Inria International program Associate Team proposal 2016-2018 Submission form

LifeForm

Title: Life Sciences need Formal Methods !

Associate Team acronym: LifeForm

Principal investigator (Inria):

Principal investigator (Main team):

Stefan HAAR DR2 INRIA Head of EPI MEXICO

Victor KHOMENKO Reader in Formal Methods Newcastle University UK

Other participants:

On the French side,

- Loïc Paulevé, LRI (U Paris-Sud) and CNRS, is a researcher in bioinformatics and systems biology
- César Rodríguez, LIPN (U Paris-Nord)

On the UK side :

• Anil WIPAT, Prof. of Integrative Bioinformatics at School of Computing Science, co-director of the *Centre for Synthetic Biology and the Bioeconomy (CSBB)*.

1 Partnership

1.1 Detailed list of participants

- 1. INRIA, MEXICO team: Concurrency Theory, Diagnosis, Unfoldings
 - Stefan HAAR, senior INRIA researcher; coordinator of LifeForm
 - Thomas CHATAIN, associate professor, ENS Cachan
 - Stefan SCHWOON, associate professor, ENS Cachan
 - Clara Scherbaum, 3rd year ERASMUS intern (Feb-Aug 2015) at ENS Cachan, student from Aachen Univ.;
 - yet to be found PhD student at ENS Cachan, starting in 2016¹

2. CNRS/LRI/U ParisSud, BioInfo team:

- Loïc PAULEVÉ, CNRS researcher; Systems Biology
- 3. Paris Nord, LIPN: Concurrency Theory, Unfoldings
 - Thi Thanh Huyen NGUYEN, PhD student at **LIPN** since Sept 2015; her thesis on distributed Petri Net unfoldings is co-supervised by:
 - César RODRÍGUEZ, Assistant Professor, Univ. Paris 13/LIPN; César has done his PhD in the MEXICO team (defence 2013) under the supervision of Stefan Schwoon and Stefan Haar.
- 4. **Newcastle University**, UK : asynchronous circuits, concurrency semantics; synthetic biology
 - Victor KHOMENKO, School of Computing Science;
 - Maciej KOUTNY, Professor at School of Computing Science;
 - Alex YAKOVLEV, Professor at School of Electrical and Electronic Engineering
 - Andrey MOKHOV, Lecturer at School of Electrical and Electronic Engineering; along with expertise in asynchronous circuit design, Andrey brings in experience with DSLs and their inclusion into Haskell.
 - Jonathan Beaumont, PhD student since September 2014, School of Electrical and Electronic Engineering; design of asynchronous circuits
 - Alessandro de Gennaro, PhD student since January 2015, School of Electrical and Electronic Engineering; research in Conditional Partial Order Graphs
 - Anil WIPAT, Prof. of Integrative Bioinformatics at School of Computing Science, associate director of the *Centre for Synthetic Biology and the Bioeconomy*.
 - Wendy SMITH, Senior Research associate at Institute for Cell and Molecular Biosciences, The Medical School.
 - Goksel MISIRLI, PhD, RA atSchool of Computing Science

 $^{^{1}}$ a DIGICOSME grant has been obtained for cut-set computation with unfoldings in 2015, but no candidate was accepted; we expect a new grant with a candidate for 2016

1.2 Nature and history of the collaboration

The cooperation between the French and UK sides has been carried by a shared passion for partial order semantics and formal verification methods built on them. Thomas Chatain, Stefan Schwoon, César Rodriguéz and Stefan Haar have been cooperating on several subjects with Victor Khomenko for almost a decade, see [GHKS14, CK07, RSK13], on efficient construction and storage of unfoldings construction, their analysis, and their use in diagnosis and other monitoring and verification tasks; in the past three years, these contacts have increasingly also involved Maciej Koutny [CHKS15], Alex Yakovlev and Andrey Mokhov, widening the subjects to boolean nets [CHKS15], biological systems [KKP14] and conditional partial order graphs [MY10]. Shorter visits have been exchanged at least once a year for four years now, resulting in joint papers and PhD juries; we emphasize that Maciej Koutny was an invited professor of ENS Cachan in the MEXICO group in 2014, as well as a reviewer of César Rodriquez' PhD thesis, defended Dec. 2013.

On the UK side, the idea of using formal methods for purposes of synthetic biology has arisen between the above group of computer scientist and electrical and electronic engineering scientists on the one hand, and the group of Anil Wipat on the other; the latter group has joined in the informal exchanges with INRIA in 2014. Independently, INRIA has become involved via cooperation with Loïc Paulevé in the issue of exact identification of attractors in regulatory networks through Petri net unfoldings, and from there to further issues in systems biology [CHJ⁺14], over the last two years.

The Newcastle side bring in their expertise with (i) compact representation of concurrent processes via merged processes and the *Punf* tool, (ii) synthesis of concurrent system models, (iii) synthesis of asynchronous electronic circuits, and (iv) very importantly, in synthetic biology on all levels from concepts to the wet lab. On the French side, one finds background on (i) exploration of concurrent semantics for ordinary *and* contextual nets, including the tools *Mole* and *Cunf*, (ii) exploitation of such semantics for purposes of monitoring and testing of systems, via novel constructs such as reveals relations [BCH13, HKS13], and last but not least on (iii) systems biology, in particular on the discrete event view on regulatory networks.

2 Scientific program

2.1 Context

This project endeavors to bring formal methods for *concurrent* systems to use for the understanding and control of biological systems. While the use of formal models in other branches, such as systems biology, is not novel in itself, there is a wide gap to be filled in *synthetic biology*. In fact, while genetic engineers have been manually designing small synthetic genetic circuits at the molecular level for many years, this practice does not scale up to the whole organism level; rigorous principles and techniques are required. As a 2013 article [Fu] puts it, "synthetic biology is still in an immature development phase ... Synergistic efforts from ... multi-disciplinary teams of biologists, engineers, mathematicians, philosophers, computer experts ... are needed to make synthetic biology breakthroughs...". Passing "from models to cells and back" is also cited as one of "the ten grand challenges of synthetic life" [PM11]. We envisage that (a) a shift from analog to digital is unavoidable for designing large-scale biological systems, (b) that formal methods for such discrete models are necessary to enable automatic synthesis and verification procedures on a large scale, and (c) that the multitude of chemical species interacting, and of processes occurring in parallel with weak interaction, makes biological systems intrinsically concurrent systems. Any advance here is thus threatened by the state space explosion, which can be fought with partial order models.

Those are the challenges addressed by the present proposal, from the formal model and the biological side.

We expect positive impact of the use of causality-based, Petri-net type models both for synthesis and analysis of biological systems. In fact, capturing their behaviour is typically done today with boolean automata models whose interconnections are given by signed digraphs, i.e. any gene can be an activator (+) or inhibitor (-) for the expression of another gene. Such qualitative models of dynamics of regulatory networks [TK01] provide an abstraction level that is appropriate for revealing the main dynamical features of the system, as is has been shown in recent work on the prediction of cell reprogramming determinants using Boolan networks [ICdS13] Also, the multitude of chemical species interacting, and of processes occurring in parallel with weak interaction, makes biological systems intrinsically concurrent systems. Even though qualitative models allow a coarse-grained analysis, capturing their behaviour remains a tremendous challenge for large scale models. State-of-the-art techniques for tackling networks with hundreds or thousands of entities rely on approximations and may be non-conclusive. However, because such models serve now as a base for driving new and costly in-vitro and invivo experiments, it becomes crucial to ensure the correctness and completude of the analyses in silico. To that end, harnessing the complexity of such networks with precise, Petri- net based methods that enable to take advantage of concurrency via a localised, causal analysis, is at the heart of this proposal. We have begun the analysis of long-run behaviours in a biological network, e.g. the possible fates of cells governed by such networks (in differentiation and mutation processes), called *attractors*, via exact, context-aware, Petri-net based methods that enable to take advantage of concurrency via a localised, causal analysis, in [CHJ⁺14]. Going further in refining the interaction causalities, one can then envisage cell reprogramming [ICdS13], i.e. devising controls to reverse differentiation and to steer the cell into a different behaviour.

2.2 Objectives (for the three years)

The idea will be to design a high-level domain-specific language (DSL; compare [BDD11, PP09, BW11, MS83, MS09]) to facilitate the construction of Petri net models of synthetic regulatory networks which can then be validated and analysed, following a modular approach [MHW14, CRM⁺10]. The resulting Petri net models will then inform the selection of pre-defined components from a library of BDBPs, so as to integrate formal verification methods, allowing to construct appropriate biological constraints for their composition and for the synthesis of controls.

The existing cooperation between both sides on unfolding-based verification and monitoring methods [CHKS15, CK07, GHKS14, RSK13] provides crucial steps into this direction. Also, a recent result [RSSK15] opens the possibility of analysing the partially-ordered state space (the unfolding) of a concurrent system in a depth-first fashion. Ongoing work by Cesar and Andrey is already exploiting this new approach. The obejctive is to integrate this new algorithmic technique with approaches steming from interpolation-based model checking, originally developed for software verification [McM06].

The results presented in [RSSK15] also enable the development of novel algorithms for distributed unfolding construction, a subject where both the UK and French side have previous experience [HKK02],[BHK06] and which is at the heart of Huyen's PhD program.

Our main use case, at this point, is the production of heterologous proteins, such as drugs, from a plasmid system in the "chassis" system Bacillus subtilis [HPSW13]. The problem is that producing a foreign protein puts a load on the cell and causes its growth rate to slow. Via analyzing and modelling the effects and causes of metabolic load in this cell, we aim at designing

a feedback circuit to slow down the production of the foreign protein if this happens. In a wider time horizon, we will prepare the ground for synthesizing bacteria capable of performing chemical repair tasks; an ambitious goal here is filling cracks in concrete with a glue produced by the bacteria from the dust found in the crack itself. Jointly with this effort to synthesis *new* biological controls, we will also study the use of to regulatory and signalling mechanisms that exist in living organisms to influence cell evolution. This effort is centered on the French side, where research is continuing (jointly with other partners in Luxemburg) on the exploration of attractors in regulatory networks [CHJ⁺14], computation of cut sets [PAK13] (sets of control variables for enforcing or preventing the cell's evolution into some particular attractor) and cellular reprogramming (means to shift a cell's state from one attractor to another). The long-run purpose is to confront the two control approaches to understand their mutual advantages and potential for being combined.

2.3 Work-program (for the first year)

During the first year, strong efforts are necessary in the following tasks:

- 1. progress quickly, in establishing a common modeling framework, based on Petri-net related formalisms, that allows for both efficient formal methods and for interfacing with the biological practice. We also expect to be able, later in 2016, to anticipate whether a connection with cell reprogramming will be feasible and warrant to envisage a joint research effort in both directions, or whether the cell reprogramming research will diverge from the LifeForm branch.
- 2. Establishing the (first version of) the formal modelling bricks, validated by capturing the dynamics of the *Bacillus subtilis* use case. In the second year, these models will be continually adapted and refined in order to help prepare the in-lab validation.
- 3. For analysing the arising model variants (in particular, boolean Petri nets (see Koutny, Chatain et al, Schwoon/Rodríguez) and Conditional Partial Order Graphs), in particular their efficient unfolding with extensions of the exisiting tools; these developments and implementations themselves will be too long and time-consuming to be completed in this year.

We plan to organize two plenary, workshop - style meetings in 2016, one early in the year in Newcastle to kick off the joint work and identify task forces, and a second one in Cachan, Saclay or another suitable meeting point in the Paris area in the summer. At this second meeting, one of the main topics will be to establish, if possible, a link between the control of synthetic biological circuits and the re-programming of existing cells. The second topic is pursued by the French team, in cooperation with a team at the University of Luxemburg (around T. Sauter and A. del Sol); the workhop will include them along with the French and UK side, to foster trilateral cooperation in view of a future, larger cooperation project.

The preliminary schedule for the first year is as follows:

1. France to UK:

- (a) Kickoff workshop in Newcastle, with all French and UK participants: Feb/ March, one week; modelling, controller synthesis in the main use case, and verification
- (b) Visit by H. Nguyen and C. Rodríguez: June, one week; construction of modular, distributed unfoldings for verification of composite systems, with V. Khomenko
- (c) Visit by S. Haar in May/June 2016, one week; modelling, monitoring, control, with M. Koutny, A. Yakovlev and V. Khomenko

- (d) Visit by T. Chatain and S. Schwoon, October / November, one week; contextual net modelling and efficient unfolding .
- 2. UK to France :
 - (a) Workshop in the Paris area, including V. Khomenko, A. Mokhov, M. Koutny, A. Wipat, W. Smith and G. Misirli; one week, June/July 2016; model validation, DSL, monitoring and control, possible link with cell reprogramming.
 - (b) A. Mokhov and G. Mizirli, 1 week, Sept/October; DSL, use of model variants, controller synthesis
 - (c) V. Khomenko: 1 week, Sept/October; Unfolding techniques and refinements

3 Budget

3.1 Budget (for the first year)

The budget is essentially determined by travel and lodging cost, where the French side takes over travel to UK, and lodging visitors from Newcastle in France.

Estimating the cost of a round trip flight Newcastle-Paris at roughly 250 Euros, and the cost of staying near Cachan/Saclay at 750 Euros per week, we arrive at a cost for the French side of about

- 1. 1500 Euros for travel to the Newcastle Workshop.
- 2. 4500 Euros for lodging Newcastle visitors to the summer workshop near Cachan/Saclay,
- 3. 1000 Euros for the remaining trips to Newcastle, see above, and
- 4. 3000 Euros for the remaining visits from Newcastle, see again the previous section,
- 5. travel or lodging for additional non-permanent staff (PhD student, PostDoc, further interns) that join the French or the UK side during the year , for approx. 2000 Euros.

This yields a total budget of 12000 Euros. Further funding is available on the French side through

- L.Paulevé's l'Univ. Paris-Sud "Attractivit'e" grant of 3 kEuros for 2016, and
- the ENS Cachan's FARMAN project ICAR (3500 Euros for the MEXICO team);

these funds can of course be only in part used within the LifeForm activities.

On the UK side, funding is available through several currently active EPSRC contracts; only the total amount is given here, to indicate that all activities to be charged on the Newcastle side are covered.

- EP/N005791/1 Computational Colloids: Engineered bacteria as computational agents in the design and manufacture of new materials and structures. £240,435
- EP/K039083/1 A New Frontier in Design: The Simulation of Open Engineered Biological Systems. £5,577,007
- EP/L011573/1 SynbiCITE an Imperial College led Innovation and Knowledge Centre (IKC) in Synthetic Biology. £5,074,187
- EP/J02175X/1 An infrastructure for platform technology in synthetic biology. £5,007,845
- EP/K001698/1: "UNderstanding COmplex system eVolution through structurEd behaviouRs (UNCOVER)". £559,122
- EP/L025507/1: "Asynchronous design for analogue electronics (A4A)". £574,524 pounds

We therefore request INRIA funding of 10000 Euros for 2016.

Note : The participation of the Luxemburg team in the summer workshop is not included in this request and will be covered by other funds.

3.2 Strategy to get additional funding

We will be monitoring opportunities in national/bilateral funding schemes such as EPSRC and ANR PCRI that allow to link the French and UK partners. Note that bilateral funding schemes of the PHC type for French-UK cooperations did exist in the past but are stopped at the moment. Further funds are being requested in a ANR PRCI proposal on cell reprogramming that will be submitted jointly with the Luxemburg team. Also, and most importantly, we expect to prepare an H2020 project proposal involving the Newcastle and French sides, probably - if the exchanges and the joint workhop are successful - also involving the Luxemburg team.

Note that a consortium which involved, among others, the partners of the present proposal, had already submitted a FETOPEN proposal (MOBEDYC) in 2014, which was not accepted. Our plan is to construct a new project that is slimmer in terms of the number of partners and of the number of scientific challenges, and focused on well-identified technological challenges. These challenges may then include (variants or extensions of) the biotechnological use cases sketched for the present proposal, and/or medical/pharmaceutical topics related to cell-reprogramming; along with the development of the formal methods, we intend to utilize the cooperation to identify which combination of use cases is most promising here.

These questions are linked also to the choice of project call in the 2016 work program of H2020 to be targeted; besides FETOPEN, a biotechnological call or a call on health issues could be appropriate, and maybe others still. Selecting the submission target, and finding further partnerships in the domain of technological or health stakeholders, is one of the tasks of the associated team, and necessary for its sustainability.

Regional funds (DIGICOSME at Saclay) will be sought to finance additional student internships and PhD theses.

4 Added value

The cooperation of MEXICO and associates with Newcastle University has been inspiring and fruitful for years; MEXICO have been delighted, but not surprised, to learn that the School of Computing Science was ranked 9th overall and 1st for socio-economic impact by the UK 2014 Research Excellence Framework (REF) national exercise. The project partners were involved in two out of four impact case studies submitted for this exercise, namely

- worldwide adoption of asynchronous circuits and improved business process modelling, and
- Novel computational approaches to discover medicines.

The potential for MEXICO in this cooperation is further enhanced by the interaction with the CSBB giving wet-lab input to our research, and allowing wet lab validation of concepts and tools developed in the cooperation.

enefit for the UK side lies in the outreach to cell reprogramming whose potential for computer science and for control in synthetic biology shall be explored during the three year period. Further benefit should come from the enhance synergy in verification techniques.

Perhaps most of all, both partners will benefit from creating a joint task force for concurrency methods in biology, through a major cooperative project on the European scale; *LifeForm* is part of an ongoing effort to build and empower this task force.

5 Other remarks

None.

6 References

6. 1. Joint publications of the partners

- [CHJ⁺14] Thomas Chatain, Stefan Haar, Loïg Jezequel, Loïc Paulevé, and Stefan Schwoon. Characterization of reachable attractors using Petri net unfoldings. In Pedro Mendes, editor, Proceedings of the 12th Conference on Computational Methods in System Biology (CMSB'14), volume 8859 of Lecture Notes in Bioinformatics, pages 129– 142, Manchester, UK, November 2014. Springer-Verlag.
- [CHKS15] Thomas Chatain, Stefan Haar, Maciej Koutny, and Stefan Schwoon. Non-atomic transition firing in contextual nets. In Raymond Devillers and Antti Valmari, editors, Proceedings of the 36th International Conference on Applications and Theory of Petri Nets (ICATPN'15), volume 9115 of Lecture Notes in Computer Science, pages 117– 136, Brussels, Belgium, June 2015. Springer.
- [CK07] Thomas Chatain and Victor Khomenko. On the well-foundedness of adequate orders used for construction of complete unfolding prefixes. *Information Processing Letters*, 104(4):129–136, November 2007.
- [GHKS14] Vasileos Germanos, Stefan Haar, Victor Khomenko, and Stefan Schwoon. Diagnosability under weak fairness. In Proceedings of the 14th International Conference on Application of Concurrency to System Design (ACSD'14), pages 132–141, Tunis, Tunisia, June 2014. IEEE Computer Society Press.
- [RSK13] César Rodríguez, Stefan Schwoon, and Victor Khomenko. Contextual merged processes. In José-Manuel Colom and Jörg Desel, editors, Proceedings of the 34th International Conference on Applications and Theory of Petri Nets (ICATPN'13), volume 7927 of Lecture Notes in Computer Science, pages 29–48, Milano, Italy, June 2013. Springer.

6.2.1. Main publications on the French side relevant to the project

- [BCH13] Sandie Balaguer, Thomas Chatain, and Stefan Haar. Building occurrence nets from reveals relations. *Fundamenta Informaticae*, 123(3):245–272, May 2013.
- [BHK06] Paolo Baldan, Stefan Haar, and Barbara König. Distributed unfolding of Petri nets. In Luca Aceto and Anna Ingólfsdóttir, editors, Proceedings of the 9th International Conference on Foundations of Software Science and Computation Structures (FoS-SaCS'06), volume 3921 of Lecture Notes in Computer Science, pages 126–141, Vienna, Austria, March 2006. Springer.
- [CHJ⁺14] Thomas Chatain, Stefan Haar, Loïg Jezequel, Loïc Paulevé, and Stefan Schwoon. Characterization of reachable attractors using Petri net unfoldings. In Pedro Mendes, editor, Proceedings of the 12th Conference on Computational Methods in System Biology (CMSB'14), volume 8859 of Lecture Notes in Bioinformatics, pages 129– 142, Manchester, UK, November 2014. Springer-Verlag.
- [HKS13] Stefan Haar, Christian Kern, and Stefan Schwoon. Computing the reveals relation in occurrence nets. *Theoretical Computer Science*, 493:66–79, July 2013.
- [RSSK15] César Rodríguez, Marcelo Sousa, Subodh Sharma, and Daniel Kroening. Unfoldingbased partial order reduction. In 26th International Conference on Concurrency Theory, CONCUR 2015, Madrid, Spain, September 1.4, 2015, pages 456–469, 2015. Best Paper Award.

6.2.2. Main publications on the UK side relevant to the project

- [CRM⁺10] Mike T. Cooling, V. Rouilly, Goksel Misirli, James R. Lawson, Tommy Yu, Jennifer Hallinan, and Anil Wipat. Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics*, 26(7):925–931, 2010.
- [HKK02] Keijo Heljanko, Victor Khomenko, and Maciej Koutny. Parallelisation of the petri net unfolding algorithm. In Tools and Algorithms for the Construction and Analysis of Systems, 8th International Conference, TACAS 2002, Held as Part of the Joint European Conference on Theory and Practice of Software, ETAPS 2002, Grenoble, France, April 8-12, 2002, Proceedings, pages 371–385, 2002.
- [KKP14] Jetty Kleijn, Maciej Koutny, and Marta Pietkiewicz-Koutny. Tissue systems and petri net synthesis. Trans. Petri Nets and Other Models of Concurrency, 9:124–146, 2014.
- [MHW14] Goksel Misirli, Jennifer Hallinan, and Anil Wipat. Composable modular models for synthetic biology. *JETC*, 11(3):22:1–22:19, 2014.
- [MY10] Andrey Mokhov and Alexandre Yakovlev. Conditional partial order graphs: Model, synthesis, and application. *IEEE Trans. Computers*, 59(11):1480–1493, 2010.

References

- [BDD11] A. Liu S. Cheung E. Weeding B. Xia M. Leguia J.C. Anderson Bilitchenko, L. and Eugene D. Densmore. A domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PLoS ONE*, 6(4):rsif.2008.0516.focus, 2011.
- [BW11] T. Lu Beal, J. and R. Weiss. Automatic compilation from high-level biologicallyoriented programming language to genetic regulatory networks. *PLoS ONE*, 6(8):e22490, 2011.
- [Fu] Pengcheng Fu. Grand challenges in synthetic biology to be accomplished. *Front.* Bioeng. Biotechnol.
- [HPSW13] C.R Harwood, S. Pohl, W. Smith, and A. Wipat. Bacillus subtilis: Model grampositive synthetic biology chassis. In *Microbial Synthetic Biology*, volume 40, pages 87–117. Academic Press, 2013.
- [ICdS13] W. Jurkowski I. Crespo, T. M. Perumal and A. del Sol. Detecting cellular reprogramming determinants by differential stability analysis of gene regulatory networks. BMC Systems Biology, 7(140), 2013.
- [McM06] Kenneth L. McMillan. Lazy abstraction with interpolants. In Computer Aided Verification, 18th International Conference, CAV 2006, Seattle, WA, USA, August 17-20, 2006, Proceedings, pages 123–136, 2006.
- [MS83] M.A. Marchisio and J. Stelling. 2011, title =Automatic Design of Digital Synthetic Gene Circuits, journal = PLoS Comput. Biol. volume = 7, number = 2, pages = e1001083, doi = 10.1371/journal.pcbi.1001083.
- [MS09] Mario A Marchisio and Jörg Stelling. Computational design tools for synthetic biology. *Current Opinion in Biotechnology*, 20(4):479–485, 2009.
- [PAK13] Loč Paulevé, Geoffroy Andrieux, and Heinz Koeppl. Under-approximating cut sets for reachability in large scale automata networks. In Natasha Sharygina and Helmut Veith, editors, Computer Aided Verification, volume 8044 of Lecture Notes in Computer Science, pages 69–84, Berlin Heidelberg, 2013. Springer.
- [PM11] de Lorenzo V et al. Porcar M, Danchin A. The ten grand challenges of synthetic life. Systems and Synthetic Biology, 5((1-2)):1–9, 2011.
- [PP09] M. Pedersen and A. Phillips. Towards programming languages for genetic engineering of living cells. *Journal of The Royal Society Interface*, 2009.
- [TK01] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. ii. logical analysis of regulatory networks in terms of feedback circuits. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 11:180–195, 2001.