approximation ratios for policies respecting the locality constraints. The algorithms work in an online fashion, improving the global behavior of the system while still keeping a low running time. These improvements rely mainly on our past experience in designing approximation algorithms. Instead of relying on complex networking models and communication patterns for estimating execution times, the communications are disconnected from the execution time. Then, the scheduling problem leads to a trade-off: optimizing locality of communications on one side and a performance objective (like the makespan or stretch) on the other side.

In the perspective of taking care of locality, other ongoing works include the study of schedulers for platforms whose interconnection network is a static structured topology (like the 3D-torus of the BlueWaters platform we work on in collaboration with the Argonne National Laboratory). One main characteristic of this 3D-torus platform is to provide I/O nodes at specific locations in the topology. Applications generate and access specific data and are thus bounded to specific I/O nodes. Resource allocations are constrained in a strong and unusual way. This problem is close for actual hierarchical platforms. The scheduler needs to compute a schedule such that I/O nodes requirements are filled for each application while at the same time avoiding communication interferences. Moreover, extra constraints can arise for applications requiring accelerators that are gathered on the nodes at the edge of the network topology.

While current results are encouraging, they are however limited in performance by the low amount of information available to the scheduler. We look forward to extend ongoing work by progressively increasing application and network knowledge (by technical mechanisms like profiling or monitoring or by more sophisticated methods like learning). It is also important to anticipate on application resource usage in terms of compute units, memory as well as network and I/Os to efficiently schedule a mix of applications with different profiles. For instance, a simple solution is to partition the jobs as "communication intensive" or "low communications". Such a tag could be achieved by the users themselves or obtained by learning techniques. We could then schedule low communications jobs using leftover spaces while taking care of high communication jobs. More sophisticated options are possible, for instance those that use more detailed communication patterns and networking models. Such options would leverage the work proposed in Section 4.2 for gathering application traces.

4.1.3. Data-Centric Processing

Exascale computing is shifting away from the traditional compute-centric models to a more data-centric one. This is driven by the evolving nature of large scale distributed computing, no longer dominated by pure computations but also by the need to handle and analyze large volumes of data. These data can be large databases of results, data streamed from a running application or another scientific instrument (collider for instance). These new workloads call for specific resource allocation strategies.

Data movements and storage are expected to be a major energy and performance bottleneck on next generation platforms. Storage architectures are also evolving, the standard centralized parallel file system being complemented with local persistent storage (Burst Buffers, NVRAM). Thus, one data producer can stage data on some nodes' local storage, requiring to schedule close by the associated analytics tasks to limit data movements. This kind of configuration, often referred as *in-situ analytics*, is expected to become common as it enables to switch from the traditional I/O intensive workflow (batch-processing followed by *post mortem* analysis and visualization) to a more storage conscious approach where data are processed as closely as possible to where and when they are produced (in-situ processing is addressed in details in section 4.3). By reducing data movements and scheduling the extra processing on resources not fully exploited yet, in-situ processing is expected to have also a significant positive energetic impact. Analytics codes can be executed in the same nodes than the application, often on dedicated cores commonly called helper cores, or on dedicated nodes called staging nodes. The results are either forwarded to the users for visualization or saved to disk through I/O nodes. In-situ analytics can also take benefit of node local disks or burst buffers to reduce data movements.

Future job scheduling strategies should take into account in-situ processes in addition to the job allocation to optimize both energy consumption and execution time. On the one hand, this problem can be reduced to an allocation problem of extra asynchronous tasks to idle computing units. But on the other hand, embedding analytics in applications brings extra difficulties by making the application more heterogeneous and imposing more constraints (data affinity) on the required resources. Thus, the main point here is to develop efficient algorithms for dealing with heterogeneity without increasing the global computational cost.

4.1.4. Learning

Another important issue is to adapt the job management system to deal with the bad effects of uncertainties, which may be catastrophic in large scale heterogeneous HPC platforms (jobs delayed arbitrarily far or jobs killed). A natural question is then: *is it possible to have a good estimation of the job and platform parameters in order to be able to obtain a better scheduling ?* Many important parameters (like the number or type of required resources or the estimated running time of the jobs) are asked to the users when they submit their jobs. However, some of these values are not accurate and in many cases, they are not even provided by the end-users. In DataMove, we propose to study new methods for a better prediction of the characteristics of the jobs and their execution in order to improve the optimization process. In particular, the methods well-studied in the field of big data (in supervised Machine Learning, like classical regression methods, Support Vector Methods, random forests, learning to rank techniques or deep learning) could and must be used to improve job scheduling in large scale HPC platforms. This topic received a great attention recently in the field of parallel and distributed processing. A preliminary study has been done recently by our team with the target of predicting the job running times (called wall times). We succeeded to improve significantly in average the reference EASY Back Filling algorithm by estimating the wall time of the jobs, however, this method leads to big delay for the stretch of few jobs. Even if we succeed in determining more precisely hidden parameters, like the wall time of the jobs, this is not enough to determine an optimized solution. The shift is not only to learn on dedicated parameters but also on the scheduling policy. The data collected from the accounting and profiling of jobs can be used to better understand the needs of the jobs and through learning to propose adaptations for future submissions. The goal is to propose extensions to further improve the job scheduling and improve the performance and energy efficiency of the application. For instance preference learning may enable to compute on-line new priorities to back-fill the ready jobs.

4.1.5. Multi-objective Optimization

Several optimization questions that arise in allocation and scheduling problems lead to the study of several objectives at the same time. The goal is then not a single optimal solution, but a more complicated mathematical object that captures the notion of trade-off. In broader terms, the goal of multi-objective optimization is not to externally arbitrate on disputes between entities with different goals, but rather to explore the possible solutions to highlight the whole range of interesting compromises. A classical tool for studying such multi-objective optimization problems is to use *Pareto curves*. However, the full description of the Pareto curve can be very hard because of both the number of solutions and the hardness of computing each point. Addressing this problem will opens new methodologies for the analysis of algorithms.

To further illustrate this point here are three possible case studies with emphasis on conflicting interests measured with different objectives. While these cases are good representatives of our HPC context, there are other pertinent trade-offs we may investigate depending on the technology evolution in the coming years. This enumeration is certainly not limitative.

Energy versus Performance. The classical scheduling algorithms designed for the purpose of performance can no longer be used because performance and energy are contradictory objectives to some extent. The scheduling problem with energy becomes a multi-objective problem in nature since the energy consumption should be considered as equally important as performance at exascale. A global constraint on energy could be a first idea for determining trade-offs but the knowledge of the Pareto set (or an approximation of it) is also very useful.

Administrators versus application developers. Both are naturally interested in different objectives: In current algorithms, the performance is mainly computed from the point of view of administrators, but the users should be in the loop since they can give useful information and help to the construction of better schedules. Hence, we face again a multi-objective problem where, as in the above case, the approximation of the Pareto set provides the trade-off between the administrator view and user demands. Moreover, the objectives are usually of the same nature. For example, *max stretch* and *average stretch* are two objectives based on the slowdown factor that can interest administrators and users, respectively. In this case the study of the norm of stretch can be also used to describe the trade-off (recall that the L_1 -norm corresponds to the average objective while the L_∞ -norm to the max objective). Ideally, we would like to design an algorithm that gives good approximate solutions at the same time for all norms. The L_2 or L_3 -norm are useful since they describe the performance of the whole schedule from the administrator point of view as well as they provide a fairness indication to the users. The hard point here is to derive theoretical analysis for such complicated tools.

Resource Augmentation. The classical resource augmentation models, i.e. speed and machine augmentation, are not sufficient to get good results when the execution of jobs cannot be frequently interrupted. However, based on a resource augmentation model recently introduced, where the algorithm may reject a small number of jobs, some members of our team have given the first interesting results in the non-preemptive direction. In general, resource augmentation can explain the intuitive good behavior of some greedy algorithms while, more interestingly, it can give ideas for new algorithms. For example, in the rejection context we could dedicate a small number of nodes for the usually problematic rejected jobs. Some initial experiments show that this can lead to a schedule for the remaining jobs that is very close to the optimal one.

4.2. Empirical Studies of Large Scale Platforms

Experiments or realistic simulations are required to take into account the impact of allocations and assess the real behavior of scheduling algorithms. While theoretical models still have their interest to lay the groundwork for algorithmic designs, the models are necessarily reflecting a purified view of the reality. As transferring our algorithm in a more practical setting is an important part of our creed, we need to ensure that the theoretical results found using simplified models can really be transposed to real situations. On the way to exascale computing, large scale systems become harder to study, to develop or to calibrate because of the costs in both time and energy of such processes. It is often impossible to convince managers to use a production cluster for several hours simply to test modifications in the RJMS. Moreover, as the existing RJMS production systems need to be highly reliable, each evolution requires several real scale test iterations. The consequence is that scheduling algorithms used in production systems are mostly outdated and not customized correctly. To circumvent this pitfall, we need to develop tools and methodologies for alternative empirical studies, from analysis of workload traces, to job models, simulation and emulation with reproducibility concerns.

4.2.1. Workload Traces with Resource Consumption

Workload traces are the base element to capture the behavior of complete systems composed of submitted jobs, running applications, and operating tools. These traces must be obtained on production platforms to provide relevant and representative data. To get a better understanding of the use of such systems, we need to look at both, how the jobs interact with the job management system, and how they use the allocated resources. We propose a general workload trace format that adds jobs resource consumption to the commonly used SWF⁰ workload trace format. This requires to instrument the platforms, in particular to trace resource consumptions like CPU, data movements at memory, network and I/O levels, with an acceptable performance impact. In a previous work we studied and proposed a dedicated job monitoring tool whose impact on the system has been measured as lightweight (0.35% speed-down) with a 1 minute sampling rate. Other tools also explore job monitoring, like TACC Stats. A unique feature from our tool is its ability to monitor distinctly jobs sharing common nodes.

⁰Standard Workload Format: <http://www.cs.huji.ac.il/labs/parallel/workload/swf.html>

Collected workload traces with jobs resource consumption will be publicly released and serve to provide data for works presented in Section 4.1. The trace analysis is expected to give valuable insights to define models encompassing complex behaviours like network topology sensitivity, network congestion and resource interferences.

We expect to join efforts with partners for collecting quality traces (ATOS/Bull, Ciment meso center, Joint Laboratory on Extreme Scale Computing) and will collaborate with the Inria team POLARIS for their analysis.

4.2.2. Simulation

Simulations of large scale systems are faster by multiple orders of magnitude than real experiments. Unfortunately, replacing experiments with simulations is not as easy as it may sound, as it brings a host of new problems to address in order to ensure that the simulations are closely approximating the execution of typical workloads on real production clusters. Most of these problems are actually not directly related to scheduling algorithms assessment, in the sense that the workload and platform models should be defined independently from the algorithm evaluations, in order to ensure a fair assessment of the algorithms' strengths and weaknesses. These research topics (namely platform modeling, job models and simulator calibration) are addressed in the other subsections.

We developed an open source platform simulator within DataMove (in conjunction with the OAR development team) to provide a widely distributable test bed for reproducible scheduling algorithm evaluation. Our simulator, named Batsim, allows to simulate the behavior of a computational platform executing a workload scheduled by any given scheduling algorithm. To obtain sound simulation results and to broaden the scope of the experiments that can be done thanks to Batsim, we did not chose to create a (necessarily limited) simulator from scratch, but instead to build on top of the SimGrid simulation framework.

To be open to as many batch schedulers as possible, Batsim decouples the platform simulation and the scheduling decisions in two clearly-separated software components communicating through a complete and documented protocol. The Batsim component is in charge of simulating the computational resources behaviour whereas the scheduler component is in charge of taking scheduling decisions. The scheduler component may be both a resource and a job management system. For jobs, scheduling decisions can be to execute a job, to delay its execution or simply to reject it. For resources, other decisions can be taken, for example to change the power state of a machine i.e. to change its speed (in order to lower its energy consumption) or to switch it on or off. This separation of concerns also enables interfacing with potentially any commercial RJMS, as long as the communication protocol with Batsim is implemented. A proof of concept is already available with the OAR RJMS.

Using this test bed opens new research perspectives. It allows to test a large range of platforms and workloads to better understand the real behavior of our algorithms in a production setting. In turn, this opens the possibility to tailor algorithms for a particular platform or application, and to precisely identify the possible shortcomings of the theoretical models used.

4.2.3. Job and Platform Models

The central purpose of the Batsim simulator is to simulate job behaviors on a given target platform under a given resource allocation policy. Depending on the workload, a significant number of jobs are parallel applications with communications and file system accesses. It is not conceivable to simulate individually all these operations for each job on large platforms with their associated workload due to implied simulation complexity. The challenge is to define a coarse grain job model accurate enough to reproduce parallel application behavior according to the target platform characteristics. We will explore models similar to the BSP (Bulk Synchronous Program) approach that decomposes an application in local computation supersteps ended by global communications and a global synchronization. The model parameters will be established by means of trace analysis as discussed previously, but also by instrumenting some parallel applications to capture communication patterns. This instrumentation will have a significant impact on the concerned application performance, restricting its use to a few applications only. There are a lot of recurrent applications executed on HPC platform, this fact will help to reduce the required number of instrumentations and captures. To assign

each job a model, we are considering to adapt the concept of application signatures as proposed in. Platform models and their calibration are also required. Large parts of these models, like those related to network, are provided by Simgrid. Other parts as the filesystem and energy models are comparatively recent and will need to be enhanced or reworked to reflect the HPC platform evolutions. These models are then generally calibrated by running suitable benchmarks.

4.2.4. Emulation and Reproducibility

The use of coarse models in simulation implies to set aside some details. This simplification may hide system behaviors that could impact significantly and negatively the metrics we try to enhance. This issue is particularly relevant when large scale platforms are considered due to the impossibility to run tests at nominal scale on these real platforms. A common approach to circumvent this issue is the use of emulation techniques to reproduce, under certain conditions, the behavior of large platforms on smaller ones. Emulation represents a natural complement to simulation by allowing to execute directly large parts of the actual evaluated software and system, but at the price of larger compute times and a need for more resources. The emulation approach was chosen in to compare two job management systems from workload traces of the CURIE supercomputer (80000 cores). The challenge is to design methods and tools to emulate with sufficient accuracy the platform and the workload (data movement, I/O transfers, communication, applications interference). We will also intend to leverage emulation tools like Distem from the MADYNES team. It is also important to note that the Batsim simulator also uses emulation techniques to support the core scheduling module from actual RJMS. But the integration level is not the same when considering emulation for larger parts of the system (RJMS, compute node, network and filesystem).

Replaying traces implies to prepare and manage complex software stacks including the OS, the resource management system, the distributed filesystem and the applications as well as the tools required to conduct experiments. Preparing these stacks generate specific issues, one of the major one being the support for reproducibility. We propose to further develop the concept of reconstructability to improve experiment reproducibility by capturing the build process of the complete software stack. This approach ensures reproducibility over time better than other ways by keeping all data (original packages, build recipe and Kameleon engine) needed to build the software stack.

In this context, the Grid'5000 (see Sec. 6.4) experimentation infrastructure that gives users the control on the complete software stack is a crucial tool for our research goals. We will pursue our strong implication in this infrastructure.

4.3. Integration of High Performance Computing and Data Analytics

Data produced by large simulations are traditionally handled by an I/O layer that moves them from the compute cores to the file system. Analysis of these data are performed after reading them back from files, using some domain specific codes or some scientific visualisation libraries like VTK. But writing and then reading back these data generates a lot of data movements and puts under pressure the file system. To reduce these data movements, **the in situ analytics paradigm proposes to process the data as closely as possible to where and when the data are produced**. Some early solutions emerged either as extensions of visualisation tools or of I/O libraries like ADIOS. But significant progresses are still required to provide efficient and flexible high performance scientific data analysis tools. Integrating data analytics in the HPC context will have an impact on resource allocation strategies, analysis algorithms, data storage and access, as well as computer architectures and software infrastructures. But this paradigm shift imposed by the machine performance also sets the basis for a deep change on the way users work with numerical simulations. The traditional workflow needs to be reinvented to make HPC more user-centric, more interactive and turn HPC into a commodity tool for scientific discovery and engineering developments. In this context DataMove aims at investigating programming environments for in situ analytics with a specific focus on task scheduling in particular, to ensure an efficient sharing of resources with the simulation.

4.3.1. Programming Model and Software Architecture

In situ creates a tighter loop between the scientist and her/his simulation. As such, an in situ framework needs to be flexible to let the user define and deploy its own set of analysis. A manageable flexibility requires to favor simplicity and understandability, while still enabling an efficient use of parallel resources. Visualization libraries like VTK or Visit, as well as domain specific environments like VMD have initially been developed for traditional post-mortem data analysis. They have been extended to support in situ processing with some simple resource allocation strategies but the level of performance, flexibility and ease of use that is expected requires to rethink new environments. There is a need to develop a middleware and programming environment taking into account in its foundations this specific context of high performance scientific analytics.

Similar needs for new data processing architectures occurred for the emerging area of Big Data Analytics, mainly targeted to web data on cloud-based infrastructures. Google Map/Reduce and its successors like Spark or Stratosphere/Flink have been designed to match the specific context of efficient analytics for large volumes of data produced on the web, on social networks, or generated by business applications. These systems have mainly been developed for cloud infrastructures based on commodity architectures. They do not leverage the specifics of HPC infrastructures. Some preliminary adaptations have been proposed for handling scientific data in a HPC context. However, these approaches do not support in situ processing.

Following the initial development of FlowVR, our middleware for in situ processing, we will pursue our effort to develop a programming environment and software architecture for high performance scientific data analytics. Like FlowVR, the map/reduce tools, as well as the machine learning frameworks like TensorFlow, adopted a dataflow graph for expressing analytics pipe-lines. We are convinced that this dataflow approach is both easy to understand and yet expresses enough concurrency to enable efficient executions. The graph description can be compiled towards lower level representations, a mechanism that is intensively used by Stratosphere/Flink for instance. Existing in situ frameworks, including FlowVR, inherit from the HPC way of programming with a thinner software stack and a programming model close to the machine. Though this approach enables to program high performance applications, this is usually too low level to enable the scientist to write its analysis pipe-line in a short amount of time. The data model, i.e. the data semantics level accessible at the framework level for error check and optimizations, is also a fundamental aspect of such environments. The key/value store has been adopted by all map/reduce tools. Except in some situations, it cannot be adopted as such for scientific data. Results from numerical simulations are often more structured than web data, associated with acceleration data structures to be processed efficiently. We will investigate data models for scientific data building on existing approaches like Adios or DataSpaces.

4.3.2. Resource Sharing

To alleviate the I/O bottleneck, the in situ paradigm proposes to start processing data as soon as made available by the simulation, while still residing in the memory of the compute node. In situ processings include data compression, indexing, computation of various types of descriptors (1D, 2D, images, etc.). Per se, reducing data output to limit I/O related performance drops or keep the output data size manageable is not new. Scientists have relied on solutions as simple as decreasing the frequency of result savings. In situ processing proposes to move one step further, by providing a full fledged processing framework enabling scientists to more easily and thoroughly manage the available I/O budget.

The most direct way to perform in situ analytics is to inline computations directly in the simulation code. In this case, in situ processing is executed in sequence with the simulation that is suspended meanwhile. Though this approach is direct to implement and does not require complex framework environments, it does not enable to overlap analytics related computations and data movements with the simulation execution, preventing to efficiently use the available resources. Instead of relying on this simple time sharing approach, several works propose to rely on space sharing where one or several cores per node, called *helper cores*, are dedicated to analytics. The simulation responsibility is simply to handle a copy of the relevant data to the node-local in situ processes, both codes being executed concurrently. This approach often lead to significantly better performance than in-simulation analytics.

For a better isolation of the simulation and in situ processes, one solution consists in offloading in situ tasks from the simulation nodes towards extra dedicated nodes, usually called *staging nodes*. These computations are said to be performed *in-transit*. But this approach may not always be beneficial compared to processing on simulation nodes due to the costs of moving the data from the simulation nodes to the staging nodes.

FlowVR enables to mix these different resources allocation strategies for the different stages of an analytics pipeline. Based on a component model, the scientist designs analytics workflows by first developing processing components that are next assembled in a dataflow graph through a Python script. At runtime the graph is instantiated according to the execution context, FlowVR taking care of deploying the application on the target architecture, and of coordinating the analytics workflows with the simulation execution.

But today the choice of the resource allocation strategy is mostly ad-hoc and defined by the programmer. We will investigate solutions that enable a cooperative use of the resource between the analytics and the simulation with minimal hints from the programmer. In situ processings inherit from the parallelization scale and data distribution adopted by the simulation, and must execute with minimal perturbations on the simulation execution (whose actual resource usage is difficult to know a priori). We need to develop adapted scheduling strategies that operate at compile and run time. Because analysis are often data intensive, such solutions must take into consideration data movements, a point that classical scheduling strategies designed first for compute intensive applications often overlook. We expect to develop new scheduling strategies relying on the methodologies developed in Sec. 4.1.5 . Simulations as well as analysis are iterative processes exposing a strong spatial and temporal coherency that we can take benefit of to anticipate their behavior and then take more relevant resources allocation strategies, possibly based on advanced learning algorithms or as developed in Section 4.1 .

In situ analytics represent a specific workload that needs to be scheduled very closely to the simulation, but not necessarily active during the full extent of the simulation execution and that may also require to access data from previous runs (stored in the file system or on specific burst-buffers). Several users may also need to run concurrent analytics pipe-lines on shared data. This departs significantly from the traditional batch scheduling model, motivating the need for a more elastic approach to resource provisioning. These issues will be conjointly addressed with research on batch scheduling policies (Sec. 4.1).

4.3.3. Co-Design with Data Scientists

Given the importance of users in this context, it is of primary importance that in situ tools be co-designed with advanced users, even if such multidisciplinary collaborations are challenging and require constant long term investments to learn and understand the specific practices and expectations of the other domain.

We will tightly collaborate with scientists of some application domains, like molecular dynamics or fluid simulation, to design, develop, deploy and assess in situ analytics scenarios, as already done with Marc Baaden, a computational biologist from LBT.

DELYS Project-Team (section vide)

DIANA Project-Team (section vide)

DIONYSOS Project-Team

4. Application Domains

4.1. Networking

Our global research effort concerns networking problems, both from the analysis point of view, and around network design issues. Specifically, this means the IP technology in general, with focus on specific types of networks seen at different levels: wireless systems, optical infrastructures, peer-to-peer architectures, Software Defined Networks, Content Delivery Networks, Content-Centric Networks, clouds.

A specific aspect of network applications and/or services based on video or voice content, is our PSQA technology, able to measure the Perceptual Quality automatically and in real time. PSQA provides a MOS value as close as it makes sense to the value obtained from subjective testing sessions. The technology has been tested in many environments, including one way communications as, for instance, in video streaming, and bi-directional communications as in IP telephony, UDP- or TCP-based systems, etc. It has already served in many collaborative projects as the measuring tool used.

4.2. Stochastic modeling

Many of the techniques developed at Dionysos are related to the analysis of complex systems in general, not only in telecommunications. For instance, our Monte Carlo methods for analyzing rare events have been used by different industrial partners, some of them in networking but recently also by companies building transportation systems. We develop methods in different areas: numerical analysis of stochastic models, bound computations in the same area, Discrete Event Simulation, or, as just mentioned, rare event analysis.

DIVERSE Project-Team (section vide)

DYOGENE Project-Team

4. Application Domains

4.1. Physical communication networks

Internet, wireless, mobile, cellular networks, transportation networks, distributed systems (cloud, call centers).
In collaboration with Nokia Bell Labs and Orange Labs.

4.2. Abstract networks

Social interactions, human communities, economic networks.

4.3. Power grids

Energy networks. In collaboration with EDF.

EASE Project-Team

4. Application Domains

4.1. Pervasive applications in Smart Building

A Smart Building is a living space 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 Pervasive Computing paradigm in the Smart Building 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*. EASE 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 complicated processing compared to the current solutions. The result is a smart environment 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 smarter environment, showing a solution for improving both comfort and energy savings.

4.2. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attracts number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organize the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future autonomous vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicles could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle, it is still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc.). From augmented perception to fully cooperative automated vehicle, the autonomy covers various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. That helps the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take advantages of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly relies on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an example of smart city automated processes that will have to share information within the surrounding to make their decisions.

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 that 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 principles 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 a 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 a large number of objects of various nature, arranged randomly in a given area or container. **Most pervasive computing applications fall in this context.**

EVA Project-Team

4. Application Domains

4.1. Industrial Process Automation

Wireless networks have become ubiquitous and are an integral part of our daily lives. These networks are present in many application domains; the most important are detailed in this section.

Networks in industrial process automation typically perform **monitoring and control** tasks. Wired industrial communication networks, such as HART⁰, have been around for decades and, being wired, are highly reliable. Network administrators tempted to “go wireless” expect the same reliability. Reliable process automation networks – especially when used for control – often impose stringent latency requirements. Deterministic wireless networks can be used in critical systems such as control loops, however, the unreliable nature of the wireless medium, coupled with their large scale and “ad-hoc” nature raise some of the most important challenges for low-power wireless research over the next 5-10 years.

Through the involvement of team members in standardization activities, protocols and techniques are proposed for the standardization process with a view to becoming the *de-facto* standard for wireless industrial process automation. Besides producing top level research publications and standardization activities, EVA intends this activity to foster further collaborations with industrial partners.

4.2. Environmental Monitoring

Today, outdoor WSNs are used to monitor vast rural or semi-rural areas and may be used to detect fires. Another example is detecting fires in outdoor fuel depots, where the delivery of alarm messages to a monitoring station in an upper-bounded time is of prime importance. Other applications consist in monitoring the snow melting process in mountains, tracking the quality of water in cities, registering the height of water in pipes to foresee flooding, etc. These applications lead to a vast number of technical issues: deployment strategies to ensure suitable coverage and good network connectivity, energy efficiency, reliability and latency, etc.

We work on such applications in an associate team “REALMS” comprising members from EVA, the university of Berkeley and the university of Michigan.

4.3. The Internet of Things

The general agreement is that the Internet of Things (IoT) is composed of small, often battery-powered objects which measure and interact with the physical world, and encompasses smart home applications, wearables, smart city and smart plant applications.

It is absolutely essential to (1) clearly understand the limits and capabilities of the IoT, and (2) develop technologies which enable user expectation to be met.

The EVA team is dedicated to understanding and contributing to the IoT. In particular, the team maintains a good understanding of the different technologies at play (Bluetooth, IEEE 802.15.4, WiFi, cellular), and their trade-offs. Through scientific publications and other contributions, EVA helps establish which technology best fits which application.

⁰Highway Addressable Remote Transducer

4.4. Military, Energy and Aerospace

Through the HIPERCOM project, EVA has developed cutting-edge expertise in using wireless networks for military, energy and aerospace applications. Wireless networks are a key enabling technology in the application domains, as they allow physical processes to be instrumented (e.g. the structural health of an airplane) at a granularity not achievable by its wired counterpart. Using wireless technology in these domains does however raise many technical challenges, including end-to-end latency, energy-efficiency, reliability and Quality of Service (QoS). Mobility is often an additional constraint in energy and military applications. Achieving scalability is of paramount importance for tactical military networks, and, albeit to a lesser degree, for power plants. EVA will work in this domain.

Smart cities share the constraint of mobility (both pedestrian and vehicular) with tactical military networks. Vehicular Ad-hoc NETWORKS (VANETs) will play an important role in the development of smarter cities.

The coexistence of different networks operating in the same radio spectrum can cause interference that should be avoided. Cognitive radio provides secondary users with the frequency channels that are temporarily unused (or unassigned) by primary users. Such opportunistic behavior can also be applied to urban wireless sensor networks. Smart cities raise the problem of transmitting, gathering, processing and storing big data. Another issue is to provide the right information at the place where it is most needed.

4.5. Emergency Applications

In an “emergency” application, heterogeneous nodes of a wireless network cooperate to recover from a disruptive event in a timely fashion, thereby possibly saving human lives. These wireless networks can be rapidly deployed and are useful to assess damage and take initial decisions. Their primary goal is to maintain connectivity with the humans or mobile robots (possibly in a hostile environment) in charge of network deployment. The deployment should ensure the coverage of particular points or areas of interest. The wireless network has to cope with pedestrian mobility and robot/vehicle mobility. The environment, initially unknown, is progressively discovered and may contain numerous obstacles that should be avoided. The nodes of the wireless network are usually battery-powered. Since they are placed by a robot or a human, their weight is very limited. The protocols supported by these nodes should be energy-efficient to maximize network lifetime. In such a challenging environment, sensor nodes should be replaced before their batteries are depleted. It is therefore important to be able to accurately determine the battery lifetime of these nodes, enabling predictive maintenance.

4.6. Types of Wireless Networks

The EVA team will distinguish between opportunistic communication (which takes advantage of a favorable state) and collaborative communication (several entities collaborate to reach a common objective). Furthermore, determinism can be required to schedule medium access and node activity, and to predict energy consumption.

In the EVA project, we will propose **self-adaptive wireless networks** whose evolution is based on:

- optimization to minimize a single or multiple objective functions under some constraints (e.g. interference, or energy consumption in the routing process).
- machine learning to be able to predict a future state based on past states (e.g. link quality in a wireless sensor network) and to identify tendencies.

The types of wireless networks encountered in the application domains can be classified in the following categories.

4.6.1. Wireless Sensor and Mesh Networks

Standardization activities at the IETF have defined an “upper stack” allowing low-power mesh networks to be seamlessly integrated in the Internet (6LoWPAN), form multi-hop topologies (RPL), and interact with other devices like regular web servers (CoAP).

Major research challenges in sensor networks are mostly related to (predictable) power conservation and efficient multi-hop routing. Applications such as monitoring of mobile targets, and the generalization of smart phone devices and wearables, have introduced the need for WSN communication protocols to cope with node mobility and intermittent connectivity.

Extending WSN technology to new application spaces (e.g. security, sports, hostile environments) could also assist communication by seamless exchanges of information between individuals, between individuals and machines, or between machines, leading to the Internet of Things.

4.6.2. Deterministic Low-Power Networks

Wired sensor networks have been used for decades to automate production processes in industrial applications, through standards such as HART. Because of the unreliable nature of the wireless medium, a wireless version of such industrial networks was long considered infeasible.

In 2012, the publication of the IEEE 802.15.4e standard triggered a revolutionary trend in low-power mesh networking: merging the performance of industrial networks, with the ease-of-integration of IP-enabled networks. This integration process is spearheaded by the IETF 6TiSCH working group, created in 2013. A 6TiSCH network implements the IEEE 802.15.4e TSCH protocol, as well as IETF standards such as 6LoWPAN, RPL and CoAP. A 6TiSCH network is synchronized, and a communication schedule orchestrates all communication in the network. Deployments of pre-6TiSCH networks have shown that they can achieve over 99.999% end-to-end reliability, and a decade of battery lifetime.

The communication schedule of a 6TiSCH network can be built and maintained using a centralized, distributed, or hybrid scheduling approach. While the mechanisms for managing that schedule are being standardized by the IETF, which scheduling approach to use, and the associated limits in terms of reliability, throughput and power consumption remain entirely open research questions. Contributing to answering these questions is an important research direction for the EVA team.

4.6.3. MANETs and VANETs

In contrast to routing, other domains in MANETs such as medium access, multi-carrier transmission, quality of service, and quality of experience have received less attention. The establishment of research contracts for EVA in the field of MANETs is expected to remain substantial. MANETs will remain a key application domain for EVA with users such as the military, firefighters, emergency services and NGOs.

Vehicular Ad hoc Networks (VANETs) are arguably one of the most promising applications for MANETs. These networks primarily aim at improving road safety. Radio spectrum has been ring-fenced for VANETs worldwide, especially for safety applications. International standardization bodies are working on building efficient standards to govern vehicle-to-vehicle or vehicle-to-infrastructure communication.

4.6.4. Cellular and Device-to-Device Networks

We propose to initially focus this activity on spectrum sensing. For efficient spectrum sensing, the first step is to discover the links (sub-carriers) on which nodes may initiate communications. In Device-to-Device (D2D) networks, one difficulty is scalability.

For link sensing, we will study and design new random access schemes for D2D networks, starting from active signaling. This will assume the availability of a control channel devoted to D2D neighbor discovery. It is therefore naturally coupled with cognitive radio algorithms (allocating such resources): coordination of link discovery through eNode-B information exchanges can yield further spectrum usage optimization.

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. Large scale networks

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale networks, and developing tools for transportation networks.

HIEPACS Project-Team

4. Application Domains

4.1. Material physics

Participants: Olivier Coulaud, Pierre Esterie.

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 of algorithms in scientific applications

Participants: Emmanuel Agullo, Aurélien Falco, Olivier Beaumont, Lionel Eyraud-Dubois, Mathieu Faverge, Marek Felsoci, Luc Giraud, Gilles Marait, Van Gia Thinh Nguyen, Pierre Ramet, Guillaume Sylvand, Alena Shilova.

4.2.1. High performance simulation for ITER tokamak

Scientific simulation for ITER tokamak modeling provides a natural bridge between theory and experimentation and is also an essential tool for understanding and predicting plasma behavior. Recent progresses in numerical simulation of fine-scale turbulence and in large-scale dynamics of magnetically confined plasma have been enabled by access to petascale supercomputers. These progresses would have been unreachable without new computational methods and adapted reduced models. In particular, the plasma science community has developed codes for which computer runtime scales quite well with the number of processors up to thousands cores. The research activities of HIEPACS concerning the international ITER challenge have started in the Inria Project Lab C2S@Exa in collaboration with CEA-IRFM and were related to two complementary studies: a first one concerning the turbulence of plasma particles inside a tokamak (in the context of GYSELA code) and a second one concerning the MHD instability edge localized modes (in the context of JOREK code). The activity concerning GYSELA was completed at the end of 2018.

Other numerical simulation tools designed for the ITER challenge aim at making a significant progress in understanding active control methods of plasma edge MHD instability Edge Localized Modes (ELMs) which represent a particular danger with respect to heat and particle loads for Plasma Facing Components (PFC) in the tokamak. The goal is to improve the understanding of the related physics and to propose possible new strategies to improve effectiveness of ELM control techniques. The simulation tool used (JOREK code) is related to non linear MHD modeling and is based on a fully implicit time evolution scheme that leads to 3D large very badly conditioned sparse linear systems to be solved at every time step. In this context, the use of PaStiX library to solve efficiently these large sparse problems by a direct method is a challenging issue.

This activity continues within the context of the EoCoE2 project, in which the PaStiX solver is identified to allow the processing of very larger linear systems for the nuclear fusion code TOKAM3X from CEA-IRFM. Contrary to the JOREK code, the problem to be treated corresponds to the complete 3D volume of the plasma torus. The objective is to be competitive, for complex geometries, compared to an Algebraic MultiGrid approach designed by one partner of EoCoE2.

4.2.2. Numerical and parallel scalable hybrid solvers in large scale calculations

Parallel and numerically scalable hybrid solvers based on a fully algebraic coarse space correction have been theoretically studied and various advanced parallel implementations have been designed. Their parallel scalability has been initially investigated on large scale problems within the EoCoE project thanks to a close collaboration with the BSC and the integration of MaPhyS within the Alya software. This activity will further develop in the EoCoE2 project. The performance has also been assessed on PRACE Tier-0 machine within a PRACE Project Access through a collaboration with CERFACS and Laboratoire de Physique des Plasmas at Ecole Polytechnique for the calculation of plasma propulsion. A comparative parallel scalability study with the Algebraic MultiGrid from Petsc has been conducted in that framework.

4.2.3. Aeroacoustics Simulation

This domain is in the context of a long term collaboration with Airbus Research Centers. Wave propagation phenomena intervene in many different aspects of systems design at Airbus. They drive the level of acoustic vibrations that mechanical components have to sustain, a level that one may want to diminish for comfort reason (in the case of aircraft passengers, for instance) or for safety reason (to avoid damage in the case of a payload in a rocket fairing at take-off). Numerical simulations of these phenomena play a central part in the upstream design phase of any such project. Airbus Central R & T has developed over the last decades an in-depth knowledge in the field of Boundary Element Method (BEM) for the simulation of wave propagation in homogeneous media and in frequency domain. To tackle heterogeneous media (such as the jet engine flows, in the case of acoustic simulation), these BEM approaches are coupled with volumic finite elements (FEM). We end up with the need to solve large (several millions unknowns) linear systems of equations composed of a dense part (coming from the BEM domain) and a sparse part (coming from the FEM domain). Various parallel solution techniques are available today, mixing tools created by the academic world (such as the Mumps and Pastix sparse solvers) as well as parallel software tools developed in-house at Airbus (dense solver SPIDO, multipole solver, \mathcal{H} -matrix solver with an open sequential version available online). In the current state of

knowledge and technologies, these methods do not permit to tackle the simulation of aeroacoustics problems at the highest acoustic frequencies (between 5 and 20 kHz, upper limits of human audition) while considering the whole complexity of geometries and phenomena involved (higher acoustic frequency implies smaller mesh sizes that lead to larger unknowns number, a number that grows like f^2 for BEM and f^3 for FEM, where f is the studied frequency). The purpose of the study in this domain is to develop advanced solvers able to tackle this kind of mixed dense/sparse linear systems efficiently on parallel architectures.

4.2.4. Optimization for Deep Convolutional Neural Networks

The training phase of Deep Convolutional Neural Networks represents nowadays a significant share of the computations performed on HPC supercomputers. It introduces several new problems concerning resource allocation and scheduling issues, because of the specific pattern of task graphs induced by the stochastic gradient descent and because memory consumption is particularly critical when performing training. As of today, the most classical parallelization methods consists in partitioning mini-batches, images, filters,... but all these methods induce high synchronization and communication costs, and only very partially resolve memory issues. Within the framework of the Inria IPL on HPC Big Data and Learning convergence, we are working on re-materialization techniques and on the use of model parallelism, in particular to be able to build on the research that has been carried out in a more traditional HPC framework on the exploitation of resource heterogeneity and dynamic runtime scheduling.

INDES Project-Team

4. Application Domains

4.1. Web

The Web is the natural application domain of the team. We are designing and implementing multitier languages for helping the development of Web applications. We are creating static and dynamic analyses for Web security. We are conducting empirical studies about privacy preservation on the Web.

4.2. Internet of Things

More recently, we have started focusing on *Internet of Things* (IoT) applications. They share many similarities with Web applications so most of the methodologies and expertises we have developed for the Web apply to IoT but the restricted hardware resources made available by many IoT devices demand new developments and new research explorations.

KERDATA Project-Team

4. Application Domains

4.1. Application Domains

The KerData team investigates the design and implementation of architectures for data storage and processing across clouds, HPC and edge-based systems, which address the needs of a large spectrum of applications. The use cases we target to validate our research results come from the following domains.

- Climate and meteorology
- Earth science
- Energy and sustainable development
- Smart cities
- Data science

MARACAS Team

4. Application Domains

4.1. 5G, 6G, and beyond

The fifth generation (5G) broadens the usage of cellular networks but requires new features, typically very high rates, high reliability, ultra low latency, for immersive applications, tactile internet, M2M communications.

From the technical side, new elements such as millimeter waves, massive MIMO, massive access are under evaluation. The initial 5G standard validated in 2019, is finally not really disruptive with respect to the 4G and the clear breakthrough is not there yet. The ideal network architecture for billions of devices in the general context of Internet of Things, is not well established and the debate still exists between several proposals such as NB-IoT, Sigfox, Lora. We are developing a deep understanding of these techniques, in collaboration with major actors (Orange Labs, Nokia Bell Labs, Sequans, Sigfox) and we want to be able to evaluate, to compare and to propose evolutions of these standards with an independent point of view.

This is why we are interested in developing partnerships with major industries, access providers but also with service providers to position our research in a joint optimization of the network infrastructure and the data services, from a theoretical perspective as well as from experimentation.

4.2. Energy sustainability

The energy footprint and from a more general perspective, the sustainability of wireless cellular networks and wireless connectivity is somehow questionable.

We develop our models and analysis with a careful consideration of the energy footprint : sleeping modes, power adaptation, interference reduction, energy gathering, ... many techniques can be optimized to reduce the energetic impact of wireless connectivity. In a *computing networks* approach, considering simultaneously transmission, storage and computation constraints may help to reduce drastically the overall energy footprint.

4.3. Smart building, smart cities, smart environments

Smart environments rely on the deployment of many sensors and actuators allowing to create interactions between the twinned virtual and real worlds. These smart environments (e.g. smart building) are for us an ideal playground to develop new models based on information theory and estimation theory to optimize the network architecture including storage, transmission, computation at the right place.

Our work can be seen as the dark side of cloud/edge computing. In collaboration with other teams expert in distributed computing or middleware (typically at CITIlab, with the team Dynamid of Frédéric Le Mouel) and in the framework of the chaire SPIE/ICS-INSA Lyon, we want to optimize the mechanisms associated to these technologies : in a multi-constrained approach, we want to design new distributed algorithms appropriate for large scale smart environments.

4.4. Machine learning based radio

During the first 6G wireless meeting which was held in Lapland, Finland in March 2019, machine learning (ML) was clearly identified as one of the most promising breakthroughs for future 6G wireless systems expected to be in use around 2030 (<https://www.6gsummit.com/>). The research community is entirely leveraging the international ML tsunami. We strongly believe that the paradigm of wireless networks is moving toward to a new era. Our view is supported by the fact that artificial Intelligence (AI) in wireless communications is not new at all. The telecommunications industry has been seeking for 20 years to reduce the operational complexity of communication networks in order to simplify constraints and to reduce costs on deployments. This obviously relies on data-driven techniques allowing the network to self-tune its own parameters. Over the successive 3GPP standard releases, more and more sophisticated network control has been introduced. This has supported increasing flexibility and further self-optimization capabilities for radio resource management (RRM) as well as for network parameters optimization.

We target the following key elements :

- Obtaining data from experimental scenarios, at the lowest level (baseband I/Q signals) in multi-user scenarios (based upon FIT/CorteXlab).
- Developing a framework and algorithms for deep learning based radio.
- Developing new reinforcement learning techniques in high dimensional state-action spaces.
- Embedding NN structures on radio devices (FPGA or m-controllers) and in FIT/CorteXlab.
- Evaluating the gap between these algorithms and fundamental limits from information theory.
- Building an application scenario in a smart environment to experiment a fully cross-layer design (e.g. within a smart-building context, how could a set of object could learn their protocols efficiently ?)

4.5. Molecular communications

Many communication mechanisms are based on acoustic or electromagnetic propagation; however, the general theory of communication is much more widely applicable. One recent proposal is molecular communication, where information is encoded in the type, quantity, or time or release of molecules. This perspective has interesting implications for the understanding of biochemical processes and also chemical-based communication where other signaling schemes are not easy to use (e.g., in mines). Our work in this area focuses on two aspects: (i) the fundamental limits of communication (i.e., how much data can be transmitted within a given period of time); and (ii) signal processing strategies which can be implemented by circuits built from chemical reaction-diffusion systems.

A novel perspective introduced within our work is the incorporation of coexistence constraints. That is, we consider molecular communication in a crowded biochemical environment where communication should not impact pre-existing behavior of the environment. This has lead to new connections with communication subject to security constraints as well as the stability theory of stochastic chemical reaction-diffusion systems and systems of partial differential equations which provide deterministic approximations.

MIMOVE Project-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. One such domain 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 has been and keeps evolving at a fast pace in close concert with the latest technology trends. It is notably worth highlighting how mobile and social network use has 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 focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In particular, we concentrate on the following specialized applications:

- **Democratization of urban data for healthy cities.** We integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances. This goes 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).
- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. We study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development of such "civic apps".

4.2. 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 quality of their Internet access, causes of potential problems and—ideally—ways to fix them. We intend to develop a set of easy-to-use home network monitoring and diagnosis tools. The development of home network monitoring and diagnosis tools brings a number of challenges. First, home networks are heterogeneous. 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, Internet application and services are also heterogeneous with very diverse network requirements. We must develop methods that can infer application quality solely from the observation of (often encrypted) application network traffic. 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. Finally, 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.

In our partnership with Princeton University (associate team HOMENET) we have deployed monitoring infrastructure within users’ homes. We are developing a mostly passive measurement system to monitor the performance of user applications, which we call Network Microscope. We are developing Network Microscope to run in a box acting as home gateway. We have deployed these boxes in 50 homes in the US and 10 in France. The US deployment was ran and financed by the Wall Street Journal. They were interested in understanding the relationship between Internet access speed and video quality. We have been discussing with Internet regulators (in particular, FCC, ACERP, and BEREK) as well as residential access ISP in how Network Microscope can help overcome the shortcomings of existing Internet quality monitoring systems.

4.3. Mobile Internet quality of experience

Mobile Internet usage has boomed with the advent of ever smarter handheld devices and the spread of fast wireless access. People rely on mobile Internet for everyday tasks such as banking, shopping, or entertainment. The importance of mobile Internet in our lives raises people’s expectations. Ensuring good Internet user experience (or Quality of Experience—QoE) is challenging, due to the heavily distributed nature of Internet services. For mobile applications, this goal is even more challenging as access connectivity is less predictable due to user mobility, and the form factor of mobile devices limits the presentation of content. For these reasons, the ability to monitor QoE metrics of mobile applications is essential to determine when the perceived application quality degrades and what causes this degradation in the chain of delivery. Our goal is to improve QoE of mobile applications.

To achieve this goal, we are working on three main scientific objectives. First, we are working on novel methods to monitor mobile QoE. Within the IPL BetterNet we are developing the HostView for Android tool that runs directly on mobile devices to monitor network and system performance together with the user perception of performance. Second, we plan to develop models to predict QoE of mobile applications. We will leverage the datasets collected with HostView for Android to build data-driven models. Finally, our goal is to develop methods to optimize QoE for mobile users. We are currently developing optimization methods for interactive video applications. We envision users walking or driving by road-side WiFi access points (APs) with full 3G/LTE coverage and patchy WiFi coverage (i.e., community Wifi or Wifi APs on Lampposts) or devices with multiple 3G/LTE links. To achieve this goal, we plan to leverage multi-path and cross-layer optimizations.

4.4. Internet Scanning

Internet-wide scanning has enabled researchers to answer a wealth of new security and measurement questions ranging from “How are authoritarian regimes spying on journalists?” to “Are security notifications effective at prompting operators to patch?” Most of these studies have used tools like ZMap, which operates naively, scanning every IPv4 address once. This simplicity enables researchers to easily answer a question once, but the methodology scales poorly when continually scanning to detect changes, as networks change at dramatically

different rates. Service configurations change more frequently on cloud providers like Amazon and Azure than on residential networks. Internet providers in developing regions often have extremely short DHCP windows. Some networks are unstable with host presence varying wildly between different hours and others have distinct periodic patterns, e.g., hosts are only available during regional business hours. A handful of large autonomous systems have not had hosts present in decades. Our work in collaboration with Stanford University is developing more intelligent Internet-wide scanning methods to then implement a system that can scan continuously. Such a system will allow for up-to-date analysis of Internet trends and threats with real-time alerts of important events.

Myriads Project-Team

4. Application Domains

4.1. Main application domains

The Myriads team investigates the design and implementation of system services. Thus its research activities address a broad range of application domains. We validate our research results with selected use cases in the following application domains:

- Smart city services,
- Smart grids,
- Energy and sustainable development,
- Home IoT applications,
- Bio-informatics applications,
- Data science applications,
- Computational science applications,
- Numerical simulations.

NEO Project-Team

4. Application Domains

4.1. Network Science

Network Science is a multidisciplinary body of knowledge, principally concerned with the emergence of global properties in a network of individual agents, from the “local” properties of this network, namely, the way agents interact with each other. The central model of “networks” is the graph (of Graph Theory/Operations Research), with nodes representing the different entities managing information and taking decisions, and the links representing the fact that entities interact, or not. Links are usually equipped with a “weight” that measures the intensity of interaction. Adding evolution rules to this quite elementary representation leads to dynamic network models, the properties of which Network Science tries to analyze.

A classical example of properties sought in networks is the famous “six degrees of separation” (or “small world”) property: how and why does it happen so frequently? Another ubiquitous property of real-life networks is the Zipf or “scale-free” distribution for degrees. Some of these properties, when properly exploited, lead to successful business opportunities: just consider the PageRank algorithm of Google, which miraculously connects the relevance of some Web information with the relevance of the other information that points to it.

4.2. Network Engineering

In its primary acceptance, Network Science involves little or no engineering: phenomena are assumed to be “natural” and emerge without intervention. However, the idea comes fast to intervene in order to modify the outcome of the phenomenon. This is where NEO is positioned. Beyond the mostly descriptive approach of Network Science, we aim at using the techniques of Operations Research so as to engineer complex networks.

To quote just two examples: controlling the spread of diseases through a “network” of people is of primarily interest for mankind. Similarly, controlling the spread of information or reputation through a social network is of great interest in the Internet. Precisely: given the impact of web visibility on business income, it is tempting (and quite common) to manipulate the graph of the web by adding links so as to drive the PageRank algorithm to a desired outcome.

Another interesting example is the engineering of community structures. Recently, thousands of papers have been written on the topic of community *detection* problem. In most of the works, the researchers propose methods, most of the time, heuristics, for detecting communities or dense subgraphs inside a large network. Much less effort has been put in the understanding of community formation process and even much less effort has been dedicated to the question of how one can influence the process of community formation, e.g. in order to increase overlap among communities and reverse the fragmentation of the society.

Our ambition for the medium term is to reach an understanding of the behavior of complex networks that will make us capable of influencing or producing a certain property in said network. For this purpose, we will develop families of models to capture the essential structure, dynamics, and uncertainty of complex networks. The “solution” of these models will provide the correspondence between metrics of interest and model parameters, thus opening the way to the synthesis of effective control techniques.

In the process of tackling real, very large size networks, we increasingly deal with large graph data analysis and the development of decision techniques with low algorithmic complexity, apt at providing answers from large datasets in reasonable time.

POLARIS Project-Team

4. Application Domains

4.1. Large Computing Infrastructures

Supercomputers typically comprise thousands to millions of multi-core CPUs with GPU accelerators interconnected by complex interconnection networks that are typically structured as an intricate hierarchy of network switches. Capacity planning and management of such systems not only raises challenges in term of computing efficiency but also in term of energy consumption. Most legacy (SPMD) applications struggle to benefit from such infrastructure since the slightest failure or load imbalance immediately causes the whole program to stop or at best to waste resources. To scale and handle the stochastic nature of resources, these applications have to rely on dynamic runtimes that schedule computations and communications in an opportunistic way. Such evolution raises challenges not only in terms of programming but also in terms of observation (complexity and dynamicity prevents experiment reproducibility, intrusiveness hinders large scale data collection, ...) and analysis (dynamic and flexible application structures make classical visualization and simulation techniques totally ineffective and require to build on *ad hoc* information on the application structure).

4.2. Next-Generation Wireless Networks

Considerable interest has arisen from the seminal prediction that the use of multiple-input, multiple-output (MIMO) technologies can lead to substantial gains in information throughput in wireless communications, especially when used at a massive level. In particular, by employing multiple inexpensive service antennas, it is possible to exploit spatial multiplexing in the transmission and reception of radio signals, the only physical limit being the number of antennas that can be deployed on a portable device. As a result, the wireless medium can accommodate greater volumes of data traffic without requiring the reallocation (and subsequent re-regulation) of additional frequency bands. In this context, throughput maximization in the presence of interference by neighboring transmitters leads to games with convex action sets (covariance matrices with trace constraints) and individually concave utility functions (each user's Shannon throughput); developing efficient and distributed optimization protocols for such systems is one of the core objectives of Theme 5.

Another major challenge that occurs here is due to the fact that the efficient physical layer optimization of wireless networks relies on perfect (or close to perfect) channel state information (CSI), on both the uplink and the downlink. Due to the vastly increased computational overhead of this feedback – especially in decentralized, small-cell environments – the ongoing transition to fifth generation (5G) wireless networks is expected to go hand-in-hand with distributed learning and optimization methods that can operate reliably in feedback-starved environments. Accordingly, one of POLARIS' application-driven goals will be to leverage the algorithmic output of Theme 5 into a highly adaptive resource allocation framework for next-generation wireless systems that can effectively "learn in the dark", without requiring crippling amounts of feedback.

4.3. Energy and Transportation

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. 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.

4.4. Social Computing Systems

Social computing systems are online digital systems that use personal data of their users at their core to deliver personalized services directly to the users. They are omnipresent and include for instance recommendation systems, social networks, online medias, daily apps, etc. Despite their interest and utility for users, these systems pose critical challenges of privacy, security, transparency, and respect of certain ethical constraints such as fairness. Solving these challenges involves a mix of measurement and/or audit to understand and assess issues, and modeling and optimization to propose and calibrate solutions.

RESIST Team

4. Application Domains

4.1. Internet

Among the different network types, the Internet is the one to link them all and is consequently our most prominent subject, not to mention its prime importance in today's society. The Internet also exhibits its own challenges due to the scale and diversity of stakeholders, applications and network technologies in use.

From a security perspective, **monitoring and analysing Internet traffic is an important part of threat prevention and predictive security**. Indeed, large network telescopes like the one we use in the High Security Laboratory⁰ allow detecting world-wide campaigns of attacks which target a specific exploit in some applications. Moreover the monitoring of the Internet traffic at the **edge** is the best way to quickly detect distributed attacks like DDoS and to mitigate them before they become effective. However, the Internet traffic analysis is made much more complicated since the **massive shift towards encryption** that happened few years ago, which requires new traffic classification methods.

The performance and resilience of services running over the Internet is also a major topic of Resist. In particular, it is very difficult to **diagnose the cause of a degradation of performance among the different actors and technologies** that are used to deliver a service over the Internet (access medium, ISP, CDN, web-browser, etc.). Networked systems deployed at Internet scale are also a natural research subject for Resist. Indeed **decentralized systems** like P2P networks or blockchains are known to be robust and scalable. However, their security and performance have to be carefully assessed because a single flaw in their design can endanger the whole system.

4.2. SDN and Data-Center Networks

As the SDN paradigm and its implementations bring new opportunities that can be leveraged in different contexts, in particular for security and performance, programmable networks are also part of the research scope of Resist. This includes data-plane **programming models and hardware offloading** that enable very flexible programming at the network level. While OpenFlow was initially designed for academic research, SDN in general has then been adopted by industrial players, above all in **data-center networks**. It supports innovations to better share load and optimize resources among processes, in particular for virtualization platforms. Contributing to the development of these technologies is primordial for us as they are key elements for monitoring and enhancing the performance and security of future data-center networks.

When defining or extending SDN technologies, the strongest constraint is to guarantee a satisfactory level of performance, i.e. enabling high flexibility in programming with a **reduced footprint of network throughput**. However, as it may also break isolation principles between multiple tenants, security has to be carefully considered, either by adding safeguard mechanisms at run-time or through a priori verification and testing.

4.3. Fog and Cloud computing

Cloud computing has largely evolved in the last years including new networking capabilities as highlighted in the previous section towards the model of XaaS or **everything-as-a-service**. Moreover, cloud computing continues to be more distributed and aims at integrating more heterogeneous resources. One particular example is **fog computing** that consists of a massively distributed number of different resources, including low-performance ones. Large network operators have a great interest in fog computing because they already operate such an infrastructure (e.g. a national operator with regional clouds and setup boxes in end users' homes). Softwarization or virtualization of all functions and services will help them to be competitive by reducing their costs. In general, intelligent orchestration of massively distributed resources will be investigated in various application domains, including **federated cloud infrastructures, fog computing, 5G networks, IoT and big data infrastructures**.

⁰<https://lhs.loria.fr>

The manageability of such largely distributed systems is a core topic with questions related to monitoring, security and orchestration of resources. Major changes and errors can have dramatic effects on a real system, that actually lead to only minor changes being carried out and slow down innovation and adoption of new propositions. Hence, **controlled and reproducible experiments are vital**.

As shown by our past work, we are able to quickly adjust to experimental needs in most areas of distributed computing and networking, such as *High Performance Computing (HPC)*, *Big Data*, *Peer-to-peer systems*, *Grid computing*, etc. However, in the context of Resist, **we will focus mainly on Software-Defined Infrastructures**, gathering *cloud computing* for compute and storage resources, *software-defined networking* and *network function virtualization* for networking. Those infrastructures share many common features: need for performance, for scalability, for resilience, all implemented using flexible software components.

Worth mentioning here is our involvement in the international testbed community (FIRE, GENI). We plan to strengthen our existing links with the Chameleon and CloudLab US projects, to leverage the recently accepted Fed4FIRE+ project on a testbed federation, and, at the national level, to contribute to the SILECS initiative for a new large-scale experimental computer science infrastructure.

4.4. Cyber-Physical Systems

Cyber-Physical Systems (CPSs) used to be well isolated and so designed accordingly. In the last decade, they have become **integrated within larger systems** and so accessible through the Internet. This is the case with **industrial systems**, like SCADA, that have been unfortunately exposed to major threats. Furthermore, the **Internet-of-Things (IoT)** has become a reality with numerous protocols, platforms and devices being developed and used to support the growing deployment of smart* services: smart home, transport, health, city... and even rather usual rigid systems such as industry 4.0.

From an academic perspective, the IoT can be seen as an evolution of sensor networks. It thus inherits from the same problems regarding security and scalability, but with a higher order of magnitude both in terms of number of devices and their capabilities, which can be exploited by attackers. Research in this area has focused on developing dedicated protocols or operating systems to guarantee security and performance, Resist aims to tackle identical problems but **assuming a more practical deployment of IoT systems composed of heterogeneous and uncontrolled devices**. Indeed, this ecosystem is very rich and **cannot be controlled by a unique entity**, e.g. services are often developed by third parties, manufacturers of embed devices are different from those providing connectivity.

As a result, managing an IoT system (monitoring, changing configuration, etc.) is very hard to achieve as most of the devices or applications cannot be directly controlled. For instance, many IoT providers rely on their own cloud services, with their own unknown **proprietary protocols** and most of the time through **encrypted channels**. Above all, the use of middle-boxes like gateways hides the IoT end-devices and applications. We will thus need to infer knowledge from **indirect and partial observations**. Likewise, control will be also indirect for example through filtering or altering communications.

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 more than 25 company members).

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 analyzing software systems to support understanding and continuous development as well as software quality analysis.

ROMA Project-Team

4. Application Domains

4.1. Applications of sparse direct solvers

Sparse direct (e.g., multifrontal solvers that we develop) solvers have a wide range of applications as they are used at the heart of many numerical methods in computational science: 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 system of linear equations involving sparse matrices. There are therefore a number of application fields, among which some of the ones cited by the users of the sparse direct solver MUMPS are: structural mechanics, seismic modeling, 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.

SOCRATE Project-Team (section vide)

SPIRALS Project-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 built, 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 Cappucino and **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 crowd-sensing 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.

STACK Project-Team

4. Application Domains

4.1. Overview

Supporting industrial actors and open-source communities in building an advanced software management stack is a key element to favor the advent of new kinds of information systems as well as web applications. Augmented reality, telemedicine and e-health services, smart-city, smart-factory, smart-transportation and remote security applications are under investigations. Although, STACK does not intend to address directly the development of such applications, understanding their requirements is critical to identify how the next generation of ICT infrastructures should evolve and what are the appropriate software abstractions for operators, developers and end-users. STACK team members have been exchanging since 2015 with a number of industrial groups (notably Orange Labs and Airbus), a few medical institutes (public and private ones) and several telecommunication operators in order to identify both opportunities and challenges in each of these domains, described hereafter.

4.2. Industrial Internet

The Industrial Internet domain gathers applications related to the convergence between the physical and the virtual world. This convergence has been made possible by the development of small, lightweight and cheap sensors as well as complex industrial physical machines that can be connected to the Internet. It is expected to improve most processes of daily life and decision processes in all societal domains, affecting all corresponding actors, be they individuals and user groups, large companies, SMEs or public institutions. The corresponding applications cover: the improvement of business processes of companies and the management of institutions (*e.g.*, accounting, marketing, cloud manufacturing, etc.); the development of large “smart” applications handling large amounts of geo-distributed data and a large set of resources (video analytics, augmented reality, etc.); the advent of future medical prevention and treatment techniques thanks to the intensive use of ICT systems, etc. We expect our contributions will favor the rise of efficient, correct and sustainable massively geo-distributed infrastructures that are mandatory to design and develop such applications.

4.3. Internet of Skills.

The Internet of Skills is an extension of the Industrial Internet to human activities. It can be seen as the ability to deliver physical experiences remotely (*i.e.*, via the Tactile Internet). Its main supporters advocate that it will revolutionize the way we teach, learn, and interact with pervasive resources. As most applications of the Internet of Skills are related to real time experiences, latency may be even more critical than for the Industrial Internet and raise the locality of computations and resources as a priority. In addition to identifying how Utility Computing infrastructures can cope with this requirement, it is important to determine how the quality of service of such applications should be defined and how latency and bandwidth constraints can be guaranteed at the infrastructure level.

4.4. e-Health

The e-Health domain constitutes an important societal application domain of the two previous areas. The STACK teams is investigating distribution, security and privacy issues in the fields of systems and personalized (aka. precision) medicine. The overall goal in these fields is the development of medication and treatment methods that are tailored towards small groups or even individual patients.

We are working, as part of the ongoing PrivGen CominLabs collaborative project on new means for the sharing of genetic data and applications in the Cloud. More generally, we are applying and developing corresponding techniques for the medical domains of genomics, immunobiology and transplantology in the international network SHLARC and the regional networks SysMics and Oncoshare: there, we investigate how to secure and preserve privacy if potentially sensitive personal data is moved and processed by distributed biomedical analyses.

We are also involved in the SyMeTRIC regional initiative where preliminary studies have been conducted in order to build a common System Medicine computing infrastructure to accelerate the discovery and validation of bio-markers in the fields of oncology, transplantation, and chronic cardiovascular diseases. The challenges were related to the need of being able to perform analyses on data that cannot be moved between distinct locations.

The STACK team will continue to contribute to the e-Health domain by harnessing advanced architectures, applications and infrastructures for the Fog/Edge.

4.5. Network Virtualization and Mobile Edge Services.

Telecom operators have been among the first to advocate the deployment of massively geo-distributed infrastructures, in particular through working groups such as Mobile Edge Computing at the European Telecommunication Standards Institute⁰. The initial reason is that geo-distributed infrastructures will enable Telecom operators to virtualize a large part of their resources and thus reduce capital and operational costs. As an example, we are investigating through the I/O Lab, the joint lab between Orange and Inria, how can a Centralized Radio Access Networks (*a.k.a.* C-RAN or Cloud-RAN) be supported for 5G networks. We highlight that our expertise is not on the network side but rather on where and how we can deploy, allocate and reconfigure software components, which are mandatory to operate a C-RAN infrastructure, in order to guarantee the quality of service expected by the end-users. Finally, working with actors from the network community is a valuable advantage for a distributed system research group such as STACK. Indeed, achievements made within one of the two communities serve the other.

⁰<http://www.etsi.org/news-events/news/1078-2016-04-etsi-mobile-edge-computing-publishes-foundation-specifications>.

STORM Project-Team

4. Application Domains

4.1. Application domains benefiting from HPC

The application domains of this research are the following:

- Bioinformatics (see ADT Gordon 9.2.2)
- Environment, in particular CO_2 capture (see Exa2PRO, 9.3.1)
- Health and heart disease analysis (see EXACARD, 9.2.1)
- Software infrastructures for Telecommunications (see AFF3CT, 7.9 , 9.2.2)
- Aeronautics (collaboration with Airbus, J.-M. Couteyen)

4.2. Application in High performance computing/Big Data

Most of the research of the team has application in the domain of software infrastructure for HPC and BigData (ANR SOLHAR 9.2.1 , Inria ADT SwLoc and Gordon 9.2.2 , IPL HAC-SPECIS and BigData 9.2.3 , PIA project ELCI 9.2 , H2020 projects INTERTWinE and Exa2Pro 9.3.1 and PRACE project PRACE5IP 9.3.1).

TADAAM Project-Team

4. Application Domains

4.1. Mesh-based applications

TADAAM targets scientific simulation applications on large-scale systems, as these applications present huge challenges in terms of performance, locality, scalability, parallelism and data management. Many of these HPC applications use meshes as the basic model for their computation. For instance, PDE-based simulations using finite differences, finite volumes, or finite elements methods operate on meshes that describe the geometry and the physical properties of the simulated objects. This is the case for at least two thirds of the applications selected in the 9th PRACE. call ⁰, which concern quantum mechanics, fluid mechanics, climate, material physic, electromagnetism, etc.

Mesh-based applications not only represent the majority of HPC applications running on existing supercomputing systems, yet also feature properties that should be taken into account to achieve scalability and performance on future large-scale systems. These properties are the following:

Size Datasets are large: some meshes comprise hundreds of millions of elements, or even billions.

Dynamicity In many simulations, meshes are refined or coarsened at each time step, so as to account for the evolution of the physical simulation (moving parts, shockwaves, structural changes in the model resulting from collisions between mesh parts, etc.).

Structure Many meshes are unstructured, and require advanced data structures so as to manage irregularity in data storage.

Topology Due to their rooting in the physical world, meshes exhibit interesting topological properties (low dimensionality embedding, small maximum degree, large diameter, etc.). It is very important to take advantage of these properties when laying out mesh data on systems where communication locality matters.

All these features make mesh-based applications a very interesting and challenging use-case for the research we want to carry out in this project. Moreover, we believe that our proposed approach and solutions will contribute to enhance these applications and allow them to achieve the best possible usage of the available resources of future high-end systems.

⁰<http://www.prace-ri.eu/prace-9th-regular-call/>

TRIBE Project-Team (section vide)

WHISPER Project-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, v4.14, comprises over 16 million lines of code, and supports 30 different families of CPU architectures, around 50 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. [41], numerous research tools have been applied to the Linux kernel, typically for finding bugs [40], [57], [67], [76] or for computing software metrics [46], [78]. In our work, we have studied generic C bugs in Linux code [10], bugs in function protocol usage [51], [52], issues related to the processing of bug reports [71] and crash dumps [45], and the problem of backporting [66], [77], 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.

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 [8], a domain-specific language for describing the low-level interface of a device. Chipounov *et al.* propose RevNic, [34] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [68], [69] 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 [25], Coverity [41], CP-Miner, [56] PR-Miner [57], and Coccinelle [9]. 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 tools for the maintenance of existing drivers.

WIDE Project-Team (section vide)

ALICE Team

4. Application Domains

4.1. Geometric Tools for Simulating Physics with a Computer

Numerical simulation is the main targeted application domain for the geometry processing tools that we develop. Our mesh generation tools will be tested and evaluated within the context of our cooperation with Hutchinson, experts in vibration control, fluid management and sealing system technologies. We think that the hex-dominant meshes that we generate have geometrical properties that make them suitable for some finite element analyses, especially for simulations with large deformations.

We also have a tight collaboration with Wan Chiu Li, a geophysical modeling specialist. He is creating a start-up company MESHSPACE whose goal is to transfer to the industry remeshing tools we developed in the ALICE team. In particular, he uses hexahedral-dominant meshes for geomechanical simulations of gas and oil reservoirs. From a scientific point of view, this use case introduces new types of constraints (alignments with faults and horizons), and allows certain types of nonconformities that we did not consider until now.

Our cooperation with RhinoTerrain pursues the same goal: reconstruction of buildings from point cloud scans allows to perform 3D analysis and studies on insolation, floods and wave propagation, wind and noise simulations necessary for urban planification.

ALMANACH Project-Team

4. Application Domains

4.1. Application domains for ALMAnaCH

ALMAnaCH's research areas cover Natural Language Processing (nowadays identified as a sub-domain of Artificial Intelligence) and Digital Humanities. Application domains are therefore numerous, as witnessed by ALMAnaCH's multiple academic and industrial collaborations, for which see the relevant sections. Examples of application domains for NLP include:

- Information extraction, information retrieval, text mining (ex.: opinion surveys)
- Text generation, text simplification, automatic summarisation
- Spelling correction (writing aid, post-OCR, normalisation of noisy/non-canonical texts)
- Machine translation, computer-aided translation
- Chatbots, conversational agents, question answering systems
- Medical applications (early diagnosis, language-based medical monitoring...)
- Applications in linguistics (modelling languages and their evolution, sociolinguistic studies...)
- Digital humanities (exploitation of text documents, for instance in historical research)

Auctus Team

4. Application Domains

4.1. Factory 4.0

The 4th industrial revolution (factory 4.0) is characterized by the integration of digital technologies into the production process, in order to meet the challenge of customizing services and products. This agility requires making manufacturing and maintenance lines flexible and versatile. This capacity for adaptation is the characteristic of the human being, which puts him at the center of the production apparatus. However, this can no longer be done at the expense of their health and well-being. How then can we reconcile the enhancement of our manual and analytical expertise, the ever desired increase in productivity and manufacturing quality, while reducing the hardship at work? Collaborative robotics, which we are seeking to build, is one of the central solutions to meet this societal challenge. By assisting humans in their most dangerous and painful tasks, it complements or replaces them in their phases of physical and cognitive fragility.

More generally, we are interested in workstation cobotization, in the manufacturing and assembly industry but also in the construction and craft industries. The application areas are related to regional needs in aeronautics, including maintenance, water and waste treatment. In most of these cases, it is possible to define the tasks, evaluate the stakes and added value of our work.

AVIZ Project-Team

4. Application Domains

4.1. Natural Sciences

As part of a CORDI PhD project, we collaborate with researchers at CERN on interactive data visualization using augmented reality, with the goal to better understand this new visualization environment and to support the physicists in analysing their 3D particle collision data. As part of another CORDI PhD project, we collaborate with researchers at the German Center for Climate Computation (DKRZ), to better understand collaborative data exploration and interaction in immersive analytics contexts. Finally, as part of the Inria IPL “Naviscope,” we collaborate with researchers at INRA (as well as other Inria teams) on interactive visualization tools for the exploration of plant embryo development.

4.2. Social Sciences

We collaborate with social science researchers from EHESS Paris on the visualization of dynamic networks; they use our systems (GeneaQuilts [56], Vistorian [64], PAOHVis [7]) and teach them to students and researchers. Our tools are used daily by ethnographers and historians to study the evolution of social relations over time. In the social sciences, many datasets are gathered by individual researchers to answer a specific question, and automated analytical methods cannot be applied to these small datasets. Furthermore, the studies are often focused on specific persons or organizations and not always on the modeling or prediction of the behavior of large populations. The tools we design to visualize complex multivariate dynamic networks are unique and suited to typical research questions shared by a large number of researchers. This line of research is supported by the DataIA “HistorIA” project, and by the “IVAN” European project. We also collaborate on the BART initiative, a joint project with IRT-SystemX on the analysis and visualization of blockchain data, in collaboration with economists from Université Paris-Sud.

4.3. Medicine

We collaborate with CMAP/Polytechnique on the analysis and visualization of CNAM Data “parcours de santé” to help referent doctors and epidemiologists make sense of French health data. In particular, we are working on a subset of the CNAM Data focused on urinary problems, and we have received a very positive feedback from doctors who can see what happens to the patients treated in France vs. what they thought happened through the literature. This project is starting but is already getting a lot of traction from our partners in medicine, epidemiology, and economy of health.

CEDAR Project-Team

4. Application Domains

4.1. Cloud Computing

Cloud computing services are strongly developing and more and more companies and institutions resort to running their computations in the cloud, in order to avoid the hassle of running their own infrastructure. Today's cloud service providers guarantee machine availabilities in their Service Level Agreement (SLA), without any guarantees on performance measures according to a specific cost budget. Running analytics on big data systems require the user not to only reserve the suitable cloud instances over which the big data system will be running, but also setting many system parameters like the degree of parallelism and granularity of scheduling. Choosing values for these parameters, and choosing cloud instances need to meet user objectives regarding latency, throughput and cost measures, which is a complex task if it's done manually by the user. Hence, we need need to transform cloud service models from availability to user performance objective rises and leads to the problem of multi-objective optimization. Research carried out in the team within the ERC project "Big and Fast Data Analytics" aims to develop a novel optimization framework for providing guarantees on the performance while controlling the cost of data processing in the cloud.

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. CEDARresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. Within the ANR ContentCheck project, and also as part of our international collaboration with the AIST institute from Japan, we work on one hand, to lay down foundations for computational data journalism and fact checking, and also work to devise concrete algorithms and platforms to help journalists perform their work better and/or faster. This work is carried in collaboration with Le Monde's "Les Décodeurs".

On a related topic, heterogeneous data integration under a virtual graph abstract model is studied within the ICODA Inria project which has started in September 2017. There, we collaborate with Les Décodeurs as well as with Ouest France and Agence France Presse (AFP). The data and knowledge integration framework resulting from this work will support journalists' effort to organize and analyze their knowledge and exploit it in order to produce new content.

CHORALE Team

4. Application Domains

4.1. Transportation of people and goods

The researches developed in CHORALE can be applied in different applications fields. We are particularly interested in *transportation of people and goods*. CHORALE contribute to the development of Autonomous Connected Vehicles (e.g. Learning, Mapping, Localization, Navigation) and the associated services (e.g. towing, platooning, taxi).

4.2. Assistance and services robots

We are also interested in *assistance and service robotics*. By adding the social level in the representation of the environment, CHORALE develops social and proactive navigation.

4.3. Survey and monitoring of environments

CHORALE studies multi-robots systems to explore, survey and monitor different kind of environments (e.g. agricultural space, forest space, destroyed space), which are of particular importance in the field of *Exploration and monitoring of poorly structured environments*.

CHROMA Project-Team

4. Application Domains

4.1. Introduction

Applications in Chroma are organized in two main domains : **i) Future cars and transportation systems and ii) Services robotics**. These domains correspond to the experimental fields initiated in Grenoble (eMotion team) and in Lyon (CITI lab). However, the scientific objectives described in the previous sections are intended to apply equally to both applicative domains. Even our work on Bayesian Perception is today applied to the intelligent vehicle domain, we aim to generalize to any mobile robots. The same remark applies to the work on multi-agent decision making. We aim to apply algorithms to any fleet of mobile robots (service robots, connected vehicles, UAVs). This is the philosophy of the team since its creation.



Figure 4. Most of the Chroma platforms: the Pepper robot, a fleet of (22) Turtlebot 2, one of the 4 Bebop drones and the equipped Toyota Lexus.

4.2. 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 such as the Google's car project and several future products announcements made by the car industry. Moreover, the legal issue starts to be addressed in USA (see for instance the recent laws in Nevada and in California authorizing autonomous vehicles on roads) and in several other countries (including France).

In this context, we are interested in the development of ADAS⁰ 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).

⁰ Advanced Driver Assistance Systems

Since about 8 years, we are collaborating with Toyota and with Renault-Nissan on these applications (bilateral contracts, PhD Theses, shared patents), but also recently with Volvo group (PhD thesis started in 2016). We are also strongly involved (since 2012) in the innovation project Perfect then now Security for autonomous vehicle of the IRT⁰ Nanoelec (transportation domain). In 2016, we have been awarded a European H2020 ECSEL project⁰ involving major European automotive constructors and car suppliers. In this project, Chroma is focusing on the embedded perception component (models and algorithms, including the certification issue), in collaboration with Renault, Valeo and also with the Inria team TAMIS (Rennes). Chroma is also involved in the ANR project "Valet" (2015-2018) coordinated by the Inria team RITS (Rocquencourt), dealing with automatic redistribution of car-sharing vehicles and parking valet; Chroma is involved in the pedestrian-vehicle interaction for a safe navigation.

In this context, Chroma has two experimental vehicles equipped with various sensors (a Toyota Lexus and a Renault Zoe, see Fig. 4 and Fig. 2 .b), which are maintained by Inria-SED⁰ and that allow the team to perform experiments in realistic traffic conditions (Urban, road and highway environments). The Zoe car has been automated in December 2016, through our collaboration with the team of P. Martinet (IRCCyN Lab, Nantes), that allow new experiments in the team.

4.3. Services robotics

Service robotics is an application domain quickly emerging, and more and more industrial companies (e.g., IS-Robotics, Samsung, LG) are now commercializing service and intervention robotics products such as vacuum cleaner robots, drones for civil or military applications, entertainment robots ... One of the main challenges is to propose robots which are sufficiently robust and autonomous, easily usable by non-specialists, and marked at a reasonable cost. We are involved in developing observation and surveillance systems, by using ground robots and aerial ones, see Fig. 4 . Since 2016, we develop solutions for 3D observation/exploration of complex scenes or environments with a fleet of UAVs (DynaFlock Inria/DGA project, Inria ADT CORDES⁰) or mobile robots (COMODYS FIL project [82]).

A more recent challenge for the coming decade is to develop robotized systems for assisting elderly and/or disabled people. In the continuity of our work in the IPL PAL⁰, we aim to propose smart technologies to assist electric wheelchair users in their displacements and also to control autonomous cars in human crowds (see Figure 17 for illustration). This concerns our recent "Hianic" ANR project. Another emerging application is humanoid robots helping humans at their home or work. In this context, we address the problem of NAMO (Navigation Among Movable Obstacles) in human populated environments (eg. PhD of B. Renault started on 2018). More generally we address navigation and reconnaissance tasks with Pepper humanoids in the context of the RoboCup-Social League.

⁰Institut de Recherche Technologique

⁰ENABLE-S3: European Initiative to Enable Validation for Highly Automated Safe and Secure Systems.

⁰Service Expérimentation et Développement

⁰Coordination d'une Flotte de Drones Connectés pour la Cartographie 3D d'édifices, led by O. Simonin.

⁰Personnaly assisted Living

COML Team

4. Application Domains

4.1. Speech processing for underresourced languages

We plan to apply our algorithms for the unsupervised discovery of speech units to problems relevant to language documentation and the construction of speech processing pipelines for underresourced languages.

4.2. Tools for the analysis of naturalistic speech corpora

Daylong recordings of speech in the wild gives rise a to number of specific analysis difficulties. We plan to use our expertise in speech processing to develop tools for performing signal processing and helping annotation of such resources for the purpose of phonetic or linguistic analysis.

DEFROST Project-Team

4. Application Domains

4.1. Industry

Robotics in the manufacturing industry is already widespread and is one of the strategies put in place to maintain the level of competitiveness of companies based in France and to avoid relocation to cheap labor countries. Yet, in France, it is considered that the level of robotization is insufficient, compared to Germany for instance. One of the challenges is the high investment cost for the acquisition of robotic arms. In recent years, this challenge has led to the development of “generic” and “flexible” (but rigid) robotic solutions that can be mass produced. But their applicability to specific tasks is still challenging or too costly. With the development of 3D printing, we can imagine the development of a complete opposite strategy: a “task-specific” design of robots. Given a task that needs to be performed by a deformable robot, we could optimize its shape and its structure to create the set of desired motions. A second important aspect is the reduction of the manufacturing cost: it is often predicted that the cost of deformable robots will be low compared to classical rigid robots. The robot could be built on one piece using rapid prototyping or 3D printers and be more adapted for collaborative work with operators. In this area, using soft materials is particularly convenient as they provide a mass/carried load ratio several orders of magnitude higher than traditional robots, highly decreasing the kinetic energy thus increasing the motion speed allowed in presence of humans. Moreover, the technology allows more efficient and ergonomic wearable robotic devices, opening the option for exo-skeletons to be used by human operators inside the factories and distribution centers. This remains to be put in place, but it can open new perspectives in robotic applications. A last remarkable property of soft robots is their adaptability to fragile or tortuous environments. For some particular industry fields (chemistry, food industry...) this could also be an advantage compared to existing rigid solutions. For instance, the German company <http://www.festo.com>, key player in the industrial robotics field, is experimenting with deformable trunk robots that exhibit great compliance and adaptability, and we are working on their accurate control.

4.2. Personal and service robotics

The personal and service robotics are considered an important source of economic expansion in the coming years. The potential applications are numerous and in particular include the challenge of finding robotic solutions for active and healthy aging at home. We plan to develop functional orthosis for which it is better not to have a rigid exoskeleton that is particularly uncomfortable. These orthosis will be ideally personalized for each patient and built using rapid prototyping. On this topic, the place of our team will be to provide algorithms for controlling the robots. We will find some partners to build these robots that would fall in the category of “wearable robots”. With this thematic we also connect with a strong pole of excellence of the region on intelligent textiles (see [Up-TeX](#)) and with the strategic plan of Inria (Improving Rehabilitation and Autonomy).

4.3. Entertainment industry and arts

Robots have a long history with entertainment and arts where [animatronics](#) have been used for decades for cinematographic shootings, theater, amusement parks ([Disney’s audio-animatronic](#)) and performing arts. We believe that soft robots could be a good support for art. As an example, last year we collaborated with the artist Dewi Brunet in the creation of animated origami structures (see <https://dewiorigami.com/>).

4.4. Medical Applications

Soft robots have many medical applications as their natural compliance makes them safer than traditional robots when interacting with humans. Such robots can be used for minimally invasive surgery, to access and act on remote parts of the body through minimal incisions in the patient. Applications include laparoscopic and brain surgery, treatment of several cancers including prostate cancer, and cardiology, for example percutaneous coronary interventions.



Figure 1. Dewi Brunet and his “Origami 2.0” exhibit at the AMV in Trélon

As an example, we received an industry grant (CIFRE) with Robocath to work on autonomous catheter navigation. See section 8.2. Another application is cochlear implant surgery.

EX-SITU Project-Team

4. Application Domains

4.1. Creative industries

We work closely with creative professionals in the arts and in design, including music composers, musicians, and sound engineers; painters and illustrators; dancers and choreographers; theater groups; game designers; graphic and industrial designers; and architects.

4.2. Scientific research

We work with creative professionals in the sciences and engineering, including neuroscientists and doctors; programmers and statisticians; chemists and astrophysicists; and researchers in fluid mechanics.

FLOWERS Project-Team

4. Application Domains

4.1. Application Domains

Neuroscience, Developmental Psychology and Cognitive Sciences The computational modelling of life-long learning and development mechanisms achieved in the team centrally targets to contribute to our understanding of the processes of sensorimotor, cognitive and social development in humans. In particular, it provides a methodological basis to analyze the dynamics of the interaction across learning and inference processes, embodiment and the social environment, allowing to formalize precise hypotheses and later on test them in experimental paradigms with animals and humans. A paradigmatic example of this activity is the Neurocuriosity project achieved in collaboration with the cognitive neuroscience lab of Jacqueline Gottlieb, where theoretical models of the mechanisms of information seeking, active learning and spontaneous exploration have been developed in coordination with experimental evidence and investigation, see <https://flowers.inria.fr/neurocuriosityproject/>.

Personal and lifelong learning 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 society. Yet, to realize this vision, important obstacles need to be overcome: these robots will have to evolve in unpredictable homes and learn new skills in a lifelong manner 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. In particular, this application domain requires advances in artificial intelligence that go beyond the current state-of-the-art in fields like deep learning. Currently these techniques require tremendous amounts of data in order to function properly, and they are severely limited in terms of incremental and transfer learning. One of our goals is to drastically reduce the amount of data required in order for this very potent field to work. We try to achieve this by making neural networks aware of their knowledge, i.e. we introduce the concept of uncertainty, and use it as part of intrinsically motivated multitask learning architectures, and combined with techniques of learning by imitation.

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.

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.

GRAPHDECO Project-Team (section vide)

GRAPHIK Project-Team

4. Application Domains

4.1. Agronomy

Agronomy is a strong expertise domain in the area of Montpellier. Some members of GraphIK are INRA researchers (computer scientists). We closely collaborate with the Montpellier research laboratory IATE, a joint unit of INRA and other organisms. A major issue for INRA and more specifically IATE applications is modeling agrifood chains (i.e., the chain of all processes leading from the plants to the final products, including waste treatment). This modeling has several objectives. It provides better understanding of the processes from begin to end, which aids in decision making, with the aim of improving the quality of the products and decreasing the environmental impact. It also facilitates knowledge sharing between researchers, as well as the capitalization of expert knowledge and “know how”. This last point is particularly important in areas strongly related to local know how (like in cheese or wine making), where knowledge is transmitted by experience, with the risk of non-sustainability of the specific skills. An agrifood chain analysis is a highly complex procedure since it relies on numerous criteria of various types: environmental, economical, functional, sanitary, etc. Quality objectives involve different stakeholders, technicians, managers, professional organizations, end-users, public organizations, etc. Since the goals of the implied stakeholders may be divergent dedicated knowledge and representation techniques are to be employed.

4.2. Data Journalism

One of today’s major issues in data science is to design techniques and algorithms that allow analysts to efficiently infer useful information and knowledge by inspecting heterogeneous information sources, from structured data to unstructured content. We take data journalism as an emblematic use-case, which stands at the crossroad of multiple research fields: content analysis, data management, knowledge representation and reasoning, visualization and human-machine interaction. We are particularly interested in issues raised by the design of data and knowledge management systems that will support data journalism. These systems include an ontology (which typically expresses domain knowledge), heterogeneous data sources (provided with their own vocabulary and querying capabilities), and mappings that relate these data sources to the ontological vocabulary. Ontologies play a central role as they act both as a mediation layer that glue together pieces of knowledge extracted from data sources, and as an inference layer that allow to draw new knowledge.

Besides pure knowledge representation and reasoning issues, querying such systems raise issues at the crossroad of data and knowledge management. In particular, although mappings have been widely investigated in databases, they need to be revisited in the light of the reasoning capabilities enabled by the ontology. More generally, the consistency and the efficiency of the system cannot be ensured by considering the components of the system in isolation (i.e., the ontology, data sources and mappings), but require to study the interactions between these components and to consider the system as a whole.

HEPHAISTOS Project-Team (section vide)

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 movie industry, content customization, video games or attractions in theme parks,
- **Construction:** with virtual mock-ups design and review, or historical/architectural visits.
- **Cultural Heritage:** with acquisition, virtual excavation, virtual reconstruction and visualization

ILDA Project-Team

4. Application Domains

4.1. Mission-critical systems

Mission-critical contexts of use include emergency response & management, and critical infrastructure operations, such as public transportation systems, communications and power distribution networks, or the operations of large scientific instruments such as particle accelerators and astronomical observatories. Central to these contexts of work is the notion of situation awareness [25], i.e., how workers perceive and understand elements of the environment with respect to time and space, such as maps and geolocated data feeds from the field, and how they form mental models that help them predict future states of those elements. One of the main challenges is how to best assist subject-matter experts in constructing correct mental models and making informed decisions, often under time pressure. This can be achieved by providing them with, or helping them efficiently identify and correlate, relevant and timely information extracted from large amounts of raw data, taking into account the often cooperative nature of their work and the need for task coordination. With this application area, our goal is to investigate novel ways of interacting with computing systems that improve collaborative data analysis capabilities and decision support assistance in a mission-critical, often time-constrained, work context.

Relevant publications by team members this year: [13], [19], [12].

4.2. Exploratory analysis of scientific data

Many scientific disciplines are increasingly data-driven, including astronomy, molecular biology, particle physics, or neuroanatomy. While making the right decision under time pressure is often less of a critical issue when analyzing scientific data, at least not on the same temporal scale as truly time-critical systems, scientists are still faced with large-to-huge amounts of data. No matter their origin (experiments, remote observations, large-scale simulations), these data are difficult to understand and analyze in depth because of their sheer size and complexity. Challenges include how to help scientists freely-yet-efficiently explore their data, keep a trace of the multiple data processing paths they considered to verify their hypotheses and make it easy to backtrack, and how to relate observations made on different parts of the data and insights gained at different moments during the exploration process. With this application area, our goal is to investigate how data-centric interactive systems can improve collaborative scientific data exploration, where users' goals are more open-ended, and where roles, collaboration and coordination patterns [48] differ from those observed in mission-critical contexts of work.

Relevant publications by team members last year: [16], [24], [14], [18].

IMAGINE Project-Team

4. Application Domains

4.1. Target applications

Our 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 theater 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. 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 7.

4.2. Visual arts

- Sculpture.
- Modeling and animation for 3D films and games.
- Virtual cinematography and tools for theater directors.

4.3. Engineering

- Industrial design.
- Mechanical & civil engineering.

4.4. Natural sciences

- Geology.
- Virtual functional anatomy.

4.5. Education and creative tools

- Sketch-based teaching.
- Creative environments for novice users.
- Museography.

LACODAM Project-Team

4. Application Domains

4.1. Introduction

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and LACODAM is not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrial partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

4.2. Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us (e.g. Energiency, Amossys).

- **Resource Consumption Analysis for Optimizing Energy Consumption and Practices in Industrial Factories (Energiency).** In order to increase their understanding of factory operation, companies introduce more and more sensors in their factories. Thus, the resource (electricity, water, etc.) consumption of engines, workshops and factories are recorded in the form of times series or temporal sequences. The person who is in charge of resource consumption optimization needs better software than classical spreadsheets for this purpose. He/she needs effective decision-aiding tools with statistical and artificial intelligence knowledge. The start-up Energiency aims at designing and offering such pieces of software for analyzing energy consumption. The CIFRE PhD thesis of Maël (defended 16/12/2019) aimed at proposing new approaches and solutions from the data mining field to tackle this issue.
- **Security (Amossys).** Current networks are faced with an increasing variety of attacks, from the classic “DDoS” that makes a server unusable for a few hours, to advanced attacks that silently infiltrate a network and exfiltrate sensitive information months or even years later. Such intrusions, called APT (Advanced Persistent Threat) are extremely hard to detect, and this will become even harder as most communications will be encrypted. A promising solution is to work on “behavioral analysis”, by discovering patterns based on the metadata of IP-packets. Such patterns can relate to an unusual sequencing of events, or to an unusual communication graph. Finding such complex patterns over a large volume of streaming data requires to revisit existing stream mining algorithms to dramatically improve their throughput, while guaranteeing a manageable false positive rate. We collaborated on this topic with the Amossys company and the EMSEC team of Irista through the co-supervision of the CIFRE PhD of Alban Siffer (located in the EMSEC team, defended 19/12/2019). Our goal was in particular to design novel anomaly detection methods making minimal assumptions on the data, and able to scale to real traffic volumes.
- **Car Sharing Data Analysis.** Peugeot-Citroën (PSA) group’s know-how encompasses all areas of the automotive industry, from production to distribution and services. Among others, its aim is to provide a car sharing service in many large cities. This service consists in providing a fleet of cars and a “free floating” system that allows users to use a vehicle, then drop it off at their convenience in the city. To optimize their fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of the cars and understand the mobility needs and behavior of their users. We tackle this subject together through the CIFRE PhD of Gregory Martin.
- **Multimodal Data Analysis for the Supervision of Sensitive Sites.** ATERMES is an international mid-sized company with a strong expertise in high technology and system integration from the upstream design to the long-life maintenance cycle. It has recently developed a new product, called BARRIER TM (“Beacon Autonomous Reconnaissance Identification and Evaluation Response”),

which provides operational and tactical solutions for mastering borders and areas. Once in place, the system allows for a continuous night and day surveillance mission with a small crew in the most unexpected rugged terrain. The CIFRE PhD of Heng Zhang aims at developing a deep learning architecture and algorithms able to detect anomalies (mainly persons) from multimodal data. The data are “multimodal” because information about the same phenomenon can be acquired from different types of detectors, at different conditions, in multiple experiments.

- **Root Cause Analysis in Networks.** AdvisorSLA is a French company specialized in software solutions for network monitoring. For this purpose, the company relies on techniques of network metrology. By continuously measuring the state of the network, monitoring solutions detect events (e.g., overloaded router) that may degrade the network’s operation and the quality of the services running on top of it (e.g., video transmission could become choppy). When a monitoring solution detects a potentially problematic sequence of events, it triggers an alarm so that the network manager can take actions. Those actions can be preventive or corrective. Some statistics show that only 40% of the triggered alarms are conclusive, that is, they manage to signal a well-understood problem that requires an action from the network manager. This means that the remaining 60% are presumably false alarms. While false alarms do not hinder network operation, they do incur an important cost in terms of human resources. Thus, the CIFRE PhD of Erwan Bourrand proposes to characterize conclusive and false alarms. This will be achieved by designing automatic methods to “learn” the conditions that most likely precede each type of alarm, and therefore predict whether the alarm will be conclusive or not. This can help adjust existing monitoring solutions in order to improve their accuracy. Besides, it can help network managers automatically trace the causes of a problem in the network.

4.3. Health

- **Care Pathways Analysis for Supporting Pharmaco-Epidemiological Studies.** Pharmaco-epidemiology applies the methodologies developed in general epidemiology to answer to questions about the uses and effects of health products, drugs [31], [30] or medical devices [23], on population. In classical pharmaco-epidemiology studies, people who share common characteristics are recruited to build a dedicated prospective cohort. Then, meaningful data (drug exposures, diseases, etc.) are collected from the cohort within a defined period of time. Finally, a statistical analysis highlights the links (or the lack of links) between drug exposures and outcomes (e.g., adverse effects). The main drawback of prospective cohort studies is the time required to collect the data and to integrate them. Indeed, in some cases of health product safety, health authorities have to answer quickly to pharmaco-epidemiology questions.

New approaches of pharmaco-epidemiology consist in using large EHR (Electronic Health Records) databases to investigate the effects and uses (or misuses) of drugs in real conditions. The objective is to benefit from nationwide available data to answer accurately and in a short time pharmaco-epidemiological queries for national public health institutions. Despite the potential availability of the data, their size and complexity make their analysis long and tremendous. The challenge we tackle is the conception of a generic digital toolbox to support the efficient design of a broad range of pharmaco-epidemiology studies from EHR databases. We propose to use pattern mining algorithms and reasoning techniques to analyse the typical care pathways of specific groups of patients.

To answer the broad range of pharmaco-epidemiological queries from national public health institutions, the PEPS⁰ platform exploits, in secondary use, the French health cross-schemes insurance system, called SNDS. The SNDS covers most of the French population with a sliding period of 3 past years. The main characteristics of this data warehouse are described in [29]. Contrary to local hospital EHR or even to other national initiatives, the SNDS data warehouse covers a huge population. It makes possible studies on unfrequent drugs or diseases in real conditions of use. To tackle the volume and the diversity of the SNDS data warehouse, a research program has been established

⁰PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

to design an innovative toolbox. This research program is focused first on the modeling of care pathways from the SNDS database and, second, on the design of tools supporting the extraction of insights about massive and complex care pathways by clinicians. In such a database a care pathway is an individual sequence of drugs exposures, medical procedures and hospitalizations.

- **Care Sequences for the Exploration of Medico-administrative Data.** The difficulty of analyzing medico-administrative data is the semantic gap between the raw data (for example, database record about the delivery at date t of drug with ATC 2 code N 02BE01) and the nature of the events sought by clinicians (“was the patient exposed to a daily dose of paracetamol higher than 3g?”). The solution that is used by epidemiologists consists in enriching the data with new types of events that, on the one side, could be generated from raw data and on the other side, have a medical interpretation. Such new abstract events are defined by clinician using proxies. For example, drugs deliveries can be translated in periods of drug exposure (drug exposure is a time-dependent variable for non-random reasons) or identify patient stages of illness, etc. A proxy can be seen as an abstract description of a care sequence.

Currently, the clinicians are limited in the expression of these proxies both by the coarse expressivity of their tools and by the need to process efficiently large amount of data. From a semantic point of view, care sequences must fully integrate the temporal and taxonomic dimensions of the data to provide significant expression power. From a computational point of view, the methods employed must make it possible to efficiently handle large amounts of data (several millions care pathways). The aim of the PhD of Johanne Bakalara is to study temporal models of sequences in order 1) to show their abilities to specify complex proxies representing care sequences needed in pharmaco-epidemiological studies and 2) to build an efficient querying tool able to exploit large amount of care pathways.

4.4. Agriculture and environment

- **Dairy Farming.** The use and analysis of data acquired in dairy farming is a challenge both for data science and animal science. The goal is to improve farming conditions, i.e., health, welfare and environment, as well as farmers’ income. Nowadays, animals are monitored by multiple sensors giving a wealth of heterogeneous data such as temperature, weight, or milk composition. Current techniques used by animal scientists focus mostly on mono-sensor approaches. The dynamic combination of several sensors could provide new services and information useful for dairy farming. The PhD thesis of Kevin Fauvel (#DigitAg grant), aims to study such combinations of sensors and to investigate the use data mining methods, especially pattern mining algorithms. The challenge is to design new algorithms that take into account data heterogeneity—in terms of nature and time units—, and that produce useful patterns for dairy farming. The outcome of this thesis will be an original and important contribution to the new challenge of the IoT (Internet of Things) and will interest domain actors to find new added value to a global data analysis. The PhD thesis, started on October 2017, takes place in an interdisciplinary setting bringing together computer scientists from Inria and animal scientists from INRA, both located in Rennes.
- **Optimizing the Nutrition of Individual Sow.** Another direction for further research is the combination of data flows with prediction models in order to learn nutrition strategies. Raphaël Gauthier started a PhD thesis (#DigitAg Grant) in November 2017 with both Inria and INRA supervisors. His research addresses the problem of finding the optimal diet to be supplied to individual sows. Given all the information available, e.g., time-series information about previous feeding, environmental data, scientists models, the research goal is to design new algorithms to determine the optimal ration for a given sow in a given day. Efficiency issues of developed algorithms will be considered since the proposed software should work in real-time on the automated feeder. The decision support process should involve the stakeholder to ensure a good level of acceptance, confidence and understanding of the final tool.
- **Ecosystem Modeling and Management.** Ongoing research on ecosystem management includes

modelling of ecosystems and anthropogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to represent these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

4.5. Education

- **Data-oriented Academic Counseling.** Course selection and recommendation are important aspects of any academic counseling system. The Learning Analytics community has long supported these activities via automatic, data-based tools for recommendation and prediction. LACODAM, in collaboration with the Ecuadorian research center CTI⁰ has contributed to this body of research with the design of a tool that allows students to select multiple courses and predict their academic performance based on historical academic data. The tool resorts to interpretable machine learning techniques, and is intended to be used by the students before the counseling sessions to plan their upcoming semester at the Ecuadorian university ESPOL. The data visualization aspects of the tool, as well as the data science considerations and our emphasis on explainability are compiled in a paper that is under revision at the Learning Analytics and Knowledge Conference (LAK'20). The data science component of the tool was developed during the M1 internship of Mohammad Poul-Doust.

4.6. Others

- **RDF Archiving.** The dynamicity of the Semantic Web has motivated the development of solutions for RDF archiving, i.e., the task of storing and querying all previous versions of an RDF dataset. Notwithstanding the value of RDF archiving for data maintainers and consumers, this field of research remains under-developed for multiple reasons. These include notably (i) the lack of usability and scalability of the existing systems, (ii) no archiving support for multi-graph RDF datasets, (iii) the absence of a standard SPARQL extension for RDF archives, and (iv) a disregard of the evolution patterns of RDF datasets. The PhD thesis of Olivier Pelgrin aims at developing techniques towards a scalable and full-fledged archiving solution for the Semantic Web. This PhD thesis is a collaboration between LACODAM and the DAISY team at Aalborg University.

⁰Centro de Tecnologías de Información, <http://cti.espol.edu.ec/>

LARSEN Project-Team

4. Application Domains

4.1. Personal Assistance

During the last fifty years, many medical advances as well as the improvement of the quality of life have resulted in a longer life expectancy in industrial societies. The increase in the number of elderly people is a matter of public health because although elderly people can age in good health, old age also causes embrittlement, in particular on the physical plan which can result in a loss of autonomy. That will force us to re-think the current model regarding the care of elderly people.⁰ Capacity limits in specialized institutes, along with the preference of elderly people to stay at home as long as possible, explain a growing need for specific services at home.

Ambient intelligence technologies and robotics could contribute to this societal challenge. The spectrum of possible actions in the field of elderly assistance is very large. We will focus on activity monitoring services, mobility or daily activity aids, medical rehabilitation, and social interactions. This will be based on the experimental infrastructure we have built in Nancy (Smart apartment platform) as well as the deep collaboration we have with OHS.⁰

4.2. Civil Robotics

Many applications for robotics technology exist within the services provided by national and local government. Typical applications include civil infrastructure services⁰ such as: urban maintenance and cleaning; civil security services; emergency services involved in disaster management including search and rescue; environmental services such as surveillance of rivers, air quality, and pollution. These applications may be carried out by a wide variety of robot and operating modality, ranging from single robots or small fleets of homogeneous or heterogeneous robots. Often robot teams will need to cooperate to span a large workspace, for example in urban rubbish collection, and operate in potentially hostile environments, for example in disaster management. These systems are also likely to have extensive interaction with people and their environments.

The skills required for civil robots match those developed in the LARSEN project: operating for a long time in potentially hostile environment, potentially with small fleets of robots, and potentially in interaction with people.

⁰See the Robotics 2020 Multi-Annual Roadmap [35].

⁰OHS (*Office d'Hygiène Sociale*) is an association managing several rehabilitation or retirement home structures.

⁰See the Robotics 2020 Multi-Annual Roadmap [35], section 2.5.

LINKMEDIA Project-Team

4. Application Domains

4.1. Asset management in the entertainment business

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 bibliothèque Nationale de France, 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. This also concerns TV on-demand and replay services as well as social TV services and multi-screen applications. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases is central here.

4.3. Data journalism

Data journalism forms an application domain where most of the technology developed by LINKMEDIA can be used. On the one hand, data journalists often need to inspect multiple heterogeneous information sources, some being well structured, some other being fully unstructured. They need to access (possibly their own) archives with either searching or navigational means. To gradually construct insight, they need collaborative multimedia analytics processes as well as elements of trust in the information they use as foundations for their investigations. Trust in the information, watching for adversarial and/or (deep) fake material, accountability are all crucial here.

LINKS Project-Team

4. Application Domains

4.1. Linked Data Integration

There are many contexts in which integrating linked data is interesting. We advocate here one possible scenario, namely that of integrating business linked data to feed what is called Business Intelligence. The latter consists of a set of theories and methodologies that transform raw data into meaningful and useful information for business purposes (from Wikipedia). 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 and businessmen need to face the problem of information explosion, due to the Internet's ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of business intelligence within the enterprise.

4.2. Data Cleaning

The second example of application of our proposal 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. We envision a publication management system that exploits both links between database elements, namely pointers to external resources and logical links. The latter can be complex relationships between local portions of data and remote resources, encoded as schema mappings. There are different tasks that such a scenario could entail such as (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 enrich the list of publications of a given research institute, and (iii) support complex queries on the corrected data combined with logical mappings.

4.3. Real Time Complex Event Processing

Complex event processing serves for monitoring nested word streams in real time. Complex event streams are gaining popularity with social networks such as with Facebook and Twitter, and thus should be supported by distributed databases on the Web. Since this is not yet the case, there remains much space for future industrial transfer related to Links' second axis on dynamic linked data.

LOKI Project-Team

4. Application Domains

4.1. Application Domains

Loki works on fundamental and technological aspects of Human-Computer Interaction that can be applied to diverse application domains.

Our 2019 research involved desktop, augmented reality, touch-based, haptics, and BCI interfaces with notable applications to medicine (analysis of fine motor control for patients with Parkinson disease), digital humanities (interpretation of handwritten historical documents), as well as creativity support tools (production of illustrations, design of Digital Musical Instruments). Our technical work also contributes to the more general application domains of software engineering and systems' design.

MAGNET Project-Team

4. Application Domains

4.1. Application Domains

Our main targeted applications are browsing, monitoring, recommending and mining in information networks. 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.

We also target applications related to decentralized learning and privacy preserving systems when users or devices are interconnected in large networks. We develop solutions based on urban and mobility data where privacy is a specific requirement.

MAGRIT 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 20 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 assist therapeutic decisions and interventional gestures. Contributions of the team focus on the development of AR tools for neuro-navigation as well as the development of simulation tools 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 MIMESIS project team.

4.3. Experimental mechanics

In experimental solid mechanics, an important problem is to characterize properties of specimens 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.

MANAO Project-Team

4. Application Domains

4.1. Physical Systems

Given our close relationships with researchers in optics, one novelty of our approach is to extend the range of possible observers to physical sensors in order to work on domains such as simulation, mixed reality, and testing. Capturing, processing, and visualizing complex data is now more and more accessible to everyone, leading to the possible convergence of real and virtual worlds through visual signals. This signal is traditionally captured by cameras. It is now possible to augment them by projecting (e.g., the infrared laser of Microsoft Kinect) and capturing (e.g., GPS localization) other signals that are outside the visible range. This supplemental information replaces values traditionally extracted from standard images and thus lowers down requirements in computational power. Since the captured images are the result of the interactions between light, shape, and matter, the approaches and the improved knowledge from *MANAO* help in designing interactive acquisition and rendering technologies that are required to merge the real and the virtual worlds. With the resulting unified systems (optical and digital), transfer of pertinent information is favored and inefficient conversion is likely avoided, leading to new uses in interactive computer graphics applications, like **augmented reality**, **displays** and **computational photography**.

4.2. Interactive Visualization and Modeling

This direction includes domains such as **scientific illustration and visualization**, **artistic or plausible rendering**, and **3D modeling**. In all these cases, the observer, a human, takes part in the process, justifying once more our focus on real-time methods. When targeting average users, characteristics as well as limitations of the human visual system should be taken into account: in particular, it is known that some configurations of light, shape, and matter have masking and facilitation effects on visual perception. For specialized applications (such as archeology), the expertise of the final user and the constraints for 3D user interfaces lead to new uses and dedicated solutions for models and algorithms.

MAVERICK Project-Team

4. Application Domains

4.1. Application Domains

The natural application domain for our research is the production of digital images, for example for movies and special effects, virtual prototyping, video games... Our research have also been applied to tools for generating and editing images and textures, for example generating textures for maps. Our current application domains are:

- Offline and real-time rendering in movie special effects and video games;
- Virtual prototyping;
- Scientific visualization;
- Content modeling and generation (e.g. generating texture for video games, capturing reflectance properties, etc);
- Image creation and manipulation.

MFx Project-Team

4. Application Domains

4.1. Digital Manufacturing

Our work addresses generic challenges related to fabrication and can thus be applied in a wide variety of contexts. Our aim is first and foremost to develop the algorithms that will allow various industrial sectors to benefit more strongly from the potential of AM. To enable this, we seek collaborations with key industry partners developing software and AM systems for a variety of processes and materials that are of interest to specific sectors (*e.g.*, dental, prosthetic, automotive, aerospace).

4.2. Medical Applications

To allow for faster transfer of our techniques and unlock novel applications, we actively seek to develop applications in the medical sector. In particular, we are involved in a project around the design of orthoses which explores how our research on elasticity control through microstructure geometries can be specifically applied to the medical sector; see §9.1.1 .

MIMETIC Project-Team

4. Application Domains

4.1. Animation, Autonomous Characters and Digital Storytelling

Computer Animation is one of the main application domain of the research work conducted in the MimeTIC team, in particular in relation to the entertainment and game industries. In these domains, creating virtual characters that are able to replicate real human motions and behaviours still highlights unanswered key challenges, especially as virtual characters are becoming more and more required to populate virtual worlds. For instance, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors, which requires to create high quality replicas that appear, move and behave both individually and collectively like real humans. The three key challenges for the MimeTIC team are therefore (i) to create natural animations (i.e., virtual characters that move like real humans), (ii) to create autonomous characters (i.e., that behave like real humans) and (iii) to orchestrate the virtual characters so as to create interactive stories.

First, our challenge is therefore to create animations of virtual characters that are natural, in the largest sense of the term of moving like a real human real would. This challenge covers several aspects of Character Animation depending on the context of application, e.g., producing visually plausible or physically correct motions, producing natural motion sequences, etc. Our goal is therefore to develop novel methods for animating virtual characters, e.g., based on motion capture, data-driven approaches, or learning approaches. However, because of the complexity of human motion (e.g., the number of degrees of freedom that can be controlled), resulting animations are not necessarily physically, biomechanically, or visually plausible. For instance, current physics-based approaches produce physically correct motions but not necessarily perceptually plausible ones. All these reasons are why most entertainment industries still mainly rely on manual animation, e.g., in games and movies. Therefore, research in MimeTIC on character animation is also conducted with the goal of validating objective (e.g., physical, biomechanical) as well as subjective (e.g., visual plausibility) criteria.

Second, one of the main challenges in terms 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). For instance, autonomous characters play the role of actors that are driven by a scenario in video games and virtual storytelling. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes. In the MimeTIC team, our focus is therefore 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.

Finally, 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.

4.2. Fidelity of Virtual Reality

VR is a powerful tool for perception-action experiments. VR-based experimental platforms allow exposing a population to fully controlled stimuli that can be repeated from trial to trial with high accuracy. Factors can be isolated and objects manipulations (position, size, orientation, appearance, ..) are easy to perform. Stimuli can be interactive and adapted to participants' responses. Such interesting features allow researchers to use VR to perform experiments in sports, motion control, perceptual control laws, spatial cognition as well as person-person interactions. However, the interaction loop between users and their environment differs in virtual conditions in comparison with real conditions. When a user interact in an environment, movement from action and perception are closely related. While moving, the perceptual system (vision, proprioception,..) provides feedback about the users' own motion and information about the surrounding environment. That allows the user to adapt his/her trajectory to sudden changes in the environment and generate a safe and efficient motion. In virtual conditions, the interaction loop is more complex because it involves several material aspects.

First, the virtual environment is perceived through a numerical display which could affect the available information and thus could potentially introduce a bias. For example, studies observed a distance compression effect in VR, partially explained by the use of Head Mounted Display with reduced field of view and exerting a weight and torques on the user's head. Similarly, the perceived velocity in a VR environment differs from the real world velocity, introducing an additional bias. Other factors, such as the image contrast, delays in the displayed motion and the point of view can also influence efficiency in VR. The second point concerns the user's motion in the virtual world. The user can actually move if the virtual room is big enough or if wearing a head mounted display. Even with a real motion, authors showed that walking speed is decreased, personal space size is modified and navigation in VR is performed with increased gait instability. Although natural locomotion is certainly the most ecological approach, the physical limited size of VR setups prevents from using it most of the time. Locomotion interfaces are therefore required. Locomotion interfaces are made up of two components, a locomotion metaphor (device) and a transfer function (software), that can also introduce bias in the generated motion. Indeed, the actuating movement of the locomotion metaphor can significantly differ from real walking and the simulated motion depends on the transfer function applied. Locomotion interfaces cannot usually preserve all the sensory channels involved in locomotion.

When studying human behavior in VR, the aforementioned factors in the interaction loop potentially introduce bias both in the perception and in the generation of motor behavior trajectories. MimeTIC is working on the mandatory step of VR validation to make it usable for capturing and analyzing human motion.

4.3. Motion Sensing of Human Activity

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 problems 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), Qualisys, 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 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.

4.4. Sports

Sport is characterized by complex displacements and motions. One main objective is to understand the determinants of performance through the analysis of the motion itself. In the team, different sports have been studied such as the tennis serve, where the goal was to understand the contribution of each segment of the body in the performance but also the risk of injuries as well as other situation in cycling, swimming, fencing or soccer. Sports motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. 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 sequentially 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 understand sports and interactions 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.

4.5. 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 VEs is to facilitate the design process, enhance the design efficiency, and reduce the costs.

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

In real working condition, motion analysis and musculoskeletal risks assessment raise also many scientific and technological challenges. Similarly to virtual reality, fidelity of the working process may be affected by the measurement method. Wearing sensors or skin markers, together with the need of frequently calibrating the assessment system may change the way workers perform the tasks. Whatever the measurement is, classical ergonomic assessments generally address one specific parameter, such as posture, or force, or repetitions..., which makes it difficult to design a musculoskeletal risk factor that actually represent this risk. Another key scientific challenge is then to design new indicators that better capture the risk of musculoskeletal disorders. However, this indicator has to deal with the tradeoff between accurate biomechanical assessment and the difficulty to get reliable and the required information in real working conditions.

4.6. Locomotion and Interactions between walkers

Modeling and simulating locomotion and interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as mathematics, cognitive sciences, physics, computer graphics, rehabilitation etc. Locomotion and interactions between walkers are by definition at the very core of our society since they represent the basic synergies of our daily life. When walking in the street, we should produce a locomotor movement while taking information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape to somebody. MimeTIC is an international key contributor in the domain of understanding and simulating locomotion and interactions between walkers. By combining an approach based on Human Movement Sciences and Computer Sciences, the team focuses on locomotor invariants which characterize the generation of locomotor trajectories, conducts challenging experiments focusing on visuo-motor coordination involved during interactions between walkers both using real and virtual set-ups. One main challenge is to consider and model not only the "average" behaviour of healthy younger adult but also extend to specific populations considering the effect of pathology or the effect of age (kids, older adults). As a first example, when patients cannot walk efficiently, in particular those suffering from central nervous system affections, it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To facilitate such evaluations, we have developed two complementary indices, one based on kinematics and the other one on muscle 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. We extend this objective evaluation by using person-person interaction paradigm which allows studying visuo-motor strategies deficit in these specific populations.

Another fundamental question is the adaptation of the walking pattern according to anatomical constraints, such as pathologies in orthopedics, or adaptation to various human and non-human primates in paleoanthropology. Hence, the question is to predict plausible locomotion according to a given morphology. This question raises fundamental questions about the variables that are regulated to control gait: balance control, minimum energy, minimum jerk...In MimeTIC we develop models and simulators to efficiently test hypothesis on gait control for given morphologies.

MOEX Project-Team (section vide)

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 scanners 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 analysis 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.

4.3. Human Motion Analysis

The recovery of dense motion information enables the combined analysis 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 analyzed. 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. Virtual and Augmented Reality

This domain has actually seen new devices emerge that enable now full 3D visualization, for instance the HTC Vive, the Microsoft Hololens and the Magic leap one. These devices create a need for adapted animated 3D contents that can either be generated or captured. We believe that captured 4D models will gain interest in this context since they provide realistic visual information on moving shapes that tend to avoid negative perception effects such as the uncanny valley effect. Besides 3D visualization devices, many recent applications also rely on everyday devices, such as mobile phones, to display augmented reality contents with free viewpoint ability. In this case, 3D and 4D contents are also expected.

MULTISPEECH Project-Team

4. Application Domains

4.1. Introduction

Approaches and models developed in the MULTISPEECH project are intended to be used for facilitating oral communication in various situations through enhancements of 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 multimodal computer interaction, annotation and processing of spoken documents, health and autonomy (more precisely aided communication and monitoring), and computer assisted learning.

4.2. Multimodal Computer Interactions

Speech synthesis has tremendous applications 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 a story, such as an audiobook, as well as for better human-machine interactions.

4.3. Annotation and Processing of Spoken Documents and Audio Archives

A 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.

A 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 such as 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, speech enhancement, prosody modification, and speaker conversion.

4.4. Aided Communication and Monitoring

Source separation techniques should help for locating and monitoring people through the detection of sound events inside apartments, and speech enhancement is mandatory for hands-free vocal interactions. A foreseen application aims at improving the autonomy of elderly or disabled people, and also fits with smartroom applications. In a longer perspective, adapting speech recognition technologies to the voice of elderly 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) can also be envisaged.

4.5. Computer Assisted Learning

Although speaking seems quite natural, learning foreign languages, or learning the mother tongue for people with language deficiencies, represents 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. Based upon an analysis of the learner's production, automatic diagnoses can be considered. However, making a reliable diagnosis on each individual utterance is still a challenge, which is dependent on the precision and quality of the segmentation of the speech utterance into phones, and of the computed prosodic parameters.

ORPAILLEUR Project-Team

4. Application Domains

4.1. Life Sciences: Agronomy, Biology, Chemistry, and Medicine

Keywords: knowledge discovery in life sciences, biology, chemistry, medicine, pharmacogenomics and precision medicine.

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. 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 the environment, or between genetic and phenotypic data, especially for public health and precision medicine (pharmacogenomics). Pharmacogenomics is one main challenge for the Orpailleur team as it considers a large panel of complex data ranging from biological to medical data, and various kinds of encoded domain knowledge ranging from texts to formal ontologies.

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 including 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 labeled graphs. Graph mining and Formal Concept Analysis methods play an important role in this application domain and can be used in an efficient and well-founded way [87].

Finally, research in agronomy is mainly based on cooperation between Inria and INRA. One research dimension is related to the characterization and the simulation of hedgerow structures in agricultural landscapes, based on Hilbert-Peano curves and Markov models [79]. Another research dimension is based on the mining of survey data for evaluating groundwater quality risks [86].

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. Remote Sensing

Remote sensing is an image acquisition paradigm in which various pieces of information from the surface of the earth are obtained by airborne or satellite imaging. In the recent years the quality and variety of sensors available on the market for such remote acquisitions has drastically improved, leaving a large number of unsolved or difficult data processing tasks on the hand of data scientists. In particular, members of the PANAMA project-team have focused on processing hyperspectral images, that are images collected at a large number of wavelengths. The tasks that have been addressed are mainly spectral unmixing, that aims at identifying the various materials on the earth surface with their relative abundances, and denoising, that aims at removing structured noise from remote acquisitions that can be polluted by e.g. sensor movement or clouds. The spectra measured at each pixel of an hyperspectral image are heavily dependent on surface features (material, slop, material mixture type) which makes these data extremely rich and interesting.

PERCEPTION Project-Team (section vide)

PERVASIVE Project-Team

4. Application Domains

4.1. Modelling of awareness and expertise from Eye Gaze and Emotion

Humans display awareness and expertise through a variety of non-verbal channels. It is increasingly possible to record and interpret such information with available technology. In the ANR CEEGE project, we have constructed an instrument for capturing and interpreting multimodal signals of humans engaged in solving challenging problems. Our instrument captures eye gaze, fixations, body postures, and facial expressions signals from humans engaged in interactive tasks on a touch screen.

An initial experiment with multi-modal observation of human experts engaged in solving problems in Chess revealed an unexpected observation of rapid changes in emotion as players attempt to solve challenging problems. In a scientific collaboration with the NeuroCognition group at the Univ Bielefeld, we have constructed a model to explain for chess experts that explains these unexpected results. This model has recently been tested in a second experiment with 22 chess players. Our results indicate that chess players associate emotions to chess chunks, and reactively use these associations to guide search in chunks for planning and problem solving. These results have recently been reported in a paper at the International Conference on Multimodal Interaction, and is the subject of the nearly completed doctoral dissertation of Thomas Guntz.

The results are currently being used in the construction of a student aware driver training device to be commercialized by SME company Sym2B financed by the SATT Linksium Project MAT: Monitoring Attention of Trainees, starting in Sept 2019. IN this project we will construct a training simulator for operation of busses and tramways.

4.2. Narrative Description of Kitchen Activities from Egocentric Video

We have developed and evaluated a system to construct situated, narrative descriptions of cooking activities including food preparation, place setting, cleaning and placing objects in storage areas. We are specifically interested in real-time, on-line techniques that recognize and interpret food types, food states and manipulation actions for transformation preparation of food. We are exploring techniques for detecting, modelling, and recognising a large vocabulary of actions and activities under different observational conditions, and describing these activities in a larger context.

A full understanding of human actions requires: recognising what action has been performed, predicting how it will affect the surrounding environment, explaining why this action has been performed, and who is performing it. Classic approaches to action recognition interpret a spatio-temporal pattern in a video sequence to tell what action has been performed, and perhaps how and where it was performed. A more complete understanding requires information about why the action was performed, and how it affects the environment. This face of understanding can be provided by explaining the action as part of a narrative.

Most work on recognition of cooking activities has concentrated on recognizing actions from the spatio-temporal patterns of hand motions. While some cooking activities may be directly recognized from motion, the resulting description is incomplete, as it does not describe the state of the ingredients, and how these have been transformed by cooking actions. A fuller description requires a description of how food ingredients have been transformed during the food preparation process.

We have addressed the automatic construction of cooking narratives by first detecting and locating ingredients and tools used in food preparation. We then recognize actions that involve transformations of ingredients, such as "slicing", and use these transformations to segment the video stream into visual events. We can then interpret detected events as a causal sequence of voluntary actions, confirmed by spatio-temporal transformation patterns, in order to automatically provide a narrative.

Our method is inspired by the intuition that object states are visually more apparent than actions from a still frame and thus provide information that is complementary to spatio-temporal action recognition. We define a state transition matrix that maps action labels into a pre-state and a post-state. We identify key frames, and use these to learn appearance models of objects and their states. For recognition, we use a modified form of VGG neural network trained via transfer learning with a specially constructed data set of images of food types and food sates. Manipulation actions are hypothesized from the state transition matrix and provide complementary evidence to spatio temporal action recognition.

4.3. Embedded Computer Vision for low-power Bolometric Imaging

In cooperation with Schneider Electric, we have developed techniques for embedded real time image analysis and tracking algorithms using Bolometric Imaging Sensors. Such sensors capture light in the far infrared and return an image where each pixel is a measurement of temperature in degree centigrade. We have designed an integrated low-cost sensor that combines an 80x80 pixel Bolometric imager with a low power micro-processor. The device provides real-time, embedded image processing for target detection and activity analysis where all sensing and interpretation is local. No images are recorded and only relevant information is about activity is communicated. The design of this system is under consideration for a patent, and thus has not been published. The software system has been registered with French APP , and is to be licensed to Schneider Electric for use a line of sensors for detecting falls and other activities for assisted living for seniors, as well as monitoring customer activities in commercial environments.

4.4. Recognizing and predicting routine activities in smart homes

Most research on smart home systems has concentrated on techniques for recognizing context. However, many categories of service require information about likely future context. We have developed an approach that uses dynamic Bayesian networks to predict future activity and context in a smart home. Our approach extends a state-of-the-art prediction model with three contributions. First, we include sensor data through aggregation nodes, instead of limiting ourselves only to higher level context data. Second, our method uses higher order relations between activities, so that past activities can have a more meaningful impact on prediction of future activities. Third, we use a latent node that estimates the cognitive state of the occupant.

4.5. User centred energy management

Participants: Amr Al-Zouhri Al-Yafi, Amine Awada, Patrick Reignier Partners: UMR G-SCOP, UMR LIG (Persuasive Interaction, IIHM), CEA Liten, PACTE, Vesta Systems and Elithis.

Work in this area explores techniques for a user centric energy management system, where user needs and tacit knowledge drive the search of solutions. These are calculated using a flexible energy model of the living areas. The system is personified by energy consultants with which building actors such as building owners, building managers, technical operators but also occupants, can interact in order to co-define energy strategies, benefiting of both assets: tacit knowledge of human actors, and measurement with computation capabilities of calculators. Putting actors in the loop, i.e. making energy not only visible but also controllable is the needed step before large deployment of energy management solutions. It is proposed to develop interactive energy consultants for all the actors, which are energy management aided systems embedding models in order to support the decision making processes. MIRROR (interactive monitoring), WHAT-IF (interactive quantitative simulation), EXPLAIN (interactive qualitative simulation), SUGGEST-AND-ADJUST (interactive management) and RECOMMEND (interactive diagnosis) functionalities will be developed.

4.6. E-Textile

Participant: Sabine Coquillart

Partner: LIMSI

Collaboration with the HAPCO team from LIMSI on e-textiles. A patent application has been filed related to this work:

- F. Bimbard, M. Bobin, M. Ammi, S. Coquillart "Procédé de conception d'un capteur de flexion textile piézorésistif à partir de fils fonctionnels", Patent Application, 2017.

4.7. Interaction with Pervasive Media

Participants: Sabine Coquillart, Jingtao Chen

Partners: Inria GRA, GIPSA, G-SCOP

Pseudo-haptic feedback is a technique aiming to simulate haptic sensations without active haptic feedback devices. Pseudo-haptic techniques have been used to simulate various haptic feedbacks such as stiffness, torques, and mass. In the framework of Jingtao Chen PhD thesis, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the EMG signals during a pseudo-haptic task. A stiffness discrimination task similar to the one published in Lecuyer's PhD thesis has been chosen. The experimental set-up has been developed, as well as the software controlling the experiment. Pre-tests are under way. They will be followed by the tests with subjects.

PETRUS Project-Team

4. Application Domains

4.1. Personal cloud, home care, IoT, sensing, surveys

As stated in the software section, the Petrus research strategy aims at materializing its scientific contributions in an advanced hardware/software platform with the expectation to produce a real societal impact. Hence, our software activity is structured around a common Secure Personal Cloud platform rather than several isolated demonstrators. This platform will serve as the foundation to develop a few emblematic applications. Several privacy-preserving applications can actually be targeted by a Personal Cloud platform, like: (i) smart disclosure applications allowing the individual to recover her personal data from external sources (e.g., bank, online shopping activity, insurance, etc.), integrate them and cross them to perform personal big data tasks (e.g., to improve her budget management) ; (ii) management of personal medical records for care coordination and well-being improvement; (iii) privacy-aware data management for the IoT (e.g., in sensors, quantified-self devices, smart meters); (iv) community-based sensing and community data sharing; (v) privacy-preserving studies (e.g., cohorts, public surveys, privacy-preserving data publishing). Such applications overlap with all the research axes described above but each of them also presents its own specificities. For instance, the smart disclosure applications will focus primarily on sharing models and enforcement, the IoT applications require to look with priority at the embedded data management and sustainability issues, while community-based sensing and privacy-preserving studies demand to study secure and efficient global query processing. Among these applications domains, one is already receiving a particular attention from our team. Indeed, we gained a strong expertise in the management and protection of healthcare data through our past DMSP (Dossier Medico-Social Partagé) experiment in the field. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. We are currently deploying 10000 boxes equipped with PlugDB in the context of the DomYcile project. In this context, we are currently setting up an Inria Innovation Lab with the Hippocad company to industrialize this platform and deploy it at large scale (see Section the bilateral contract OwnCare II-Lab).

POTIOC Project-Team

4. Application Domains

4.1. Education

Education is at the core of the motivations of the Potioc group. Indeed, we are convinced that the approaches we investigate—which target motivation, curiosity, pleasure of use and high level of interactivity—may serve education purposes. To this end, we collaborate with experts in Educational Sciences and teachers for exploring new interactive systems that enhance learning processes. We are currently investigating the fields of astronomy, optics, and neurosciences. We have also worked with special education centres for the blind on accessible augmented reality prototypes. Currently, we collaborate with teachers to enhance collaborative work for K-12 pupils. In the future, we will continue exploring new interactive approaches dedicated to education, in various fields. Popularization of Science is also a key domain for Potioc. Focusing on this subject allows us to get inspiration for the development of new interactive approaches.

4.2. Art

Art, which is strongly linked with emotions and user experiences, is also a target area for Potioc. We believe that the work conducted in Potioc may be beneficial for creation from the artist point of view, and it may open new interactive experiences from the audience point of view. As an example, we have worked with colleagues who are specialists in digital music, and with musicians. We have also worked with jugglers and we are currently working with a scenographer with the goal of enhancing interactivity of physical mockups and improve user experience.

4.3. Entertainment

Similarly, entertainment is a domain where our work may have an impact. We notably explored BCI-based gaming and non-medical applications of BCI, as well as mobile Augmented Reality games. Once again, we believe that our approaches that merge the physical and the virtual world may enhance the user experience. Exploring such a domain will raise numerous scientific and technological questions.

4.4. Well-being

Finally, well-being is a domain where the work of Potioc can have an impact. We have notably shown that spatial augmented reality and tangible interaction may favor mindfulness activities, which have been shown to be beneficial for well-being. More generally, we explore *introspectibles* objects, which are tangible and augmented objects that are connected to physiological signals and that foster introspection. We explore these directions for general public, including people with special needs.

RAINBOW Project-Team (section vide)

RITS Project-Team

4. Application Domains

4.1. Introduction

While the preceding section focused on methodology, in connection with automated guided vehicles, it should be stressed that the evolution of the problems which we deal with remains often guided by the technological developments. We enumerate three fields of application whose relative importance varies with time and which have strong mutual dependencies: driving assistance, cars available in self-service mode and fully automated vehicles (cybercars).

4.2. Driving assistance

Several techniques will soon help drivers. One of the first immediate goal is to improve security by alerting the driver when some potentially dangerous or dangerous situations arise, i.e. collision warning systems or lane tracking could help a bus driver and surrounding vehicle drivers to more efficiently operate their vehicles. Human factors issues could be addressed to control the driver workload based on additional information processing requirements. Another issue is to optimize individual journeys. This means developing software for calculating optimal (for the user or for the community) paths. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component in account.

4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs. However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these “lines of desire” or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time. To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost. Electric or gas vehicles available in self-service, as in the Praxitèle system, bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

- ease empty car moves to better distribute them;
- better use of information systems inboard and on ground;
- better integrate this system in the global transportation system.

These systems are now operating. 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 to 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. In order to rely on a professional and maintained solution, we use the RTMaps SDK development platform for our developments and demonstrations. 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 EVA 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. Deep Semantic Analysis

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.

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.

SIROCCO Project-Team

4. Application Domains

4.1. Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

4.2. Compression of emerging imaging modalities

Compression of visual content 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 imaging modalities, e.g. high dynamic range (HDR) images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

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 (FVV). 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

Free-viewpoint Television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

Broadband fixed and mobile access networks with different radio access technologies 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).

4.4. Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of “consuming” visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. 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. Our activities around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

Stars Project-Team (section vide)

THOTH Project-Team

4. Application Domains

4.1. Visual applications

Any solution to automatically understanding images and videos on a semantic level will have an immediate impact on a wide range of applications. For example:

- Semantic-level image and video access is highly relevant for visual search on the Web, in professional archives and personal collections.
- Visual data organization is applicable to organizing family photo and video albums as well as to large-scale information retrieval.
- Visual object recognition has potential applications ranging from surveillance, service robotics for assistance in day-to-day activities as well as the medical domain.
- Action recognition is highly relevant to visual surveillance, assisted driving and video access.
- Real-time scene understanding is relevant for human interaction through devices such as HoloLens, Oculus Rift.

4.2. Pluri-disciplinary research

Machine learning is intrinsically pluri-disciplinary. By developing large-scale machine learning models and algorithms for processing data, the Thoth team became naturally involved in pluri-disciplinary collaborations that go beyond visual modelling. In particular,

- extensions of unsupervised learning techniques originally developed for modelling the statistics of natural images have been deployed in neuro-imaging for fMRI data with the collaboration of the Parietal team from Inria.
- similarly, deep convolutional data representations, also originally developed for visual data, have been successfully extended to the processing of biological sequences, with collaborators from bio-informatics.
- Thoth also collaborates with experts in natural language and text processing, for applications where visual modalities need to be combined with text data.

TITANE Project-Team

4. Application Domains

4.1. Domain 1

In addition to tackling enduring 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 deliverable of our research will be 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 first 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. Consolidation being best achieved with the help of an engineer, we will search for additional funding. 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.

Our second ambition is to increase the research momentum of companies through advising Cifre Ph.D. theses and postdoctoral fellows on topics that match our research program.

TYREX Project-Team

4. Application Domains

4.1. Querying Large Graphs

Increasingly large amounts of graph-structured data become available. The methods we develop apply for the efficient evaluation of graph queries over large — and potentially distributed — graphs. In particular, we consider the SPARQL query language, which is the standard language for querying graphs structured in the Resource Description Format (RDF). We also consider other increasingly popular graph query languages such as Cypher queries for extracting information from property graphs.

We compile graph queries into lower-level distributed primitives found in big data frameworks such as Apache Spark, Flink, etc. Applications of graph querying are ubiquitous and include: large knowledge bases, social networks, road networks, trust networks and fraud detection for cryptocurrencies, publications graphs, web graphs, recommenders, etc.

4.2. Predictive Analytics for Healthcare

One major expectation of data science in healthcare is the ability to leverage on digitized health information and computer systems to better apprehend and improve care. The availability of large amounts of clinical data and in particular electronic health records opens the way to the development of quantitative models for patients that can be used to predict health status, as well as to help prevent disease and adverse effects.

In collaboration with the CHU Grenoble, we explore solutions to the problem of predicting important clinical outcomes such as patient mortality, based on clinical data. This raises many challenges including dealing with the very high number of potential predictor variables and very resource-consuming data preparation stages.

VALDA Project-Team

4. Application Domains

4.1. Personal Information Management Systems

We recall that Valda's focus is on human-centric data, i.e., data produced by humans, explicitly or implicitly, or more generally containing information about humans. Quite naturally, we have used as a privileged application area to validate Valda's results that of personal information management systems (Pims for short) [33].

A Pims is a system that allows a user to integrate her own data, e.g., emails and other kinds of messages, calendar, contacts, web search, social network, travel information, work projects, etc. Such information is commonly spread across different services. The goal is to give back to a user the control on her information, allowing her to formulate queries such as "What kind of interaction did I have recently with Alice B.?", "Where were my last ten business trips, and who helped me plan them?". The system has to orchestrate queries to the various services (which means knowing the existence of these services, and how to interact with them), integrate information from them (which means having data models for this information and its representation in the services), e.g., align a GPS location of the user to a business address or place mentioned in an email, or an event in a calendar to some event in a Web search. This information must be accessed intensionally: for instance, costly information extraction tools should only be run on emails which seem relevant, perhaps identified by a less costly cursory analysis (this means, in turn, obtaining a cost model for access to the different services). Impacted people can be found by examining events in the user's calendar and determining who is likely to attend them, perhaps based on email exchanges or former events' participant lists. Of course, uncertainty has to be maintained along the entire process, and provenance information is needed to explain query results to the user (e.g., indicate which meetings and trips are relevant to each person of the output). Knowledge about services, their data models, their costs, need either to be provided by the system designer, or to be automatically learned from interaction with these services, as in [68].

One motivation for that choice is that Pims concentrate many of the problems we intend to investigate: heterogeneity (various sources, each with a different structure), massive distribution (information spread out over the Web, in numerous sources), rapid evolution (new data regularly added), intensionality (knowledge from Wikidata, OpenStreetMap...), confidentiality and security (mostly private data), and uncertainty (very variable quality). Though the data is distributed, its size is relatively modest; other applications may be considered for works focusing on processing data at large scale, which is a potential research direction within Valda, though not our main focus. Another strong motivation for the choice of Pims as application domain is the importance of this application from a societal viewpoint.

A Pims is essentially a system built on top of a user's *personal knowledge base*; such knowledge bases are reminiscent of those found in the Semantic Web, e.g., linked open data. Some issues, such as ontology alignment [71] exist in both scenarios. However, there are some fundamental differences in building personal knowledge bases vs collecting information from the Semantic Web: first, the scope is quite smaller, as one is only interested in knowledge related to a given individual; second, a small proportion of the data is already present in the form of semantic information, most needs to be extracted and annotated through appropriate wrappers and enrichers; third, though the linked open data is meant to be read-only, the only update possible to a user being adding new triples, a personal knowledge base is very much something that a user needs to be able to edit, and propagating updates from the knowledge base to original data sources is a challenge in itself.

4.2. Web Data

The choice of Pims is not exclusive. We also consider other application areas as well. In particular, we have worked in the past and have a strong expertise on Web data [38] in a broad sense: semi-structured, structured, or unstructured content extracted from Web databases [68]; knowledge bases from the Semantic Web [71];

social networks [66]; Web archives and Web crawls [54]; Web applications and deep Web databases [47]; crowdsourcing platforms [41]. We intend to continue using Web data as a natural application domain for the research within Valda when relevant. For instance [45], deep Web databases are a natural application scenario for intensional data management issues: determining if a deep Web database contains some information requires optimizing the number of costly requests to that database.

A common aspect of both personal information and Web data is that their exploitation raises ethical considerations. Thus, a user needs to remain fully in control of the usage that is made of her personal information; a search engine or recommender system that ranks Web content for display to a specific user needs to do so in an unbiased, justifiable, manner. These ethical constraints sometimes forbid some technically solutions that may be technically useful, such as sharing a model learned from the personal data of a user to another user, or using blackboxes to rank query result. We fully intend to consider these ethical considerations within Valda. One of the main goals of a Pims is indeed to empower the user with a full control on the use of this data.

WILLOW 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. Social Semantic Web

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 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: SMILK Joint Laboratory, Corese, DBpedia.fr.

4.3. Assisting Web-based Epistemic Communities

In parallel with linked open data on the Web, social Web applications also spread virally (e.g. Facebook growing toward 1.5 billion users) first giving the Web back its status of a social read-write media and then putting it back on track 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.

By reasoning on the Linked Data and the semantics of the schemas used to represent social structures and Web resources, we provide applications supporting communities of practice and interest and fostering their interactions in many different contexts (e-learning, business intelligence, technical watch, etc.).

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: EduMICS, OCKTOPUS, Vigiglobe, Educlever, Gayatech.

4.4. Linked Data for a Web of Diversity

We intend to build on our results on explanations (provenance, traceability, justifications) and to continue our work on opinions and arguments mining toward the global analysis of controversies and online debates. One result would be to provide new search results encompassing the diversity of viewpoints and providing indicators supporting opinion and decision making and ultimately a Web of trust. Trust indicators may require collaborations with teams specialized in data certification, cryptography, signature, security services and protocols, etc. This will raise the specific problem of interaction design for security and privacy. In addition, from the point of view of the content, this requires to foster the publication and coexistence of heterogeneous data with different points of views and conceptualizations of the world. We intend to pursue the extension of formalisms to allow different representations of the world to co-exist and be linked and we will pay special attention to the cultural domain and the digital humanities. Examples of collaboration and applied projects: Zoomathia, Seempad, SMILK,

4.5. Artificial Web Intelligence

We intend to build on our experience in artificial intelligence (knowledge representation, reasoning) and distributed artificial intelligence (multi-agent systems - MAS) to enrich formalisms and propose alternative types of reasoning (graph-based operations, reasoning with uncertainty, inductive reasoning, non-monotonic, etc.) and alternative architectures for linked data with adequate changes and extensions required by the open nature of the Web. There is a clear renewed interest in AI for the Web in general and for Web intelligence in particular. Moreover distributed AI and MAS provide both new architectures and new simulation platforms for the Web. At the macro level, the evolution accelerated with HTML5 toward Web pages as full applications and direct Page2Page communication between browser clearly is a new area for MAS and P2P architectures. Interesting scenarios include the support of a strong decentralization of the Web and its resilience to degraded technical conditions (downscaling the Web), allowing pages to connect in a decentralized way, forming a neutral space, and possibly going offline and online again in erratic ways. At the micro level, one can imagine the place RDF and SPARQL could take as data model and programming model in the virtual machines of these new Web pages and, of course, in the Web servers. RDF is also used to serialize and encapsulate other languages and becomes a pivot language in linking very different applications and aspects of applications. Example of collaboration and applied projects: MoreWAIS, Corese, Vigiglobe collaboration.

4.6. Human-Data Interaction (HDI) on the Web

We need more interaction design tools and methods for linked data access and contribution. We intend to extend our work on exploratory search coupling it with visual analytics to assist sense making. It could be a continuation of the Gephi extension that we built targeting more support for non experts to access and analyze data on a topic or an issue of their choice. More generally speaking SPARQL is inappropriate for common users and we need to support a larger variety of interaction means with linked data. We also believe linked data and natural language processing (NLP) have to be strongly integrated to support natural language based interactions. Linked Open Data (LOD) for NLP, NLP for LOD and Natural Dialog Processing for querying, extracting and asserting data on the Web is a priority to democratize its use. Micro accesses and micro contributions are important to ensure public participation and also call for customized interfaces and thus for methods and tools to generate these interfaces. In addition, the user profiles are being enriched now with new data about the user such as her current mental and physical state, the emotion she just expressed or her cognitive performances. Taking into account this information to improve the interactions, change the behavior of the system and adapt the interface is a promising direction. And these human-data interaction

means should also be available for “small data”, helping the user to manage her personal information and to link it to public or collective one, maintaining her personal and private perspective as a personal Web of data. Finally, the continuous knowledge extractions, updates and flows add the additional problem of representing, storing, querying and interacting with dynamic data. Examples of collaboration and applied projects: QAKIS, Sychonext collaboration, ALOOF, DiscoveryHub, WASABI, MoreWAIS.

Web-augmented interactions with the world: The Web continues to augment our perception and interaction with reality. In particular, Linked Open Data enable new augmented reality applications by providing data sources on almost any topic. The current enthusiasm for the Web of Things, where every object has a corresponding Web resource, requires evolutions of our vision and use of the Web architecture. This vision requires new techniques as the ones mentioned above to support local search and contextual access to local resources but also new methods and tools to design Web-based human devices interactions, accessibility, etc. These new usages are placing new requirements on the Web Architecture in general and on the semantic Web models and algorithms in particular to handle new types of linked data. They should support implicit requests considering the user context as a permanent query. They should also simplify our interactions with devices around us jointly using our personal preferences and public common knowledge to focus the interaction on the vital minimum that cannot be derived in another way. For instance the access to the Web of data for a robot can completely change the quality of the interactions it can offer. Again, these interactions and the data they require raise problems of security and privacy. Examples of collaboration and applied projects: ALOOF, AZKAR, MoreWAIS.

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 and CIRAD. However, we are also addressing other scientific domains (e.g. astronomy, oil extraction, music processing) through our international collaborations.

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 has hundreds of TB. 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** 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 performance in sports. This can be done thanks to sensors, communicating devices or even connected glasses. 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 with 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.
- **Biological data integration and analysis.**

Biology and its applications, from medicine to agronomy and ecology, are now producing massive data, which is revolutionizing the way life scientists work. For instance, using plant phenotyping platforms such as PhenoDyn and PhenoArch at INRA Montpellier, quantitative genetic methods allow to identify genes involved in phenotypic variation in response to environmental conditions. These methods produce large amounts of data at different time intervals (minutes to months), at different sites and at different scales ranging from small tissue samples to the entire plant until whole plant population. Analyzing such big data creates new challenges for data management and data integration.

- **Audio heritage preservation.**

Since the end of the 19th century, France has commissioned ethnologists to record the world's immaterial audio heritage. This results in datasets of dozens of thousands of audio recordings from all countries and more than 1200 ethnies. Today, this data is gathered under the name of 'Archives du CNRS — Musée de l'Homme' and is handled by the CREM (Centre de Recherche en Ethno-Musicologie). Professional scientists in digital humanities are accessing this data daily for their investigations, and several important challenges arise to ease their work. The KAMoulox project, lead by A. Liutkus, targets at offering online processing tools for the scientists to automatically restore this old material on demand.

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.