Activity Report 2014

Project-Team AOSTE

Models and methods of analysis and optimization for systems with real-time and embedding constraints

IN COLLABORATION WITH: Laboratoire informatique, signaux systèmes de Sophia Antipolis (I3S)
## Table of contents

1. Members ........................................................................................................... 1
2. Overall Objectives ........................................................................................... 2
3. Research Program ............................................................................................. 2
   3.1. Models of Computation and Communication (MoCCs) .......................... 2
       3.1.1. K-periodic static scheduling and routing in Process Networks ...... 3
       3.1.2. Endochrony and GALS implementation of conflict-free polychronous programs 3
   3.2. Logical Time in Model-Driven Embedded System Design ....................... 3
   3.3. The AAA (Algorithm-Architecture Adequation) methodology and Real-Time Scheduling .......................... 4
       3.3.1. Algorithm-Architecture Adequation ............................................ 4
       3.3.2. Distributed Real-Time Scheduling and Optimization ...................... 5
4. Application Domains ......................................................................................... 6
   4.1. Multicore System-on-Chip design ............................................................. 6
   4.2. Automotive and avionic embedded systems .............................................. 7
5. New Software and Platforms ............................................................................ 7
   5.1. TimeSquare ............................................................................................... 7
   5.2. K-Passa ...................................................................................................... 8
   5.3. SynDEx ...................................................................................................... 8
   5.4. Lopht ......................................................................................................... 9
   5.5. SAS .......................................................................................................... 9
6. New Results ....................................................................................................... 10
   6.1. Languages, Models of Computation and Metamodeling using logical clock constraints ........................................... 10
   6.2. Experiments with Architecture and Application modeling ...................... 10
   6.3. Multiview modeling with performance and power aspects ....................... 11
   6.4. Heterogeneous Languages Coordination with Concurrency and Time ........ 11
   6.5. Performance study of Massively Parallel Processor Array (MPPA) SoC architecture ........................................... 11
   6.7. Uniprocessor Real-Time Scheduling .......................................................... 12
   6.8. Multiprocessor Real-Time Scheduling ........................................................ 12
       6.8.1. Multiprocessor Partitioned Scheduling with Exact Preemption Cost ... 13
       6.8.2. Multiprocessor Parallel Directed Acyclic Graph (DAG) scheduling ... 13
       6.8.3. Gateway with Modeling Languages for Certified Code Generation ... 13
       6.8.4. SynDEx updates .......................................................................... 13
   6.10. Off-line (static) mapping and WCET analysis of real-time applications onto NoC-based many-cores .......................................................... 14
   6.11. Real-time scheduling and code generation for time-triggered platforms ....... 15
7. Bilateral Contracts and Grants with Industry ..................................................... 15
   7.1. Kontron PhD CIFRE thesis ....................................................................... 15
   7.2. Airbus PhD CIFRE thesis ......................................................................... 15
   7.3. Astrium/CNES PostDoc ......................................................................... 16
8. Partnerships and Cooperations ......................................................................... 16
   8.1. Regional Initiatives .................................................................................... 16
   8.2. National Initiatives .................................................................................... 16
       8.2.1. ANR ............................................................................................. 16
           8.2.1.1. HOPE .............................................................................. 16
           8.2.1.2. GeMoC ............................................................................ 16
       8.2.2. FUI .............................................................................................. 17
           8.2.2.1. FUI P ............................................................................... 17
           8.2.2.2. FUI CLISTINE ............................................................... 17
8.2.3. Investissements d’Avenir
   8.2.3.1. DEPARTS
   8.2.3.2. CLARITY
   8.2.3.3. Capacites
8.3. European Initiatives
8.4. International Initiatives
   8.4.1. Inria International Labs
   8.4.2. Inria Associate Teams
8.5. International Research Visitors
   8.5.1. Visits of International Scientists
   8.5.2. Visits to International Teams
9. Dissemination
   9.1. Promoting Scientific Activities
      9.1.1. Scientific events organisation
      9.1.2. Scientific events selection
         9.1.2.1. responsable of the conference program committee
         9.1.2.2. member of conference program committee
      9.1.3. Journal
9.2. Teaching - Supervision - Juries
   9.2.1. Teaching
   9.2.2. Supervision
   9.2.3. Juries
10. Bibliography
Project-Team AOSTE

Keywords: Embedded Systems, Synchronous Languages, Model-Driven Engineering, Real-Time, Scheduling, Concurrency, Models Of Computation

Aoste is a joint team with the University of Nice/Sophia-Antipolis (UNS) and the UMR CNRS 7270: I3S. It is also co-located between the two Inria centers of Sophia-Antipolis and Rocquencourt.

Creation of the Project-Team: 2004 July 01.

1. Members

Research Scientists
Robert de Simone [Team leader, Inria, Senior Researcher, Sophia Antipolis - Méditerranée, HdR]
Yves Sorel [Team leader, Inria, Senior Researcher, Paris - Rocquencourt]
Liliana Cucu-Grosjean [Inria, Researcher, Paris - Rocquencourt, HdR]
Dumitru Potop Butucaru [Inria, Researcher, Paris - Rocquencourt]
Robert Davis [Inria International Chair, from October 2014, Paris - Rocquencourt]

Faculty Members
Julien Deantoni [UNS, Associate Professor, Sophia Antipolis - Méditerranée]
Frédéric Mallet [UNS, Professor, on Inria delegation from Sept, Sophia Antipolis - Méditerranée, HdR]
Marie Agnès Peraldi Frati [UNS, Associate Professor, Sophia Antipolis - Méditerranée]
Sid Touati [UNS, Professor, Sophia Antipolis - Méditerranée, HdR]
Laurent George [U. Paris 12 Associate Professor, external collaborator, Paris - Rocquencourt, HdR]

Engineers
Abderraouf Benyahia [Inria, FUI P project, Paris - Rocquencourt]
Daniel de Rauglaudre [Inria, Paris - Rocquencourt]
Adriana Gogonel [Inria, FP7 PROXIMA project, Paris - Rocquencourt]
Nicolas Chleq [Dream ADT, until November 2014, Sophia Antipolis - Méditerranée]
Luc Hogie [CNRS Research Engineer (part time), Sophia Antipolis - Méditerranée]
Codé Lo [Inria, FP7 PROXIMA project, Paris - Rocquencourt]
Walid Talaboumla [Inria, LEOC CAPACITES, from November 2014, Paris - Rocquencourt]
Meriem Zidouni [Inria, until February 2014, Paris - Rocquencourt]
Cristian Maxim [Inria, FP7 PROXIMA project, until March 2014, Paris - Rocquencourt]

PhD Students
Mohamed Bergach [CIFRE funding (Kontron), Sophia Antipolis - Méditerranée]
Yann Bondeu [Inria, Clarity project, from Dec, Sophia Antipolis - Méditerranée]
Thomas Carle [Inria, until October 2014, Paris - Rocquencourt]
Manel Djemal [Inria, until May 2014, Paris - Rocquencourt]
Amani Khecharem [Inria, ANR HOPE project funding, Sophia Antipolis - Méditerranée]
Émilien Kofman [UNS, UCN Labex, jointly with UMR LEAT, Sophia Antipolis - Méditerranée]
Cristian Maxim [AIRBUS contract, starting April 2014, Paris - Rocquencourt]
Falou Ndoye [Inria, until February 2014, Paris - Rocquencourt]
Amin Oueslati [Inria, FUI Clistine project funding, Sophia Antipolis - Méditerranée]
Salah Saidi [IFP Energies nouvelles, from Dec 2014, Paris - Rocquencourt]
Matias Vara Larsen [CNRS, ANR GeMoC funding, Sophia Antipolis - Méditerranée]

Post-Doctoral Fellow
Raúl Gorcitz [CNES, Paris - Rocquencourt]

Visiting Scientist
Qingguo Xu [Associate Professor, Shanghai U., from Jul 2014, Sophia Antipolis - Méditerranée]
2. Overall Objectives

2.1. Embedded System Design

Typical embedded software applications display a mix of multimedia signal/data processing with modal interfaces, resulting in heterogeneous concurrent data-flow streaming models, and often stringent real-time constraints. Similarly, embedded architectural platforms are becoming increasingly parallel, with dedicated hardware accelerators and manycore processors. The optimized compilation of such kinds of applications onto such execution platforms involves complex mapping issues, both in terms of spatial distribution and in terms of temporal scheduling. Currently, it is far from being a fully automatic compilation process as in the case of commodity PC applications. Models are thus needed, both as formal mathematical objects from theoretical computer science to provide foundations for embedded system design, and also as engineering models to support an effective design flow.

Our general approach is directly inspired from the theories of synchronous languages, process networks, and of real-time distributed scheduling. We insist on the introduction of logical time as functional design ingredient to be explicitly considered as first-class modeling element of systems. Logical time is based on logical clocks, where such a clock can be defined as any meaningful sequence of event occurrences, usually meant as activation/trigging conditions for actions and operations in the systems. So logical time can be multiform, a global partial order built from local total orders of clocks. In the course of the design flow time refinement takes place, as decision are made towards placement and timing of various tasks and operations. This solves in part the constraints between clocks, committing to schedule and placement decisions. The final version should be totally ordered, and then subject to physical timing verification as to physical constraints.

The general (logical) Time Model has been standardized as part of the OMG profile for Modeling and Analysis of Real-Time Embedded systems (MARTE).

Work on polychronous formalisms (descending from Esterel), on a Clock Constraint Specification Language handling logical time, on Application-Architecture Adequation approach and real-time scheduling results has been progressed over the years, resulting in software environments such as SynDex or TimeSquare.

3. Research Program

3.1. Models of Computation and Communication (MoCCs)

Participants: Julien Deantoni, Robert de Simone, Frédéric Mallet, Jean-Vivien Millo, Dumitru Potop Butucaru.

Esterel, SyncCharts, synchronous formalisms, Process Networks, Marked Graphs, Kahn networks, compilation, synthesis, formal verification, optimization, allocation, refinement, scheduling.
Formal Models of Computation form the basis of our approach to Embedded System Design. Because of the growing importance of communication handling, it is now associated with the name, MoCC in short. The appeal of MoCCs comes from the fact that they combine features of mathematical models (formal analysis, transformation, and verification) with those of executable specifications (close to code level, simulation, and implementation). Examples of MoCCs in our case are mainly synchronous reactive formalisms and dataflow process networks. Various extensions or specific restrictions enforce respectively greater expressivity or more focused decidable analysis results.

DataFlow Process Networks and Synchronous Reactive Languages such as ESTEREL/SYNCHRONY and SIGNAL/POLYCHRONY [53], [54], [48], [15], [4], [13] share one main characteristics: they are specified in a self-timed or loosely timed fashion, in the asynchronous data-flow style. But formal criteria in their semantics ensure that, under good correctness conditions, a sound synchronous interpretation can be provided, in which all treatments (computations, signaling communications) are precisely temporally mapped. This is referred to as clock calculus in synchronous reactive systems, and leads to a large body of theoretical studies and deep results in the case of DataFlow Process Networks [49], [47] (consider SDF balance equations for instance [56]).

As a result, explicit schedules become an important ingredient of design, which ultimately can be considered and handled by the designer him/herself. In practice such schedules are sought to optimize other parts of the design, mainly buffering queues: production and consumption of data can be regulated in their relative speeds. This was specially taken into account in the recent theories of Latency-Insensitive Design [50], or N-synchronous processes [51], with some of our contributions [6].

Explicit schedule patterns should be pictured in the framework of low-power distributed mapping of embedded applications onto manycore architectures, where they could play an important role as theoretical formal models on which to compute and optimize allocations and performances. We describe below two lines of research in this direction. Striking in these techniques is the fact that they include time and timing as integral parts of early functional design. But this original time is logical, multiform, and only partially ordering the various functional computations and communications. This approach was radically generalized in our team to a methodology for logical time based design, described next (see 3.2).

3.1. K-periodic static scheduling and routing in Process Networks

In the recent years we focused on the algorithm treatments of ultimately k-periodic schedule regimes, which are the class of schedules obtained by many of the theories described above. An important breakthrough occurred when realizing that the type of ultimately periodic binary words that were used for reporting static scheduling results could also be employed to record a completely distinct notion of ultimately k-periodic route switching patterns, and furthermore that commonalities of representation could ease combine them together. A new model, by the name of K-periodical Routed marked Graphs (KRG) was introduced, and extensively studied for algebraic and algorithmic properties [5].

The computations of optimized static schedules and other optimal buffering configurations in the context of latency-insensitive design led to the K-Passa software tool development 5.2.

3.1.2. Endochrony and GALS implementation of conflict-free polychronous programs

The possibility of exploring various schedulings for a given application comes from the fact that some behaviors are truly concurrent, and mutually conflict-free (so they can be executed independently, with any choice of ordering). Discovering potential asynchronous inside synchronous reactive specifications then becomes something highly desirable. It can benefit to potential distributed implementation, where signal communications are restricted to a minimum, as they usually incur loss in performance and higher power consumption. This general line of research has come to be known as Endochrony, with some of our contributions [11].

3.2. Logical Time in Model-Driven Embedded System Design

Participants: Julien Deantoni, Frédéric Mallet, Marie Agnès Peraldi Frati, Robert de Simone.
Starting from specific needs and opportunities for formal design of embedded systems as learned from our work on MoCCs (see 3.1), we developed a Logical Time Model as part of the official OMG UML profile MARTE for Modeling and Analysis of Real-Time Embedded systems. With this model is associated a Clock Constraint Specification Language (CCSL), which allows to provide loose or strict logical time constraints between design ingredients, be they computations, communications, or any kind of events whose repetitions can be conceived as generating a logical conceptual clock (or activation condition). The definition of CCSL is provided in [1].

Our vision is that many (if not all) of the timing constraints generally expressed as physical prescriptions in real-time embedded design (such as periodicity, sporadicity) could be expressed in a logical setting, while actually many physical timing values are still unknown or unspecified at this stage. On the other hand, our logical view may express much more, such as loosely stated timing relations based on partial orderings or partial constraints.

So far we have used CCSL to express important phenomena as present in several formalisms: AADL (used in avionics domain), EAST-ADL2 (proposed for the AutoSar automotive electronic design approach), IP-Xact (for System-on-Chip (SoC) design). The difference here comes from the fact that these formalisms were formerly describing such issues in informal terms, while CCSL provides a dedicated formal mathematical notation. Close connections with synchronous and polychronous languages, especially Signal, were also established; so was the ability of CCSL to model dataflow process network static scheduling.

In principle the MARTE profile and its Logical Time Model can be used with any UML editor supporting profiles. In practice we focused on the PYPYRUS open-source editor, mainly from CEA LIST. We developed under Eclipse the TIMESQUARE solver and emulator for CCSL constraints (see 5.1), with its own graphical interface, as a stand-alone software module, while strongly coupled with MARTE and Papyrus.

While CCSL constraints may be introduced as part of the intended functionality, some may also be extracted from requirements imposed either from real-time user demands, or from the resource limitations and features from the intended execution platform. Sophisticated detailed descriptions of platform architectures are allowed using MARTE, as well as formal allocations of application operations (computations and communications) onto platform resources (processors and interconnects). This is of course of great value at a time where embedded architectures are becoming more and more heterogeneous and parallel or distributed, so that application mapping in terms of spatial allocation and temporal scheduling becomes harder and harder. This approach is extensively supported by the MARTE profile and its various models. As such it originates from the Application-Architecture-Adequation (AAA) methodology, first proposed by Yves Sorel, member of Aoste. AAA aims at specific distributed real-time algorithmic methods, described next in 3.3.

Of course, while logical time in design is promoted here, and our works show how many current notions used in real-time and embedded systems synthesis can naturally be phrased in this model, there will be in the end a phase of validation of the logical time assumptions (as is the case in synchronous circuits and SoC design with timing closure issues). This validation is usually conducted from Worst-Case Execution Time (WCET) analysis on individual components, which are then used in further analysis techniques to establish the validity of logical time assumptions (as partial constraints) asserted during the design.

### 3.3. The AAA (Algorithm-Architecture Adequation) methodology and Real-Time Scheduling

**Participants:** Laurent George, Dumitru Potop Butucaru, Yves Sorel.

Note: The AAA methodology and the SynDEx environment are fully described at [http://www.syndex.org/](http://www.syndex.org/), together with relevant publications.

#### 3.3.1. Algorithm-Architecture Adequation

The AAA methodology relies on distributed real-time scheduling and relevant optimization to connect an Algorithm/Application model to an Architectural one. We now describe its premises and benefits.
The Algorithm model is an extension of the well-known data-flow model from Dennis [52]. It is a directed acyclic hyper-graph (DAG) that we call “conditioned factorized data dependence graph”, whose vertices are “operations” and hyper-edges are directed “data or control dependences” between operations. The data dependences define a partial order on the operations execution. The basic data-flow model was extended in three directions: first infinite (resp. finite) repetition of a sub-graph pattern in order to specify the reactive aspect of real-time systems (resp. in order to specify the finite repetition of a sub-graph consuming different data similar to a loop in imperative languages), second “state” when data dependences are necessary between different infinite repetitions of the sub-graph pattern introducing cycles which must be avoided by introducing specific vertices called “delays” (similar to $z^{-n}$ in automatic control), third “conditioning” of an operation by a control dependence similar to conditional control structure in imperative languages, allowing the execution of alternative subgraphs. Delays combined with conditioning allow the programmer to specify automata necessary for describing “mode changes”.

The Architecture model is a directed graph, whose vertices are of two types: “processor” (one sequencer of operations and possibly several sequencers of communications) and “medium” (support of communications), and whose edges are directed connections.

The resulting implementation model [9] is obtained by an external compositional law, for which the architecture graph operates on the algorithm graph. Thus, the result of such compositional law is an algorithm graph, “architecture-aware”, corresponding to refinements of the initial algorithm graph, by computing spatial (distribution) and timing (scheduling) allocations of the operations onto the architecture graph resources. In that context “Adequation” refers to some search amongst the solution space of resulting algorithm graphs, labelled by timing characteristics, for one algorithm graph which verifies timing constraints and optimizes some criteria, usually the total execution time and the number of computing resources (but other criteria may exist).

The next section describes distributed real-time schedulability analysis and optimization techniques for that purpose.

### 3.3.2. Distributed Real-Time Scheduling and Optimization

We address two main issues: uniprocessor and multiprocessor real-time scheduling where constraints must mandatorily be met, otherwise dramatic consequences may occur (hard real-time) and where resources must be minimized because of embedded features.

In the case of uniprocessor real-time scheduling, besides the classical deadline constraint, often equal to a period, we take into consideration dependences between tasks and several, latencies. The latter are complex related “end-to-end” constraints. Dealing with multiple real-time constraints raises the complexity of the scheduling problems. Moreover, because the preemption leads, at least, to a waste of resources due to its approximation in the WCET (Worst Execution Time) of every task, as proposed by Liu and Leyland [57], we first studied non-preemptive real-time scheduling with dependences, periodicities, and latencies constraints. Although a bad approximation of the preemption cost, may have dramatic consequences on real-time scheduling, there are only few researches on this topic. We have been investigating preemptive real-time scheduling since few years, and we focus on the exact cost of the preemption. We have integrated this cost in the schedulability conditions that we propose, and in the corresponding scheduling algorithms. More generally, we are interested in integrating in the schedulability analyses the cost of the RTOS (Real-Time Operating System), for which the cost of preemption is the most difficult part because it varies according to the instance (job) of each task.

In the case of multiprocessor real-time scheduling, we chose at the beginning the partitioned approach, rather than the global approach, since the latter allows task migrations whose cost is prohibitive for current commercial processors. The partitioned approach enables us to reuse the results obtained in the uniprocessor case in order to derive solutions for the multiprocessor case. We consider also the semi-partitioned approach which allows only some migrations in order to minimize the overhead they involve. In addition to satisfy the multiple real-time constraints mentioned in the uniprocessor case, we have to minimize the total execution time (makespan) since we deal with automatic control applications involving feedback loops. Furthermore, the domain of embedded systems leads to solving minimization resources problems. Since these optimization
problems are NP-hard we develop exact algorithms (B & B, B & C) which are optimal for simple problems, and heuristics which are sub-optimal for realistic problems corresponding to industrial needs. Long time ago we proposed a very fast “greedy” heuristics [8] whose results were regularly improved, and extended with local neighborhood heuristics, or used as initial solutions for metaheuristics.

In addition to the spatial dimension (distributed) of the real-time scheduling problem, other important dimensions are the type of communication mechanisms (shared memory vs. message passing), or the source of control and synchronization (event-driven vs. time-triggered). We explore real-time scheduling on architectures corresponding to all combinations of the above dimensions. This is of particular impact in application domains such as automotive and avionics (see 4.2).

The arrival of complex hardware responding to the increasing demand for computing power in next generation systems exacerbates the limitations of the current worst-case real-time reasoning. Our solution to overcome these limitations is based on the fact that worst-case situations may have a extremely low probability of appearance within one hour of functioning ($10^{-45}$), compared to the certification requirements for instance ($10^{-9}$ for the highest level of certification in avionics). Thus we model and analyze the real-time systems using probabilistic models and we propose results that are fundamental for the probabilistic worst-case reasoning over a given time window.

4. Application Domains

4.1. Multicore System-on-Chip design

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

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

Also, Network on Chip (NoC) design may call for computable switching patterns, just like computable scheduling patterns were used in (predictable) Latency-Insensitive Design. Here again formal models, such as Cyclo-static dataflow graphs and extended Kahn networks with explicit routing schemes, are modeling elements of choice for a real synthesis/optimization approach to the design of systems. New parallel architecture paradigms, such as GPU co-processors or Massively Parallel Processor Arrays (MPPA) form natural targets as NoC-based platforms.

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

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

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

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

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

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

5. New Software and Platforms

5.1. TimeSquare

Participants: Nicolas Chleq, Julien Deantoni, Frédéric Mallet [correspondant].

TimeSquare is a software environment for the modeling and analysis of timing constraints in embedded systems. It relies specifically on the Time Model of the MARTE UML profile (see section 3.2), and more accurately on the associated Clock Constraint Specification Language (CCSL) for the expression of timing constraints.

TimeSquare offers five main functionalities:

1. graphical and/or textual interactive specification of logical clocks and relative constraints between them;
2. definition and handling of user-defined clock constraint libraries;
3. automated simulation of concurrent behavior traces respecting such constraints, using a Boolean solver for consistent trace extraction;
4. call-back mechanisms for the traceability of results (animation of models, display and interaction with waveform representations, generation of sequence diagrams...);
5. compilation to pure java code to enable embedding in non eclipse applications or to be integrated as a time and concurrency solver within an existing tool.

In practice TimeSquare is a set of plug-ins developed for the Eclipse modeling framework. The software is registered by the Agence pour la Protection des Programmes, under number IDDIN.FR.R001.170007.000.S.P.2009.001.10600. It can be downloaded from the site http://timesquare.inria.fr/. It has been integrated in the OpenEmbeDD ANR RNTL platform and recently in the Gemoc Studio.
5.2. K-Passa

Participans: Jean-Vivien Millo [correspondant], Robert de Simone.

This software is dedicated to the simulation, analysis, and static scheduling of Event/Marked Graphs, SDF and KRG extensions. A graphical interface allows to edit the Process Networks and their time annotations (latency, ...). Symbolic simulation and graph-theoretic analysis methods allow to compute and optimize static schedules, with best throughputs and minimal buffer sizes. In the case of KRG the (ultimately k-periodic) routing patterns can also be provided and transformed for optimal combination of switching and scheduling when channels are shared. KPASSA also allows for import/export of specific description formats such as UML-MARTE, to and from our other TimeSquare tool.

The tool was originally developed mainly as support for experimentations following our research results on the topic of Latency-Insensitive Design. This research was conducted and funded in part in the context of the CIM PACA initiative, with initial support from ST Microelectronics and Texas Instruments.

KPASSA is registered by the Agence pour la Protection des Programmes, under the number IDDN.FR.001.310003.000.S.P.2009.000.20700. It can be downloaded from the site http://www-sop.inria.fr/aoste/index.php?page=software/kpassa.

5.3. SynDEx

Participans: Yves Sorel [correspondant], Meriem Zidouni.

SynDEx is a system level CAD software implementing the AAA methodology for rapid prototyping and for optimizing distributed real-time embedded applications. Developed in OCaml it can be downloaded free of charge, under Inria copyright, from the general SynDex site http://www.syndex.org.

The AAA methodology is described in section 3.3. Accordingly, SYDEx explores the space of possible allocations (spatial distribution and temporal scheduling), from application elements to architecture resources and services, in order to match real-time requirements; it does so by using schedulability analyses and heuristic techniques. Ultimately it generates automatically distributed real-time code running on real embedded platforms. The last major release of SYNDEX (V7) allows the specification of multi-periodic applications.

Application algorithms can be edited graphically as directed acyclic task graphs (DAG) where each edge represents a data dependence between tasks, or they may be obtained by translations from several formalisms such as Scicos (http://www.scicos.org), Signal/Polychrony (http://www.irisa.fr/espresso/Polychrony/download.php), or UML2/MARTE models (http://www.omg.org/technology/documents/profile_catalog.htm). Architectures are represented as graphical block diagrams composed of programmable (processors) and non-programmable (ASIC, FPGA) computing components, interconnected by communication media (shared memories, links and busses for message passing). In order to deal with heterogeneous architectures it may feature several components of the same kind but with different characteristics.

Two types of non-functional properties can be specified for each task of the algorithm graph. First, a period that does not depend on the hardware architecture. Second, real-time features that depend on the different types of hardware components, ranging amongst execution and data transfer time, memory, etc.. Requirements are generally constraints on deadline equal to period, latency between any pair of tasks in the algorithm graph, dependence between tasks, etc.

Exploration of alternative allocations of the algorithm onto the architecture may be performed manually and/or automatically. The latter is achieved by performing real-time multiprocessor schedulability analyses and optimization heuristics based on the minimization of temporal or resource criteria. For example while satisfying deadline and latency constraints they can minimize the total execution time (makespan) of the application onto the given architecture, as well as the amount of memory. The results of each exploration is visualized as timing diagrams simulating the distributed real-time implementation.
Finally, real-time distributed embedded code can be automatically generated for dedicated distributed real-time executives, possibly calling services of resident real-time operating systems such as Linux/RTAI or Osek for instance. These executives are deadlock-free, based on off-line scheduling policies. Dedicated executives induce minimal overhead, and are built from processor-dependent executive kernels. To this date, executive kernels are provided for: TMS320C40, PIC18F2680, i80386, MC68332, MPC555, i80C196 and Unix/Linux workstations. Executive kernels for other processors can be achieved at reasonable cost following these examples as patterns.

5.4. Lopht

Participants: Thomas Carle, Manel Djemal, Dumitru Potop Butucaru [correspondant].

The Lopht (Logical to Physical Time Compiler) has been designed as an implementation of the AAA methodology. Like SynDEx, Lopht relies on off-line allocation and scheduling techniques to allow real-time implementation of dataflow synchronous specifications onto multiprocessor systems. But there are several originality points: a stronger focus on efficiency, which results in the use of a compilation-like approach, a focus on novel target architectures (many-core chips and time-triggered embedded systems), and the possibility to handle multiple, complex non-functional requirements covering real-time (release dates and deadlines possibly different from period, major time frame, end-to-end flow constraints), ARINC 653 partitioning, the possibility to preempt or not each task, and finally SynDEx-like allocation.

Improved efficiency is attained through the use of classical and novel data structures and optimization algorithms pertaining to 3 fields: synchronous language compilation, classical compiler theory, and real-time scheduling. A finer representation of execution conditions allows us to make a better use of double resource reservation and thus improve latency and throughput. The use of software pipelining allows the improvement of computation throughput. The use of post-scheduling optimizations allows a reduction in the number of preemptions. The focus on novel architectures means that architecture descriptions need to define novel communication media such as the networks-on-chips (NoCs), and that real-time characteristics must include those specific to a time-triggered execution model, such as the Major Time Frame (MTF). Attaining efficiency also requires a fine-grain description of more classical platform resources, such as the multi-bank RAMs, to allow efficient allocation during scheduling.

Significant contributions to the Lopht tool have been brought by T. Carle (the extensions concerning time-triggered platforms), M. Djemal (the extensions concerning many-core platforms), and Zhen Zhang under the supervision of D. Potop Butucaru. The tool has been used and extended during the PARSEC project. It is currently used in the direct collaboration with Airbus Defence and Space and the CNES, in the IRT SystemX/FSF project, and in the CAPACITES project. It has been developed in OCaml.

5.5. SAS

Participants: Daniel de Rauglaudre [correspondant], Yves Sorel.

The SAS (Simulation and Analysis of Scheduling) software allows the user to perform the schedulability analysis of periodic task systems in the monoprocessor case.

The main contribution of SAS, when compared to other commercial and academic softwares of the same kind, is that it takes into account the exact preemption cost between tasks during the schedulability analysis. Beside usual real-time constraints (precedence, strict periodicity, latency, etc.) and fixed-priority scheduling policies (Rate Monotonic, Deadline Monotonic, Audsley++, User priorities), SAS additionally allows to select dynamic scheduling policy algorithms such as Earliest Deadline First (EDF). The resulting schedule is displayed as a typical Gantt chart with a transient and a permanent phase, or as a disk shape called "dameid", which clearly highlights the idle slots of the processor in the permanent phase.
For a schedulable task system under EDF, when the exact preemption cost is considered, the period of the
permanent phase may be much longer than the least commun multiple (LCM) of the periods of all tasks, as
often found in traditional scheduling theory. Specific effort has been made to improve display in this case. The
classical utilization factor, the permanent exact utilization factor, the preemption cost in the permanent phase,
and the worst response time for each task are all displayed when the system is schedulable. Response times of
each task relative time can also be displayed (separately).

SAS is written in OCaML, using CAMLP5 (syntactic preprocessor) and OLIBRT (a graphic toolkit under X).
Both are written by Daniel de Rauglaudre. It can be downloaded from the site http://pauillac.inria.fr/~ddr/sas-
dameid/.

6. New Results

6.1. Languages, Models of Computation and Metamodelling using logical clock constraints

Participants: Julien Deantoni, Robert de Simone, Frédéric Mallet, Marie Agnès Peraldi Frati.

A revised and updated version of our previous work on UML MARTE Time Model was written in survey
textbook form for a larger audience, and published in [38]. Same was done for the more applied specific
findings of the ARTEMIS PRESTO European project [39]. Also, a research report finalizing the denotational
semantics of the logical clock constraint languages was issued for reference [44].

6.2. Experiments with Architecture and Application modeling

Participants: Robert de Simone, Émilien Kofman, Jean-Vivien Millo, Amine Oueslati, Mohamed Bergach.

We submitted for publication our theoretical results on formal mapping of an application written as a process
network dataflow graph onto an abstract architecture model involving a network-on-chip and manycore
processor arrays [24].

In the context of the FUI Clistine collaborative project (which aims at building a cheap supercomputer by
assembling low-cost, general-purpose and network processors interconnected by a time-predictable, on-board
network), we considered the issue of classifying general application types, in the fashion inherited from
UC. Berkeley’s 13 "dwarfs" [46]. Meanwhile, the modeling of desired architecture was slightly postponed
due to hesitations from the main industrial partner (that will build the prototype itself). This work was the
topic of Amine Oueslati’s first year PhD. The classification, and the use of distinct type properties for efficient
and natural encoding, was applied on typical application programs provided by partners (Galerkin methods
for electromagnetic simulation by the Nachos Inria team, ray-tracing algorithms by the Optis/Simplysim SME
design company).

In the context of Mohammed Bergach’s CIFRE PhD contract with Kontron Toulon, we conducted an advanced
modeling exercise on how to best fit large DFT (Discrete Fourier Transform) modules onto a specific processor
architecture (first Intel Sandybridge, then Haswell) that offers computing compromise costs (in performance
vs power) between regular CPUs and GPU hardware accelerators. There were two issues: first, how to best
dimension the size of the largest FFT block that may be performed locally on a corresponding GPU compute
block; second, how to distribute the many such optimal size FFT block needed in a typical radar application,
using the GPU and CPU features at the best of their capacity, with account of the slow data transfer latencies
across memory banks (to and from the GPU registers).

As a side-effect, people from Kontron are now using and distributing to their customers the FFT GPU libraries
with ad-hoc FFT variants matching the GPU block memory sizes. The development, rather lengthy in the case
of Sandybridge, was quickly adjusted and ported for Haswell. A new workshop paper is under submission.
6.3. Multiview modeling with performance and power aspects

**Participants:** Julien Deantoni, Ameni Khecharem, Robert de Simone, Emilien Kofman, Carlos Gomez.

In the context of the ANR HOPE project and The CIM PACA Design Platform, we continued our work on joint modeling and co-simulation of abstract architecture and application (use case scenario) models, together with non-functional aspect views such as performance, power and temperature. The goal of the HOPE project is to consider hierarchical {Power/Performance/Temperature} Management Units (MU), and our target is to connect our IDM modeling with dedicated tools such as Synopsys Platform Architect or Docea Power AcePlorer. The IP_Xact interface format for IP blocks is also aimed for compositional assembly representation, including non-functional properties and timing semantics constraints for co-simulation. This work is mostly continued from the former PhD thesis of Carlos Gomez to a new framework by Ameni Khecharem, as part of her PhD. Practical co-simulation trends are also investigated. Current results were reported in [29].

6.4. Heterogeneous Languages Coordination with Concurrency and Time

**Participants:** Julien Deantoni, Matias Vara Larsen, Robert de Simone, Frédéric Mallet.

In the context of the ANR GEMOC project and in closely related to the mutiview approach of the team, we focused on how to deal with analysis and simulation of heterogeneous languages. Supporting coordinated use of heterogeneous domain specific languages leads to what we called the globalization of modeling language [22]. Concretely, we proposed to define a language behavioral interface to exhibit the concurrency and time aspects of the semantics of a language. The concurrency and time aspects are described by a formal extension of CCSL, named MoCCML (Model of Concurrency and Communication Modeling Language [45], [28]). Any models that conform such language exhibit a symbolic representation of all its acceptable schedules. Based on this, we shown that it is possible to coordinate heterogeneous models. To avoid redundant model coordination activities, we reified the know-how about model coordination in BCOoL (Behavioral Coordination Operator Language [34]), a language dedicated to language coordination. This work is mainly realized by Matias Vara Larsen, as part of his PhD. In this context, we organize the community around such subject for the second year in an international workshop [43] with an increasing number of participants.

6.5. Performance study of Massively Parallel Processor Array (MPPA) SoC architecture

**Participants:** Sid Touati, Franco Pestarini.

From a previous collaboration programme, we (Aoste Sophia) possess a MPPA manycore chip, designed and produced by the company Kalray, in Grenoble. The chip integrates 256 cores, composed of 16 clusters (themselves each with 16 cores), and a powerfull network-on-chip interconnect mesh structure. This architecture is oriented towards high performance embedded application, with real time constraints. The cores and NoC were designed to deliver predictable performance.

Our current project, during Franco Pestarini Inria International Intern period, was to test the performance of the NoC, trying to obtain better knowledge of its behavior. We put up a set of microbenchmarks to exercice the network under different specific scenarios (low overhead network traffic, high traffic), and analyzed the experimental results.

We produced a detailed deliverable report explaining under which conditions the NoC could deliver stable and predictable performances. We identified potential configurations where the network becomes unstable (leading to variable and impredictable performances and bandwith).

Meanwhile, the textbook on low-level code optimization, written between Sid Touati and Kalray CTO, appeared in published form [42]. Its content reports on some of the techniques used inside the MPPA compilation environment, and beyond.

**Participant:** Sid Touati.

This activity is conducted by Sid Touati in collaboration with Julien Worms, an associate professor in Mathematics at the university of Versailles Saint Quentin. It was started under the consideration that the performances of programs are hardly ever represented by a gaussian distribution. So, our purpose here is to study parametric statistics for analyzing the performances of programs. We are interested in modelling program performances by gaussian mixtures (using mixmod method). After a statistical modeling, we deduce multiple performance tests and performance criteria to decide with a high degree of confidence about the "best" program run version. This is still work-in-progress: we are implementing a free software for analysis based on our approach, and we are writing a rather complete research report prior to further publications in conferences and journals.

6.7. Uniprocessor Real-Time Scheduling

**Participants:** Falou Ndoye, Yves Sorel, Walid Talaboulma.

6.7.1. Real-Time Scheduling with Exact Preemption Cost

Previous years, we worked on schedulability analyses of dependent tasks, executed on a uniprocessor, which take into account the exact preemption cost and more generally the cost of the real-time operating system. Indeed, this cost is composed of the cost of the preemptions and the cost of the scheduler. Our approach is based on an offline real-time schedulability analysis, proved sustainable, that produces a scheduling table. This latter contains the next instants (activation and completion of tasks) when the scheduler will be called, being aware of the instants where tasks are preempted and then resumed. This approach allows the schedulability analysis to account preemption costs involved by other preemptions. The scheduling table contains also the address of the next task to execute preventing the scheduler to choose it in the ready tasks list, unlike with classical on-line scheduler. The theoretical results in the uniprocessor case, are given in the Falou Ndoye’s PhD thesis [19] defended in April this year. This approach has been implemented through an offline scheduler that is triggered by a timer when this latter is equal to zero and loaded with the next instant contained in the scheduling table, according to a time trigger approach.

We carried out two kinds of implementations. Actually, the first one is a simulation since our time trigger offline scheduler is modelled as a high priority task running uppon an existing operating system. We experimented this approach with Vanilla Linux (not modified) and Linux/Xenomai, real-time versions of the Linux operating system, with low latency characteristics. Of course, these implementations were only able to show that the theoretical results were correct, but did not provide good real-time performances nor a robust way to measure time without influencing the usefull code. Therefore, we implemented our scheduler on a bare metal ARM968E-S processor based on an ARM9 architecture since it is widespread in the industry world, and we experimented this processor few years ago to determine the cost of classical online schedulers.

For this purpose we used a MCB2929 development board, from Keil, containing the LPC2929 SoC including the ARM968E-S, an accurate timer, and various peripherals. The scheduling table is generated offline for a set of tasks, and stored in the memory as an array of couples, each composed of the task to execute and the duration elapsing until the next scheduler call. This duration is used to set the timer counter. When it hits zero it triggers a high priority interruption that is serviced by calling again the scheduler to choose the next couple (task, duration), and so on up to the end of the scheduling table. This will repeat infinitely from the beginning of the scheduling table.

We tested different set of tasks with multiple preemption scenarios, that can yield to deadlines misses. We measured for a 12Mhz CPU and timer clock frequencies a value of 28 µs for the scheduler cost, and of 1 µs for the preemption cost.

6.8. Multiprocessor Real-Time Scheduling

**Participants:** Aderraouf Benyahia, Laurent George, Falou Ndoye, Dumitru Potop-Butucaru, Yves Sorel, Meriem Zidouni.
6.8.1. Multiprocessor Partitioned Scheduling with Exact Preemption Cost

Since we chose a multiprocessor partitioned scheduling approach, we can take advantage of the results we obtained in the case of uniprocessor real-time Scheduling accounting for the cost of the real-time operating system, i.e. the cost of preemptions and of the scheduler. From the point of view of the off-line real-time schedulability analysis we only have to consider in addition to activation and completion instants, reception of data instants. This latter instant is determined by supposing that the cost of every data transfer is known for every possible communication medium. Indeed, when two dependent tasks are allocated to two different processors, the consuming task will have to wait for the data sent by the producing task. The theoretical results in the multiprocessor case, are given in the Falou Ndoye’s PhD thesis [19] defended in April this year. We chose the message passing protocol for interprocessor communications achieved through a switched ethernet network. In order to determine precisely the cost of data transfers, we started to investigate the possible approaches to synchronize the send and receive tasks located in two different processors and to schedule them with the other tasks allocated to the same processor. This synchronization protocol will be taken into account to determine the interprocessor communication costs. Concerning these communication costs, we consider FIFO and FIFO* schedulings in the switches, the later is a FIFO scheduling based on the release time of frames at their source node. We have first corrected the trajectory approach (recently shown to be optimistic for corner cases) with FIFO scheduling to compute worst case end-to-end communication costs. Then, we have extended the trajectory approach to FIFO* scheduling. We want to implement our off-line scheduler on every processor of a multiprocessor architecture composed of at least three processors, communicating through an ethernet switch.

Concerning the delay of communications, we consider FIFO and FIFO* schedulings in the switches, the later is a FIFO scheduling based on the release time of frames at their source node. We have first corrected the trajectory approach (recently shown to be optimistic for corner cases) with FIFO scheduling to computed worst case end-to-end communication delays. Then, we have extended the trajectory approach to FIFO* scheduling.

6.8.2. Multiprocessor Parallel Directed Acyclic Graph (DAG) scheduling

We are interested in studying the hard real-time scheduling problem of parallel Directed Acyclic Graph (DAG) tasks on multiprocessor systems. In this model, a task is defined as a set of dependent subtasks that execute under precedence constraints. The execution order of these subtasks is dynamic, i.e., a subtask can execute either sequentially or in parallel with its siblings based on the decisions of the real-time scheduler. To this end, we analyze two DAG scheduling approaches to determine the execution order of subtasks: the Model Transformation and the Direct Scheduling approaches. We consider global preemptive multiprocessor scheduling algorithms to be used with the scheduling approaches, such as Earliest Deadline First (EDF) and Deadline Monotonic (DM).

6.8.3. Gateway with Modeling Languages for Certified Code Generation

This work was carried out in the P FUI project 8.2.2 We continued the work on the gateway between the P formalism and SynDex, started the last two years. We have integrated in the gateway the IF and FOR blocks of Simulink that were missing in the functional specification, except for particular cases where the IF block is nested in the FOR block, or the opposite. The integration of the MERGE and MUX blocks are still to be done. We extended the P formalism with architectural elements that SynDex needs to perform schedulability analyses on functional specifications. These architectural elements are hardware resources (processor, bus, shared memory, router) and timing characteristics (deadline, period, WCET, WCTT). We developed a new part in the gateway which transforms an architectural model described with the P formalism in the input format of SynDex. We developed also a third part in the gateway which feedbacks the schedulability analysis results obtained with SynDex (the scheduling table) and stores them into models described with the P formalism. Finally, we have collaborated with the industrial partners to test our gateway on their use cases.

6.8.4. SynDex updates
The first tests on the alpha version of SynDEx V8, released last year, shown some bugs that we fixed. This first release did not include a code generator. Thus, we worked to interface the distributed real-time embedded code generator of SynDEx V7 with SynDEx V8.

6.9. Probabilistic Real-Time Systems

Participants: Liliana Cucu-Grosjean, Robert Davis, Adriana Gogonel, Codé Lo, Dorin Maxim, Cristian Maxim.

The advent of complex hardware, in response to the increasing demand for computing power in next generation systems, exacerbates some of the limitations of static timing analysis for the estimation of the worst-case execution time (WCET) estimation. In particular, the effort of acquiring (1) detail information on the hardware to develop an accurate model of its execution latency as well as (2) knowledge of the timing behaviour of the program in the presence of varying hardware conditions, such as those dependent on the history of previously executed instructions. These problems are also known as the timing analysis walls. The probabilistic timing analysis, a novel approach to the analysis of the timing behaviour of next-generation real-time embedded systems, provides answers to timing analysis walls. In [23] we have described the vision of FP7 IP PROXIMA, project that is interested in the introduction of randomization of the architectures at cache level. For this type of architecture static probabilistic timing analysis is possible [20] by providing bounds on the probabilistic execution time of a task. An industrial case study from avionics is detailed in [32]. Such distribution is then used as input for probabilistic scheduling as described in [37], [36].

This year we have also provided a complete state of the art of the probabilistic real-time systems in [17].

6.10. Off-line (static) mapping and WCET analysis of real-time applications onto NoC-based many-cores

Participants: Dumitru Potop Butucaru, Thomas Carle, Manel Djemal, Robert de Simone, Zhen Zhang.

Modern computer architectures are increasingly relying on multi-processor system-on-chips (MPSoCs, also called chip-multiprocessors), with data transfers between cores and RAM banks managed by on-chip networks (NoCs). This reflects in part a convergence between embedded, general-purpose PC, and high-performance computing (HPC) architecture designs.

In past years we have identified and compared the hardware mechanisms supporting precise timing analysis and efficient resource allocation in existing NoCs. We have determined that the NoC should ideally provide the means of enforcing a global communications schedule that is computed off-line and which is synchronized with the scheduling of computations on CPU cores. Furthermore, if in addition the computation and memory resources of the MPSoC have support for real-time predictability, then parallel applications can be developed that allow very precise WCET analysis of parallel code. WCET analysis of parallel code is joint work with Isabelle Puaut of Inria, EPI ALF.

This year we have completed our mapping (allocation and scheduling) and code generation technique and tool for NoC-based MPSoCs. NoCs pose significant challenges to both on-line (dynamic) and off-line (static) real-time scheduling approaches [25]. They have large numbers of potential contention points, have limited internal buffering capabilities, and network control operates at the scale of small data packets. Therefore, efficient resource allocation requires scalable algorithms working on hardware models with a level of detail that is unprecedented in real-time scheduling.

We considered a static (off-line) scheduling approach, and we targeted massively parallel processor arrays (MPPAs), which are MPSoCs with large numbers (hundreds) of processing cores. We proposed a novel allocation and scheduling method capable of synthesizing such global computation and communication schedules covering all the execution, communication, and memory resources in an MPPA. To allow an efficient use of the hardware resources, our method takes into account the specificities of MPPA hardware and implements advanced scheduling techniques such as pre-computed preemption of data transmissions[26] and pipelined scheduling [21].
Our method has been implemented within the Lopht tool presented in section 5.4, and first results are presented in [26], [25], and in extenso in the PhD thesis of manel Djemal [18]. One of the objectives of the starting CAPACITES project is the evaluation of the possibility of porting Lopht and the WCET analysis technique for parallel code onto the Kalray MPPA platform.

6.11. Real-time scheduling and code generation for time-triggered platforms

Participants: Dumitru Potop Butucaru, Thomas Carle, Raul Gorcitz, Yves Sorel.

We have continued this year the work on real-time scheduling and code generation for time-triggered platforms. Much of this work was carried out as part of a bilateral collaboration with Airbus DS and the CNES, which fund the post-doctorate of Raul Gorcitz, and in our collaboration with the IRT SystemX, project FSF:

The objective is to facilitate the development of complex time-triggered systems by automating the allocation, scheduling, and code generation steps. We show that full automation is possible while taking into account all the specification elements required by a complex, real-life embedded control system. The main originality of our work is that it takes into account at the same time multiple complexity elements: functional specifications with conditional execution and multiple modes and various types of non-functional properties: real-time (release dates, deadlines, major time frame, end-to-end flows), ARINC 653 partitioning (which we can fully or partially synthesize), task preemptability, allocation. Our algorithms allow the automatic allocation and scheduling onto multi-processor (distributed) systems with a global time base, taking into account communication costs.

While the past years were mainly dedicated to the development of this scheduling and code generation technique, this year the technique and the associated tool have matured enough to allow the publication of the first results concerning the optimized scheduling algorithms [21] and its application on large case studies. Ongoing work by the post-doc Raul Gorcitz, funded by Airbus DS and the CNES aims at evaluating the applicability of our methods on embedded platforms that are being considered for the future european space launchers. The Lopht tool is also used in the IRT SystemX, project FSF as part of the proposed design flow. All extensions have been implemented in the Lopht tool. All this work has been presented in extenso in the PhD thesis of Thomas Carle [16].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Kontron PhD CIFRE thesis

Participants: Mohamed Bergach, Robert de Simone.

Kontron Toulon (formerly Thales Computers) has a strong interest in finding optimal (or at least efficient) mapping of applications extensively based on FFT computations (mostly radar detection), onto GPGPU architectures of the Intel IvyBridge/Haswell family (that ar ethen integrated into avionic subsystems at Kontron). This is the main topic of Mohamed Bergach PhD thesis, which should be defended in late Spring 2015. A publication is under submission.

7.1.2. Airbus PhD CIFRE thesis

Participants: Liliana Cucu-Grosjean, Cristian Maxim.

As part of a larger collaborative programme between Inria and Airbus, the PhD thesis of Cristian Maxim has started in March 2014. This thesis will propose a methodology for obtaining probabilistic worst-case execution times distributions by characterizing the appropriate properties of Airbus applications and platforms. This first year is dedicated to the familiarization of Cristian Maxim to the Airbus applications and platforms.
7.1.3. Astrium/CNES PostDoc

The objective of our collaboration with Airbus Defence and Space and the CNES is to determine how the design and implementation of embedded software and system/network configuration can be largely automated in an aerospace context. The objective is to reduce the design and validation costs (especially in case of system evolutions), while preserving an assurance level superior to that of the Ariane 5 flight program. We are exploring automation of the real-time allocation, scheduling, and code generation using the novel algorithms developed and implemented in the Lopht tool. The application of such techniques also requires extensions at the level of system specification formalisms. This collaboration has funded the post-doctoral period of Raul Gorcitz (started in September 2013, reconducted for one year).

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. CIM PACA Design Platform

Participants: Robert de Simone, Ameni Khecharem, Carlos Gomez Cardenas, Emilien Kofman.

This ambitious regional initiative is intended to foster collaborations between local PACA industry and academia partners on the topics of microelectronic design, though mutualization of equipments, resources and R&D concerns. We are active in the Design Platform (one of three platforms), of which Inria is a founding member. This provides opportunities for interactions with local companies, leading indirectly to more formal collaborations at times. Phase 3 of the CIM PACA programme should be launched in 2015, and was subject of extensive preparation at the end of 2014.

The ANR HOPE project 8.2.1.1 is conducted under the auspices of the CIM PACA Design Platform, which also hosts prototype and commercial software products contributed by project members (Synopsys, Docea Power, and Magillem, see 8.2.1.1). Similarly, the CLISTINE FUI project was labeled by the platform as microelectronic branch of the SCS competitiveness cluster.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. HOPE

Participants: Carlos Gomez Cardenas, Ameni Khecharem, Emilien Kofman, Robert de Simone.

The ANR HOPE project focuses on hierarchical aspects for the high-level modeling and early estimation of power management techniques, with potential synthesis in the end if feasible. Although this project was officially started in November 2013, it was in part postponed due to the replacement of a major partner (Texas Instruments) by another one (Intel). Current partners are CNRS/UNS UMR LEAT, Intel, Synopsys, Docea Power, Magillem, and ourselves. A publication on multiview modeling (including performance, power, and temperature) was presented at FDL’2014, reflecting Ameni Khecharem ongoing PhD work.

8.2.1.2. GeMoC

Participants: Matias Vara Larsen, Julien Deantonii, Frédéric Mallet.

This project is administratively handled by CNRS for our joint team, on the UMR I3S side. Partners are Inria (Triskell EPI), ENSTA-Bretagne, IRIT, Obeo, Thales TRT. The project focuses on the modeling of heterogeneous systems using Models of Computation and Communication for embedded and real-time systems, described using generic means of MDE techniques (and in our case the MARTE profile, and most specifically its Time Model, which allows to specify precise timely constraints for operational semantic definition).
8.2.2. FUI

8.2.2.1. FUI P

Participants: Abderraouf Benyahia, Dumitru Potop Butucaru, Yves Sorel.

The goal of project P is to support the model-driven engineering of high-integrity embedded real-time systems by providing an open code generation framework able to verify the semantic consistency of systems described using safe subsets of heterogeneous modeling languages, then to generate optimized source code for multiple programming (Ada, C/C++) and synthesis (VHDL, SystemC) languages, and finally to support a multi-domain (avionics, space, and automotive) certification process by providing open qualification material. Modeling languages range from behavioural to architectural languages and present a synchronous and asynchronous semantics (Simulink/Scilab, Scicos, Xcos, SysML, MARTE, UML).

See also: http://www.open-do.org/projects/p/

Partners of the project are: industrial partners (Airbus, Astrium, Continental, Rockwell Collins, Safran, Thales), SMEs (AdaCore, Altair, Scilab Enterprise, STI), service companies (ACG, Aboard Engineering, Atos Origins) and research centers (CNRS, ENPC, Inria, ONERA).

8.2.2.2. FUI CLISTINE

Participants: Robert de Simone, Amin Oueslati, Emilien Kofman.

This project was started in Oct 2013, a provides PhD funding for Amine Oueslati. Partners are SynergieCAD (coordinator), Avantis, Optis, and the two EPs Aoste and Nachos. The goal is to study the feasibility of building a low-cost, low-power "supercomputer", reusing ideas from SoC design, but this time with out-of-chip network "on-board", and out-of-the-shelf processor elements organized as an array. The network itself should be time predictable and highly parallel (far more than PCI-e for instance). We started a thorough classification of parallel program types (konown as "Dwarfs" in teh literature), to provide benchmarks to evaluate the platform design options.

8.2.3. Investissements d’Avenir

8.2.3.1. DEPARTS

Participants: Liliana Cucu-Grosjean, Adriana Gogonel, Codé Lo, Cristian Maxim.

This project is funded by the BGLE Call (Briques Logicielles pour le Logiciel Embarqué) of the national support programme Investissements d’Avenir. Formally started on October 1st, 2012 with the kick-off meeting held on April, 2013 for administrative reasons. Research will target solutions for probabilistic component-based models, and a Ph.D. thesis should start at latest on September 2015. The goal is to unify in a common framework probabilistic scheduling techniques with compositional assume/guarantee contracts that have different levels of criticality.

8.2.3.2. CLARITY

Participants: Yann Bondue, Julien Deantoni, Robert de Simone, Marie Agnès Peraldi-Frati.

This project is funded by the LEOC Call (Logiciel Embarqué et Objets Connectés) of the national support programme Investissements d’Avenir. It was started in September 2014, and a kick-off meeting was held on October 9th. Partners are: Thales (several divisions), Airbus, Areva, Altran, All4Tec, Artal, the Eclipse Fondation, Scilab Enterprises, CESAMES, U. Rennes, and Inria. The purpose of teh project is to develop and promote an open-source version of the ARCADIA Melody system design environment from Thales, renamed CAPPELLA for that purpose.

8.2.3.3. Capacites

Participants: Liliana Cucu-Grosjean, Dumitru Potop-Butucaru, Yves Sorel, Walid Talaboulma.

This project is funded by the LEOC Call (Logiciel Embarqué et Objets Connectés) of the national support programme Investissements d’Avenir. It has started on November 1st, 2014 with the kick-off meeting held on November, 12th 2014. The project cordinator is Kalray, and teh objective of the project is to stydy relevance of Kalray-style MPPA processor array for real-time computation in the avionic domain (with partners such as Airbus for instance).
8.3. European Initiatives

8.3.1. Collaborations in European Programs, except FP7 & H2020

8.3.1.1. ARTEMIS PRESTO

Participants: Frédéric Mallet, Arda Goknil, Julien Deantoni, Marie Agnès Peraldi Frati, Robert de Simone, Jean-Vivien Millo.

Type: ARTEMIS
Project title: PRESTO
Duration: April 2011 - March 2014
Coordinator: Miltech (Greece)
Others partners: TELETEL S.A. (Greece), THALES Communications (France), Rapita Systems Ltd. (United Kingdom), VTT (Finland), Sogeti (France), THALES (Italy), MetaCase (Finland), Inria (France), University of L’aAquila (Italy), MILTECH HELLAS S.A (Greece), FragmaDev (France), PrismaTech (United Kingdom), Sarokal Solutions (Finland).

See also: http://www.cesarproject.eu/

Abstract: The PRESTO project aims at improving test-based embedded systems development and validation, while considering the constraints of industrial development processes. This project is based on the integration of test traces exploitation, along with platform models and design space exploration techniques. Such traces are obtained by execution of test patterns, during the software integration design phase, meant to validate system requirements. The expected result of the project is to establish functional and performance analysis and platform optimisation at early stage of the design development. The approach of PRESTO is to model the software/hardware allocation, by the use of modelling frameworks, such as the UML profile for model-driven development of Real Time and Embedded Systems (MARTE). The analysis tools, among them timing analysis including Worst Case Execution Time (WCET) analysis, scheduling analysis and possibly more abstract system-level timing analysis techniques will receive as inputs on the one hand information from the performance modelling of the HW/SW-platform, and on the other hand behavioural information of the software design from tests results of the integration test execution.

8.4. International Initiatives

8.4.1. Inria International Labs

8.4.1.1. HADES LIAMA project

This joint project is held in collaboration with ECNU Shanghai, together with the Scale Inria team, and extends in scope the Associated Team DAESD (see below). As part of this project Frédéric Mallet spends a sabbatical year at ECNU Shanghai, partly funded by an Inria delegation programme.

We attended a number of LIAMA meetings, both in France and in Beijing, most often in confcall form.

8.4.2. Inria Associate Teams

8.4.2.1. DAESD

Title: Distributed/Asynchronous and Embedded/synchronous Systems Development
Inria principal investigator: Robert de Simone (Aoste) / Eric Madelaine (Scale)
International Partner (Institution - Laboratory - Researcher):
   East China Normal University (China) - SEI-Shone - Robert De Simone
Duration: 2012 - 2014

See also: https://team.inria.fr/DAESD/
The development of concurrent and parallel systems has traditionally been clearly split in two different families: distributed and asynchronous systems on one hand, now growing very fast with the recent progress of the Internet towards large scale services and clouds; embedded, reactive, or hybrid systems on the other hand, mostly of synchronous behaviour. The frontier between these families has attracted less attention, but recent trends, e.g. in industrial systems, in Cyber-Physical systems (CPS), or in the emerging Internet of Things, give a new importance to research combining them.

The aim of the DAESD associate team is to combine the expertise of the Oasis and Aoste teams at Inria, the SEI-Shone team at ECNU-Shanghai, and to build models, methods, and prototype software tools inheriting from synchronous and asynchronous models. We plan to address modelling formalisms and tools, for this combined model; to establish a method to analyze temporal and spatial consistency of embedded distributed real-time systems; to develop scheduling strategies for multiple tasks in embedded and distributed systems with mixed constraints.

A dedicated Summer School was organized this year in Shanghai (July 8-11), with participation of Julien Deantoni and Frédéric Mallet from Aoste.

DAESD is strongly linked with the LIAMA project HADES, that it supports.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

8.5.1.1. Invited Professor

Qingguo XU

Date: July 2014 to June 2015
Institution: Shanghai University (China)

8.5.2. Visits to International Teams

8.5.2.1. Sabbatical programme

Mallet Frédéric

Date: Sep 2014 - Aug 2015
Institution: ECNU (China)

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. general chair, scientific chair

- Julien Deantoni: GeMoc’2014 (workshop) [43]
- Marie-Agnès Peraldi-Frati: CIEL’2014 [35]

9.1.2. Scientific events selection

9.1.2.1. responsable of the conference program committee


9.1.2.2. member of conference program committee


9.1.3. Journal

9.1.3.1. member of the editorial board

Yves Sorel: Traitement du Signal Journal
9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: Robert de Simone, Models of Computation for Networks-on-Chips (MoCs for NoCs), 36h, M2 Internationbal, UNS.
Master: Robert de Simone, Functional and Temporal Correctness, 36h, M1 Internationbal, UNS.
Master: Yves Sorel, Optimization of distributed real-time embedded systems, 24H, M2, University Paris Sud
Master: Yves Sorel, Distributed real-time systems, 26H, M2, University Paris Est
Master: Yves Sorel, Specification and formal models for embedded systems, 28H, M2, ENSTA, Paris
Master: Yves Sorel, Correct by construction design of reactive systems, 18H, M2, ESIEE, Noisy-Le-Grand
Master : Julien Deantoni, Encadrement Projet en équipe, 31h, M1, Polytech’Nice
Master : Julien Deantoni, Encadrement Apprentissage, 8h, M1, Polytech’Nice
Master : Julien Deantoni, Architecture machine, 24h, niveau M1, Polytech’Nice, France.
Master : Julien Deantoni, Systèmes embarqués et Ambient, 10h, M2, Polytech’Nice, France.
Master : Julien Deantoni, Langage C++, 88h, M1,Polytech’Nice, France.
Master : Julien Deantoni, Ingénierie Dirigée par les modèles par la pratique, 24h, M2,Polytech’Nice, France.
Master : Julien Deantoni, Ingénierie Dirigée par les modèles avancée, 9h, M2 Esprit Tunis, Tunisia.
Master: Dumitru Potop Butucaru, Une approche synchrone des systèmes embarqués temps réel,12h, M1, EPITA Paris
Master: Dumitru Potop Butucaru and Thomas Carle, L’approche synchrone de la construction des systèmes embarqués temps réel, 12h, M2, Polytech Paris UPMC.
Licence: Laurent George, Java and Shell programming 48h, L1, IUT RT UPEC, France
Licence: Laurent George, Distributed Real-Time Systems, 24h, M2, UPEC, France
Licence : Marie-Agnes Peraldi-Frati, Algoritms and programming 60h,L1, UNS Institute of technology.
Licence : Marie-Agnes Peraldi-Frati, System and Networks administration 80h, L2, UNS Institute of technology .
Licence : Marie-Agnes Peraldi-Frati, Web Programming 50 h, L2, UNS Institute of technology.
Licence: Frédéric Mallet, Architecture des ordinateurs, 45h, L3, UNS.
Master: Frédéric Mallet, Programmation Avancée et Design Patterns, 93h, M1, UNS.
Master: Frédéric Mallet, Java pour l’Informatique Industrielle, 24h, M1, UNS.
Master: Frédéric Mallet, Architectures des ordinateurs, 12h, M1, UNS.
Master: Frédéric Mallet, Formal Models for Network-On-Chips, 3h, M2, UNS.

9.2.2. Supervision

PhD in progress: Cristian Maxim, Probabilistic Real-Time Systems, University of UMPC-Paris-Sorbonne, co-supervised by Liliana Cucu-Grosjean with Benoit Triquet (Airbus).
PhD: Falou Ndoye, Multiprocessor real-time scheduling taking into account preemption cost [19]; defended April 2014, supervised by Yves Sorel
PhD: Manel Djemal, Reconciling performance and predictibility on a NoC-based MPSoC using offline scheduling techniques [18], EDITE/UPMC, defended June 27th 2014, co-supervised by Alix Munier (UPMC/Lip6) and D. Potop-Butucaru.

PhD in progress: Salah Saidi, *Real-time scheduling for distributed simulation of numerical models*, University of UMPC-Paris-Sorbonne, co-supervised by Nicolas Pernet (IFPEN) and Yves Sorel

PhD in progress: Matias Vara-Larsen, *Toward a formal and hierarchical timed model for concurrent heterogeneous model*, UNS, started November 2012, supervised by Frédéric Mallet, co-supervised by Julien Deantoni.

PhD in progress: Ameni Khecharem, *High-Level modeling of hierarchical power management policies in SoCs*, UNS, started October 2012, supervised by Robert de Simone.

PhD in progress: Ying Lin, *Formal Analysis of polychronous models with MARTE/CCSL*, East China Normal University, started September 2011, supervised by Jing Liu (ECNU), co-supervised by Frédéric Mallet.


PhD in progress: Yann Bondue, *Towards formal multimodal languages*, UNS, started December 2014, supervised by Julien Deantoni.

### 9.2.3. Juries

- Robert de Simone: PhD examiner of Mariem Abdelmoula (UNS)
- Yves Sorel: PhD examiner of Philippe Thierry (University Paris Est)
- Laurent George: PhD reviewer of Georges-Arnaud KEMAYO (ENSMA Poitiers), Vincent LE GOUT (Telecom ParisTech) and Gurulingesh RARAVI (University of Porto).
- Liliana Cucu-Grosjean: PhD reviewer of Maxime Cheramy (University of Toulouse-INSA) and Vladimir-Alexandru Paun (Ecole Polytechnique, Paris)
- Dumitru Potop Butucaru: PhD examiner of Chan Ngo (Univ. Rennes 1)

### 10. Bibliography

**Major publications by the team in recent years**


Publications of the year

Doctoral Dissertations and Habilitation Theses


[18] M. DIJEMAL. *Reconciling performance and predictability on a NoC-based MPSoC using off-line scheduling techniques*, Université Pierre et Marie Curie (Paris VI), June 2014, https://hal.inria.fr/tel-01088003


**Articles in International Peer-Reviewed Journals**


**International Conferences with Proceedings**


National Conferences with Proceedings


Scientific Books (or Scientific Book chapters)


on Embedded Systems Design", A. Bagnato, L. S. Indrusiak, I. R. Quadri, M. Rossi (editors), IGI Global, June 2014, pp. 283–300 [DOI : 10.4018/978-1-4666-6194-3.ch011], https://hal.inria.fr/hal-01088439

[40] D. Potop-Butucaru, Y. Sorel. Approche synchrone et ordonnancement, in "Ordonnancement dans les systèmes temps réel", M. Chetto (editor), ISTE, June 2014, https://hal.inria.fr/hal-01096245


Books or Proceedings Editing


Research Reports

[44] J. Deantonio, C. André, R. Gascon. CCSL denotational semantics, Inria ; I3S, November 2014, n° RR-8628, 29 p., https://hal.inria.fr/hal-01082274


References in notes


