**APPENDIX 1 - The funding application** (to be completed in English)

**C. Phase II – The scientific profile of the project leader**

**C1. Significant and representative scientific achievements (maximum 2 pages)**

The author has a Ph. D. from University of Rochester (U.S.A, 1999.) and has functioned as a postdoctoral fellow, then researcher at Los Alamos National Laboratory between 1999-2007. Since 2007 the P.I. has returned to Romania, at first with a *Marie Curie International Reintegration grant* from the E.U. under the FP6 program. He is now a researcher at the e-Austria Institute Timișoara and an associate professor at the West University, habilitated to direct Ph.D. theses since 2015.

The following elements **point to excellence in the overall scientific research activity** of the P.I.:

- He published papers in prestigious venues such as *ACM-SIAM Symposium on Discrete Algorithms (SODA), International Colloquium on Algorithms, Languages and Programming (ICALP), IEEE Conf. on Computational Complexity (CCC), Satisfiability Conference (SAT), Approx-RANDOM conference, IFIP Networking Conference, Random Structures and Algorithms, Combinatorics, Probability and Computing, Theoretical Computer Science.*

- *He published j*oint papers with <u>leading theoretical computer scientists</u>, such as Samuel Buss (UC San Diego), Moshe Vardi (Rice), Cristopher Moore (Santa Fe Institute), Leslie Ann Goldberg (Oxford), Mark Jerrum (Queen Mary), Dimitris Achlioptas (UC Santa Cruz).

- **Result** from paper G. Istrate *The phase transition în random Horn satisfiability and its algorithmic implication*s, Random Structures and Algorithms (2002) was **included (as an exercise) in D. Knuth *The Art of Computer Programming,* vol. 4, Addison Wesley (2008)**.

- The work of the P.I. has been <u>cited in **leading venues**</u> such as *STOC, FOCS, SODA, AAAI, Combinatorics, Probability and Computing, SIAM Journal of Computing, Random Structures and Algorithms, IEEE-ACM Transactions on Networking, Physical Review E, Combinatorica, Machine Learning*, books published by *Oxford/Cambridge University Press, Springer, North Holland/Elsevier, World Scientific*, Ph.D. theses from the U.S.A. Canada, France, U.K, Spain, Netherlands, Sweden, Czech Republic, Romania, including notable universities such as UC Berkeley, Carnegie Mellon, Rice, UC San Diego, UC Santa Cruz, Toronto, Waterloo, Simon Fraser, ENS Cachan, Chalmers.

The first part of the discussion of authors' scientific achievements mostly relate to his interests during the first period of his scientific career (also the subject of his Ph.D. thesis), that of *Phase Transitions in Combinatorial Optimization:*

- in a paper with conference version in Proceedings of SODA’99 and journal version in *Random Structures and Algorithms* (2002) he analyzed the (lack of a) phase transition in random Horn Satisfiability. The result in that paper has been included (as an exercise) in volume 4 of Donald Knuth’s monumental *The Art of Computer Programming* (Addison Wesley, 2008).
- in a paper with Dimitris Achlioptas and Cristopher Moore (SODA’01) he characterized the threshold of random 1-in-k SAT, an NP-complete problem with an "easy" phase transition.
- in a paper published in the proceedings of the IEEE Conference on Computational Complexity (CCC’00) he obtained a *complete classification of sharpness of phase transitions* in a clausal model of CSP.
- another classification theorem, for a model of random CSP due to Molloy (STOC ) was published by Nadia Creignou and (independently) the P.I. in a paper in Discrete Applied Mathematics (2005).
- in a paper published in Random Structures and Algorithms (2007), the author and his collaborators were the first ones to display a new type of phase transition in a problem related to automated reasoning, a so-called "waterfall" model.
- in a paper published in Journal of Mathematical Physics in 2008 (after his return to Romania, and with part of the work done in this country) the author and collaborators were the first to suggest (through nonrigorous, "physical" arguments) that the random graph bisection problem has a first-order phase transition. This was later confirmed through advanced methods from Statistical Physics.

The work we are proposing relies on concepts and methods from Computational Complexity. The PI is well qualified, having done his Ph.D. with a well-known complexity theorist (Mitsunori Ogihara), and *with* previous papers with complexity-theoretic content (e.g. in Proceedings of Constraint Programming conference). In particular recent work [1,2] from the next section belongs to this area – more precisely to proof complexity.

The proposed work has algorithmic aspects as well: let us point the fact that the PI has results in this area exact: he published several results on the combinatorics of packet reordering in TCP traces, that were published in journals such as *Discrete Applied Mathematics*, *Journal of Combinatorial Optimization*, *Scientific Annals in Computer Science.*
- one accomplishment, published in the *Proceedings of the First Workshop on Foundations and Algorithms for Wireless Networking* - FAWN'06 (satellite of PERCOM'2006) deals with (distributed) approximation algorithms for distance-2 coloring problem in graphs, a combinatorial problem with strong motivations from mobile computing.
-the theme of approximation algorithms is also represented in recent work [8] (see the list in next section)

**C2. Defining elements of the outstanding scientific achievements of the project leader in the last 10 years, 2006 – present[1] (maximum 3 pages)**

*Achivevements outlined in Section C1 correspond to the original research interests of the author, phase transitions.* ***In the past 3-4 years the research interests of the PI have shifted towards Combinatorial Optimization and Computational Complexity.*** *Despite the novelty of these interests,* ***he has obtained some results whose impact qualify them as outstanding:***

- in paper [2] in the list below, the PI together with a team member **first introduced a problem from Combinatorial Topology to the study of Proof Complexity**. The two authors studied the so-called Kneser-Lovasz theorem, showing that lower bounds for the Pigeonhole principle extend to the propositional formulation of this problem. Furthermore, cases k=2 and k=3 of Kneser-Lovasz theorem have polynomial size Frege (extended Frege) proofs. The paper was presented at SAT 2014 conference, part of the Vienna Summer of Logic 2014 "the largest event in the history of logic", being **one of only two papers from Romania accepted to the conferences with competitive accepting process** (CAV, RTA, CSL-LICS, SAT, IJCAR, CSF, KR) on the main track of VSL.

- based on the insights from [2], and in collaboration with an international team composed of scientists from U.C. San Diego (U.S.A.) and Universitat Polytecnica de Catalunya, in paper [1] we recently obtained what we regard as **our most significant research result of the last 10 years:** the (propositional formulation of the) Kneser-Lovasz theorem has quasipolynomial size Frege proofs and polynomial size extended Frege proofs. On the other hand the topological principle employed to prove the Kneser-Lovasz theorem, the so-called *octahedral Tucker lemma* has a "miniaturization" that produces a class of formulas that are candidates for separating Frege and extended Frege systems. The technical ingredient behind the proof is an extension of the Erdos-Ko-Rado theorem, proving the existence for every fixed k, of effective, purely combinatorial (mathematical) proofs of the Kneser-Lovasz theorem. This is an **unexpected result**, that **deserves further investigation.** The paper has been published in the proceedings of ICALP 2015. ICALP (International Colloquium on Automata, Languages and Programming), a conference of the European Association for Theoretical Computer Science (EATCS) is the most prestigious conference in theoretical computer science in Europe. Major prizes in the field (e.g. Godel prize) are frequently presented at this conference (together with the American conference STOC). **No scientist based in Romania had published a paper in this conference since 1981.** The last ones to do so are now professors in good American Universities (e.g. Sorin Istrail, now at Brown).

[1] Section C2 of the application will be published on UEFISCDI website. This will be uploaded in the online platform for submission, both as a pdf. file as well as part of the funding application - phase II.

**-** a third **significant achievement of the P.I.**, together with another member of the team, Cosmin Bonchiș, is in the area of combinatorial optimization. In paper [7] from the list below, they studied the concept of *heapable sequence*, a concept introduced in a paper in ANALCO'11 conference [BHMZ11] by a group of American authors, including Michael Mitzenmacher (Harvard).

In [BI15] we gave an algorithm for minimal decomposition of integer sequences into heapable subsequences, pointed out the intriguing asymptotic behavior of this parameter, apparently scaling as $\frac{1+\sqrt{5}}{2} \cdot \ln(n)$, and the connection with Hammersley-type interacting particle systems. The paper was popularized on the influential (professional) blog of Michael Mitzenmacher "My biased coin". It already stimulated further research [BGGS16]. Subsequent developements [BI16] form the subject of an **invited talk by the PI** at the DCFS'16 conference, scheduled for early July 2016.

**1. Articles**


1. J. Aisenberg, M.L. Bonet, S. Buss, A. Crăciun, G. Istrate. Short Proofs of Kneser-Lovasz Principles. *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP 2015),* Lecture Notes in Computer Science vol. 9135, 2015.

2. G. Istrate, A. Crăciun. Proof Complexity and the Kneser-Lovász Theorem. *Proceedings of 17th International Conference on Theory and Applications of Satisfiability Testing–SAT 2014*, Lecture Notes in Computer Science 8561, 138-153, Springer, 2014.

3. C. Moore, G. Istrate, D. Demopoulos and M. Vardi. *A continuous-discontinuous second order transition in the satisfiability of a class of random Horn formulas*. Random Structures and Algorithms 31(2), pp. 173-185, 2007.

4. G. Istrate, M.V. Marathe, S.S. Ravi. Adversarial Scheduling în Discrete Models of Social Dynamics. *Mathematical Structures in Computer Science* 22 (05), 788-815, 2012.

5. G. Istrate. Reachability and recurrence in a modular generalization of annihilating random walks (and lights-out games) to hypergraphs. *Theoretical Computer Science, 580,83-93, 2015.*

6. G. Istrate, C. Bonchiș. Partition into heapable sequences, heap tableaux and a multiset extension of Hammersley's process, *Proceesings of 26th Annual Symposium on Combinatorial Pattern Matching (CPM'15).* Lecture Notes in Computer Science 9135, 261-271, 2015.

7. A. Percus, G. Istrate, B. Tavares Goncalves, R. Sumi, S. Boettcher *The Peculiar Phase Structure of Random Graph Bisection, Journal of Mathematical Physics* vol. 49, no. 12 (2008).

8. G. Istrate, C. Bonchiș, L.P. Dinu. The Minimum Entropy Submodular Set Cover Problem. *Proceedings of the 10th International Conference on Language and Automata Theory (LATA'2016),* Lecture Notes in Computer Science 9618, 295-306, Springer, 2016.

9. G. Istrate, A. Hansson. *Counting preimages of TCP reordering patterns.* Discrete Applied Mathematics 156 (17), 3187-3193 (2008).

10. G. Istrate, A. Hansson, S.P. Kasiviswanathan. Combinatorics of TCP reordering. *Journal of Combinatorial Optimization 12(1-2), 57-70, 2006.*

**2. Books/ chapters (including monographs) :** A. Percus, G. Istrate, C. Moore (ed). *Computational Complexity and Statistical Physics*. Oxford University Press, 2006.

**3. Scientific presentations**: In the past 10 years has been an *invited speaker* as follows:

- Invited Speaker, DCFS 2016, Bucharest.
- Invited Speaker, SWORDS 2014 Workshop, University of Szeged.
- Invited (Tutorial) Speaker DACS 2014 Workshop, Bucharest.
- Invited presentation, Dagstuhl Workshop 14201 *Horn formulas, directed hypergraphs, lattices and closure systems: related formalisms and applications*, Schloss Dagstuhl, Germany, May 2014.
- Invited presentation, CRM Workshop *Statistical Physics în Combinatorial Optimization*, Bellatera (Barcelona), Spain, October 2009.
- Invited presentation, *International Workshop on Stochastic Phenomena*, (Cluj, Romania, 2008)
- Invited presentation, *International Workshop on Complex Systems* (Sovata, Romania, 2007)
- Invited presentation, IPAM Workshop *Crime Hot Spots: Behavioral, Computational and Mathematical Models*, (Los Angeles CA, January 2007).
- invited speaker to universities in Europe (Sevilla, Szeged, Belgrade) and *contributed* speaker to events in Austria, Belgium, Greece, Germany, Hungary, Italy, Spain, Switzerland, U.S.A.

**4. Reseach projects**

The project director has successfully transitioned from a scientific career in the United States to an academic/research career in Romania. This has been achieved through **Marie Curie International Reintegration Fellowship** PHASETRANS, IRG-046573, "Phase transitions in Computational Complexity and Formal Verification: Towards Generic and Realistic Approaches". 2007 – 2010.

- He was also a member of the Management Committee of COST action IC-0901 "Rich Model Toolkit: An Infrastructure for Reliable Computer Systems", 2007-2011.
- The PI led the IDEI project STRUCTCOMB (2012-2016) **"Structure and computational difficulty in combinatorial optimization: an interdisciplinary approach",** PCE-2011-3-0981, 2012-2016. that led to the recent achievements reported above. He was also in charge of the e-Austria Research Institute team on several projects, including NATCOMP.
- he was a (co)-PI of various projects at Los Alamos National Laboratory, including ``New Approaches to fault-tolerance" (Oct. 2005- 2006) , Generic Cities R&D Project (co-PI 2005-June 2006), LDRD-ER, Advanced methods in discrete simulation (2001-2004). Member în projects at LANL including LDRD-DR ``Statistical Mechanics of Infractructure Networks", LDRD-ER ``Foundations of local search".

*D.* **Research project description** *(maximum 10 pages and maximum 2 pages bibliography)*

***In what follows, numerical references are to our own work listed in Section C2.***

**D1. Issues.** Computational Complexity plays a particular position in Computer Science as a scientific discipline: on one hand it is an area of great intellectual progress and applicability, with results relevant to a variety of disciplines, from Quantum Physics to Economics or Political Science. This is exemplified by the recent incorporation of computer scientists on the scientific board of a leading journal in Economics, *Games and Economic Behavior.* Also, despite having a more applied outlook, computational complexity has contributed to the solution of open problems from various areas of Pure Mathematics: for instance, techniques developed in connection to the study of expander graphs are behind the recent solution of the so-called *Kadison-Singer conjecture* in Operator Theory. Leading pure mathematicians, such as Fields medalists Michael Freedman, Terrence Tao or Tim Gowers have recently worked on topics related to Computational Complexity, publishing in venues specific to the field, such as the FOCS and STOC conferences [GV15, TV07].

On the other hand Computational Complexity is an area where many tantalizingly simple problems are still open: in contrast to the information-theoretic lower bound (due to Shannon) on the circuit complexity of almost all boolean functions, the best concrete lower bound on the circuit complexity of a particular function has been *3n* for more than 30 years (and was recently improved by only a tiny margin [FGHK15]). In proof complexity [K95] we are still very far from proving lower bounds on the complexity of natural proof systems, and the problem of separating so-called *Frege* and *extended Frege* proof systems is still wide open.

Recent advances have brought, however, new life to some old problems in the area: many classical lower bounds (e.g. Hastad's results on the boolean complexity of parity [Has86]) have been revisited with improved techniques that led to an *average-case hierarchy circuit complexity* [RST15]. New separations results have been obtained [Wil14], and the complexity of important problems has been settled [CDT09].

The division in apperoaches between optimization and computational complexity is no longer that clearcut in recent work: techniques employed for complexity theoretic purposes often also have applications to combinatorial optimization as well: *Fourier analysis of boolean functions* [oD14] often yields learning algorithms as well [LMN93]. Work on the proof complexity of boolean and quantified boolean formulas often has implications (and offers insights) for solvers based on modern techniques (e.g. *conflict driven clause learning*), stimulating the recent interaction between the SAT solving and proof complexity communities (e.g. a recent workshop that took place in 2014 in Banff, Canada). In the other direction, results in combinatorial optimization have found applications to computational complexity: algorithmic (symbolic) methods have been proposed as proof systems in proof complexity. Another example: *submodularity*, a concept that has many practical algorithmic applications to areas such as document summarization, machine learning, is also useful to investigations related to the dichotomy theorem for CSP [CCJ08]. Finally, several of these modern techniques may be applied the same problem, often in conjunction. One example: Fourier analysis of boolean functions and submodularity are both useful in machine learning.

Sadly, such techniques seem largely unknown in the Romanian research community. A primary cause is **their absence from Romanian research in theoretical computer science:** while Romanian scientists abroad have certainly contributed to their development (just one example: distributed submodular optimization [BENW]). To our knowledge there is no (other) research in theoretical computer science inside Romania dealing with similar techniques and objectives.

**D2. Objectives.** The general objectives of the project are to *contribute to the development and application of timely, sophisticated techniques for improving the analysis of problems in Combinatorial Optimization and Computational Complexity.* Building on recent progress by the PI and members of the team [1, 2, 6, 8, IB16], we aim to

- provide a better understanding of the *proof complexity* of combinatorial principles and of noneffective topological proof methods.
- devise methods for investigating the complexity of *smoothed* [ST04, BM15] and *continuously perturbed* combinatorial problems, including versions of satisfiability.
- contribute to the t*heory of submodular optimization* by studying extensions of the notion (and their algorithmic implications), as well as *structural parameters* (e.g. VC dimension) that may lead to different algorithms than the greedy/local search algorithms that currently give the best guarantees for submodular maximization.
- advance the theory of *heapability of integer sequences and partial orders*, along the lines of the *longest increasing subsequence* problem [Rom15] by developing and analyzing notions of generalized width of (random) partial orders and the connection with concepts such as interacting particle systems, Young tableaux and Robinson-Schensted type bijections.

With the exception of the last one (motivated by a problem with exciting preliminary results [6] and significant ramifications), **our objectives are method, rather than problem oriented:** proof complexity, smoothed analysis, submodularity are timely subjects (as witnessed by recent STOC/FOCS/SODA/ICALP proceedings), that attract to a significant worldwide interest. **We intend to contribute to the significant development of these exciting directions.**

An objective that is strategic, rather than technical, is to bring to Romanian research in Theoretical Computer Science a host of problems, approaches and techniques that are synchronized with current top research in the area, and **increase the visibility of Romanian research in top venues in Theoretical Computer Science.**

**D3. Impact.** We aim to work on timely problems that significantly impact theoretical computer science, and make scientific progress that is **competitive at the world level**.

A concurrent aim is **to promote the use of the new, sophisticated, techniques we are concerned with too more applied areas** in Romanian Computer Science: while we could have applications of the techniques we investigate with a more practical orientation, the limited time duration of this project makes such an objective unrealistic. Instead, *we believe that it is important to first build the theoretical expertise (currently missing) that we hope to eventually percolate to the whole Computer Science research community in Romania.* A strategic goal of the project is **to make our team a realistic source of such expertise.**

**D4. Metodology.** We have structured our research plan in two technical Work Packages, plus an extra work package dealing with project management and dissemination. Each technical work package is divided into three activities: two activities that build on existing team capabilities, and an exploratory activity, destined to try new, more risky directions. The ideas falling under these last activities are not vague, ad-hoc ideas (see below): it is simply that we cannot substantiate them just as much through existing publications. In any case, the exploratory activities will comprise about 20% of the total research time of each of the work packages (and are marked light blue)

The last six months of the project (corresponding to the year 2019) are a mix of all types of activities, *intended to enable completion of tall he work started during the first two years.*

The work plan of the project is summarized in the following chart:

| Activity / year | | A | A1 | A2 | A3 | B | B1 | B2 | B3 | C | C1 | C2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2017 | Q1 | | | | | | | | | | | |
| | Q2 | | | | | | | | | | | |
| | Q3 | | | | | | | | | | | |
| | Q4 | | | | | | | | | | | |
| 2018 | Q5 | | | | | | | | | | | |
| | Q6 | | | | | | | | | | | |
| | Q7 | | | | | | | | | | | |
| | Q8 | | | | | | | | | | | |
| 2019 | Q9 | | | | | | | | | | | |
| | Q10 | | | | | | | | | | | |

**Work Package A:** **Advanced techniques in Computational Complexity.**

**Activity A1.** *Search and proof complexity of combinatorial principles.* This activity is based on recent advances by the PI [1,2]. In these papers we proposed the Kneser-Lovasz theorem as a combinatorial principle, showing that in inherits all complexity-theoretic lower bounds from the Pigeonhole Principle, and that it has polynomial size Extended Frege (and quasipolynomial size Frege) proofs. The technical ingredient for this achievement is a proof of the existence of **effective purely combinatorial mathematical proofs of the Kneser-Lovasz theorem,** based on a*n improved version of the Erdos-Ko-Rado theorem, that reduces (for every fixed k) the proof of the Kneser-Lovasz theorem to a finite number of cases.* Our joint result led to new insights in the complexity of search problems, due to our UCSD collaborators: they proved that the 2-dimensional Borsuk-Ulam theorem is complete for the complexity class PPA [ABB15], correcting an error from the classical paper of Papadimitriou [Pap94].

**We intend (likely, in cooperation with our partners at UCSD/UPC) to further develop the applications of combinatorial topology techniques to proof complexity.** Some concrete problems that could be investigated under this framework include:

- *"de-topologization" of results from Combinatorial Topology.* This would entail identifying combinatorial principles that lead to effective combinatorial proofs, in the style of our result from [1]. Several good candidates for target results include Schijver's generalization of the Kneser-Lovasz Theorem and the so-called *Neckace Splitting Theorem* [A87]. The combinatorial principles remain, of course, to be investigated.
- *the search complexity of principles from Combinatorial Topology.* Major open questions in this area include the search complexities of the Octahedral Tucker Lemma and that of the Necklace Splitting Theorem, not known to be complete for the class PPA(D). Furthermore, our work [1] led (journal version, forthcoming in Information and Computation) to new search principles based on our Truncated Tucker lemma that are very hard (hard for the search complexity class PPP [Ais16]).
- *the complexity of miniaturizations of results in Combinatorial Topology.* The prime example is the Truncated Tucker lemma we introduced in [1]. It is unclear what the real power of this result is (e.g. what other results that follow from ordinary Tucker lemma can it prove). Also, it would be useful if other principles in Combinatorial Topogy had versions with efficient propositional translations.
- *bounded arithmetic frameworks for non-effective proof systems.* The framework of *implicit proofs* was recently proposed by Krajicek [K04] as a way to generate hierarchies proof systems going beyond (extended) Frege systems. We aim to better characterize the complexity classes in

Krajicek's EF hierarchy from the standpoint of Bounded Arithmetic, as well as situating existing nonconstructive proof approaches (such as those from Computational Topology) in this hierarchy.

**Activity A2.** *Techniques for analyzing the Smoothed and Continuously Perturbed Complexity of Combinatorial Problems.* Smoothed analysis [ST04, BM15] has been proposed as a hybrid of worst-case and average-case analysis. The main motivation for its definition was to provide intuition for the behavior of real-life algorithms, like the simplex method (or local search methods) that have exponential complexity worst-case instances, but good "practical" behavior. Smoothed analysis has found application to continuous problems, but also to discrete problems such as graph and satisfiability problems [CFKV09]. The work we propose aims to develop methods for analyzing smoothed complexity of problems. Candidate directions are:

- *logical foundations for smoothed complexity.* While the beginnings of a complexity-theoretic study of smoothed complexity exist [BM15], little is known on general frameworks that guarantee tractability in the smoothed sense. It has been conjectured by Roughgarden [Rou14] that the search complexity class PLS might have some sort of appropriately defined "smoothed tractability" property. This appears to be true for problem MAX-CUT [ER14]. The fact that problem PLS and its generalization feature importantly in the theory of Bounded Arithmetic [BK94] , motivate the investigation of logical frameworks for smoothed complexity, an approach which, to our knowledge has not been tried.

- *analyzing algorithm behavior on (random) instances subject to perturbations.* Analyzing algorithm evolution on random instances using differential equations is a well-established technique [MM11] (we have used it in previous work, e.g. [3], related to the area of phase transitions), and is based on modeling the mean trajectory of algorithms with differential equations. One argues, then, using Chernoff-bound type ingredients, that with high probability the system stays close to its mean trajectory. The first extension we have in mind is adapting the differential equations method to smoothed (random) instances. This extension would employ results on stability of systems of differential equations. Moreover, we also propose to deal with (continuous) perturbations associated to smoothed instances by employing stochastic differential equations. If this were possible, it would be t**he first time when SDE's are useful in the analysis of algorithms.**

- *stochastic stability in smoothed analysis.* Stochastic stability is a concept employed in the theory of Markov chains that explores the possibility of making such chains ergodic via small perturbations to transition probabilities. We have already proved some results using the stochastic stability framework [IMR08], and propose its systematic study in this context.

**Activity A3.** *Exploratory approaches in Computational Complexity.* We list under this activity includes several approaches in Computational Complexity (primarily proof/circuit complexity).

- the use of *Fourier methods for the analysis of boolean functions* [oD14] in complexity theory. Although tthis is not documented by previous work, the P.I. is well-versed in such methods, as they are useful to prove the existence of phase transitions in combinatorial optimization.

- *complexity results for other proof systems* (e.g. algebraic proof systems, such as polynomial calculus, cutting planes, ideal proof system), as well as proof *systems for problems with complexity different from NP (e.g. QBF).*

- approaches to lower bounding the complexity of boolean functions. An especially attractive problem is improve the recent lower bounds on the complexity of boolean functions [FGHK15].

**Work Package B: Advanced techniques in Combinatorial Optimization.**

***Activity B1.*** *(Generalized) submodularity in complexity and approximation. Concepts, parameters, applications.* Submodularity is a discrete analog of convexity that encodes the concept of *diminishing returns*, and plays nowadays a central role in a variety of problems related to clustering, mobile computing, data summarization, machine learning, etc. Submodular has featured as a subject of special workshops and tutorials at leading Machine Learning and general Artificial Intelligence conferences (ICML, IJCAI). Efficient algorithms for various variants of submodular optimization (minimization, maximization, distributed algorithms) are known [BV14, Sch00, BENW15]. On the other hand there are various aspects of submodular optimization that go beyond the classical framework that we propose to study:

- optimization using nonstandard objective functions. Motivated by applications in Computational Biology, in [8] we have studied the problem of minimum entropy submodular minimization. Many open problems related to the strength of the methods in [8] exist. We propose to continue this direction and adapt it to other nonstandard objective functions for submodular optimizations (and other algorithms - for instance we believe that the so-called continuous greedy algorithm, that offers better guarantees for submodular maximization, should be analyzed for problem in [8] as well.

- relaxations and extensions of the notion of submodularity. Various practical problems yield concepts related to but subdifferent from the classical concept of submodularity. We give only one example: result diversification in web search and document summarization [BLY12] (see also sarXiv.org preprint 1401.6697 for the associated variant of submodularity, called weak submodularity). These variations may have different algorithms/guarantees with respect to the classical version. We propose their investigation.

- <u>parameters, algorithms and guarantees for submodular optimization.</u> Monotone, integral, submodular functions (so called integer polymatroids) can be viewed as an extension of the set cove r problem subject to independence constraints of matroid type. The theory of approximability of the set cover problem extends in a natural way to submodular set cover. On the other hand for set cover a different set of algorithms provides better guarantees in the presence of a bound on a structural parameter, the Vapnik-Chervonenkis dimension [BG95]. Many practical settings, including geometric coverage scenarios in mobile computing, impose such restrictions on the resulting instances of set cover. We propose to extend the theory of Vapnik-Chervonenkis dimension to integral polymatroids and analyze the resulting algorithms.

**<u>Activity B2.</u>** *Heapability of (random) partial orders: combinatorial and algorithmic approaches.*

A sequence of integers is *heapable* [BHMZ11] if it can be inserted into a binary tree (not necessarily complete) respecting the min-heap property. We briefly discussed existing research on the heapability of integer sequences in section C2. Heapability can be seen as a relaxation of the notion of increasing sequence. The longest increasing subsequence problem is a classical problem in combinatorial optimization with a rich theory [Rom15] and connections to many areas of science, including random matrix theory and statistical physics. The concept can be naturally extended to partial orders [IB16], and it leads to an extension of the width parameter for such partial orders.

We aim to scale to heapable partial orders the rich theory of LIS [Rom15], including the following:
- the complexity of the l*<u>ongest heapable subsequence</u>* problem, left open in [BHMZ11].
- *<u>address the conjecture in [6]</u>*, concerning the scaling constant of the decomposition into a minimal number of heapable subsequences.
- study heapability of *<u>random partial orders</u>* [IB16], for which scaling behavior is not elucidated.
- the study of *<u>heapability of intervals</u>*. This is related to Gallai type results for intervals and known results on the longest subsequence of increasing intervals [JSW90].
- the study of *<u>Hammersley-type systems</u>* for LIS/heapability of partial orders. Even for LIS of intervals, for which the scaling behavior is known [JSW90], such a system has **not** been studied.
- *<u>multiset generalizations of Young tableaux, hook formulas and Robinson-Schensted algorithms.</u>* The LIS problem is related to the topics above via the so-called Robinson-Schenstd-(Knuth) algorithm. Insertion of elements into the first row of Young tableaux corresponds to the (ordinary) Hammersley process. In [6] we defined a multiparticle generalization of Hammersley process naturally associated to heapability of integer sequences (see [BGGS16], where the process is generalized under the name *Hammersley's trees*). It is natural to inquire on the existence of

generalizations of Young tableaux, hook formulas, R-S processes dual to this generalizations. The problem has further far reaching connections in algebraic combinatorics that could be explored.

**Activity B3.** *Exploratory techniques in Combinatorial Optimization.* Under this activity we could deal with topics such as

- the analysis of (rapidly mixing) Markov chains. Note that the PI has research experience in this direction. This topic also connects well with the use of Fourier theoretic methods outlined in Activity C3, as well as the theory of perturbations of dynamical systems. It would be interesting if the techniques developed for Activity A2 would find applications to this setting as well.

**Work Package C: Management and Dissemination**

**Activity C1: Dissemination.** Dissemination will be realized in the following three ways:

- publication of the research outcomes and deliverables on the project web page and (if appropriate) on preprint servers such as arXiv.org or Electronic Colloquium on Computational Complexity. Of course, these results will be eventually published in appropriate venues (journals or conferences).
- participation to scientific conferences. We have allocated an appropriate budget to this goal.
- we intend to organize a workshop on proof complexity, satellite to SYNASC conference, scheduled to take place in September 2017 in Timișoara. **We intend this workshop to bring world-class researchers in proof complexity to Timișoara**, and will use the funding to bring some of the leading researchers as invited speakers.

**Activity C2: Reporting, administration and management.** This is a continuous activity. The PI will take care of all the reporting requirements associated with the administration and management of the project. This includes maintenance of a web page for the project, and all the reporting.

**Deliverables:** All papers written as part of the project, together with the periodic scientific report, will be published on the project webpage.

**Research Team Composition and Capabilities:** The team is composed of **four or five members**, the P.I. (Gabriel Istrate), two senior researchers (Cosmin Bonchis and Adrian Craciun) and *one or two masters/Ph.D. student(s) TBD.*

- The Principal Investigator, Gabriel Istrate, is an established researcher in Algorithms and Computational Complexity, **as described in sections B and C.**
- Cosmin Bonchiș (a senior researcher member of the team) and the PI have recently published joint papers [6,8] that are directly relevant to the scope of the project, that of submodular minimum entropy optimization. Cosmin Bonchis will work with the PI primarily on the aspects related to combinatorial optimization (WP2 below).

- the other member of the team, Adrian Crăciun has a solid background in logic, proof complexity, SAT solving and theorem proving. He is the coauthor (together with the PI) of papers [1,2] on proof complexity, and will collaborate with the PI primarily on the complexity-theoretic aspects (WP 1).
- we have set up one full time position for masters/Ph.D. students. **This position is left open.** Depending on the availability and quality, we may choose to involve two students (rather than one) on the project. Let us remark that **the P.I. has a habilitation to direct Ph.D. studies, (as well as teaching masters' level courses), therefore he has access to students at both levels.**

We envision for all members of the team an involvement in both work packages of the project. Given the small size of the team and the reduced duration of the project, we believe this choice ensures cohesion and contributes to the overall success of the project.

**Risk mitigation** The risks associated with the project arise from a) the ambitious scientific goals, combined with b) the fairly short duration of the project (30 months), and c) uncertainty in the availability and timeliness of funding.

**a) Ambitious scientific goals:** As stated, the research agenda is **very ambitious:** rather than confine ourselves to a limited set of research objectives we stated a maximal agenda, especially given the available time frame. This is, we believe, <u>normal for a exploratory research project.</u> The scientific success of the project should be measured, we believe, by the extent to which we made significant appropriate progress on the scientific objectives stated above, **especially when compared to projects with a similar level of financing in Western institutions.** <u>We mitigate this risk</u> by the appropriate division of time between lower risk and high risk activities (80/20). Also, team involvement in both work packages will act as a mitigation factor.

**b) The fairly short duration of the project:** It is, frankly, difficult to guarantee a certain number of publications in the short time frame of the project: such an approach often leads to compromises in quality of said publications, **an approach we don't want to take.** <u>We strongly believe that in an exploratory research quality is measured by the international relevance of the resulting work, rather than sheer number.</u>

**c) Uncertainty in funding.** Frankly, our project is critically dimensioned: a small team, competitively compensated (with respect to market in informatics), with a short duration. **Any issues in funding (such as those experienced since 2011) could harm the success of the project. T**his factor is beyond our control: we urge the funding agency to stick by its stated goals.

**D5. Ethical aspects (if appropriate)**

The proposal is a foundational one, with primarily a theoretical content, and does not pose significant ethical problems.

**D6. Resources and budget**

**Supporting infrastructure.** The e-Austria Research Institute is well-positioned to deal with the computational requirements of the project: in addition to the existing computing infrastructure, if simulations will require it, the team can benefit from access to the HPC infrastructure of the West University of Timisoara, see https://erris.gov.ro/UVT-HPC-Center for details. The University is a founding member of IEAT.

**Budget.** The project team will cover the estimated duration of each activity and subactivity as described in the work plan. Given the available sum, we decided to forgo the acquisition of new equipment, and only budget modest amounts for materials and supplies as part of logistics. Given that much of the relevant publication in Theoretical Computer Science takes place in scientific conferences, we have allocated about 16.000 euro on travel expenses for the duration of the project. Of these, an important amount in 2017 (about 4000 euro) is reserved for travel and lodging for the invited speakers to the workshop we organize. There is one type of budgeted expense which will be subcontracted (hence it does not count into establishing the overhead): 4.500 RON for the meal/coffee break expenses associated to the workshop. **Due to shortcomings of the electronic system for proposal submission, it appears there as Logistics/Equipment (the only category not taken into account in the system when computing the overhead). This is incorrect: we intend no equipment acquisition, but subcontracting.**

Budget Breakdown (euro, for the whole project, assuming an exchange rate 1 Euro = 4.5 RON)

| **Budget chapter (expenses)** | **Total budget 2017 - 2019** |
|---|---|
| **Personnel** | **133333.33** |
| **Logistical** | **2777.78** |
| **Travel** | **16000** |
| **Indirect** | **36777.78** |
| **TOTAL** | **188888.89** |

Budget Breakdown (lei):

| **Budget chapter (expenses)** | **2017 (lei)** | **2018 (lei)** | **2019 (lei)** | **Total (lei)** |
|---|---|---|---|---|
| **Personnel** | **240000** | **240000** | **120000** | **600000** |
| **Logistical** | **7500** | **3000** | **2000** | **12500** |
| **Travel** | **40000** | **22000** | **10000** | **72000** |
| **Indirect** | **66250** | **66250** | **33000** | **165500** |
| **Total** | **353750** | **331250** | **165000** | **850000** |

**D7. Bibliography** *(maximum 2 pages)*


[A87] Alon, Noga. "Splitting necklaces." *Advances in Mathematics* 63.3 (1987): 247-253.

[Ais16] J. Aisenberg. *Ph. D. Thesis, University of California, San Diego*, 2016.

[ABB15] J. Aisenberg, M.L. Bonet, S. Buss. 2-D Tucker is PPA Complete. *Manuscript*, available on Sam Buss's webpage.

[BENW15] R. Barbosa, R., A. Ene, H.L. Nguyen, & J. Ward. The power of randomization: Distributed submodular maximization on massive datasets. In *International Conference on Machine Learning* (ICML'15) pp. 1236-1244, 2015.

[BG95] Brönnimann, Hervé, and Michael T. Goodrich. "Almost optimal set covers in finite VC-dimension." *Discrete & Computational Geometry* 14.4 (1995): 463-479.

[BK94] S. Buss, and J. Krajicek. "An application of Boolean complexity to separation problems in bounded arithmetic." *Proceedings of the London Mathematical Society* 69.1 (1994): 1-21.

[Bri93] G. Brightwell. "Models of random partial orders." *Surveys in combinatorics* 5383 (1993).

[BGGS16] Basdevant, A. L., Gerin, L., Gouere, J. B., & Singh, A. (2016). From Hammersley's lines to Hammersley's trees. *arXiv preprint arXiv:1605.02981*.

[BHMZ11] Byers, J., Heeringa, B., Mitzenmacher, M., & Zervas, G. *Heapable sequences and subseqeuences.* In Proceedings of the Meeting on Analytic Algorithmics and Combinatorics (ANALCO 2011), pp. 33-44.

[BLY12] Borodin, Allan, Hyun Chul Lee, and Yuli Ye. "Max-sum diversification, monotone submodular functions and dynamic updates." *Proceedings of the 31st ACM PODS*. ACM, 2012.

[BM15] Bläser, Markus, and Bodo Manthey. "*Smoothed complexity theory*." *ACM Transactions on Computation Theory (TOCT)* 7.2 (2015): 6.

[BV14] A. Badanidiyuru, A., & J. Vondrák, J. Fast algorithms for maximizing submodular functions. In *Proc. of the 25th ACM-SIAM Symposium on Discrete Algorithms,* 1497-1514, 2014.

[CCJ08] D. Cohen, M. Cooper, and P. Jeavons. "Generalising submodularity and Horn clauses: Tractable optimization problems defined by tournament pair multimorphisms." *Theoretical Computer Science* 401.1 (2008): 36-51.

[CFKV09] Coja-Oghlan, A., Feige, U., Frieze, A., Krivelevich, M., & Vilenchik, D. On smoothed k-CNF formulas and the Walksat algorithm. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*, pp. 451-460, 2009.

[CDT09] X. Chen, X. Deng, and S.H. Teng. "Settling the complexity of computing two-player Nash equilibria." *Journal of the ACM (JACM)* 56.3 (2009): 14.

[ER14] M. Etscheid, H. Roglin**.** *Smoothed analysis of local search for the maximum-cut problem.* Proc. of 25th ACM-SIAM Symposium on Discrete Algorithms (SODA'14), 882-889, 2014.

[FGHK15] M.Find, A. Golovnev, E.A. Hirsch, A.Kulikov. A better-than-3n lower bound for the circuit complexity of an explicit function. *Electronic Colloq. Comp. Complexity* 22: 166 (2015)

[GV15] T. Gowers, E. Viola. The communication complexity of interleaved group products. In STOC 2015, *Proc. of the 47th annual ACM Symposium on Theoty of Computing:* 351-360, 2015.

[Has86] Hastad, Johan. "Almost optimal lower bounds for small depth circuits." *Proceedings of the eighteenth annual ACM symposium on Theory of computing*. ACM, 1986.

[IB16] G. Istrate, C. Bonchiș. Heapability, interacting particle systems, partial orders: results and open problems. In *Proceedings of DCFS'2016 (forthcoming),* Springer Verlag, 2016.

[IMR08] G. Istrate, G., M. Marathe, S. Ravi, S. Adversarial scheduling analysis of game-theoretic models of norm diffusion. In *Computability in Europe* (CIE'08) pp. 273-282, Springer Verlag.

[JSW90] J. Justicz, E. Schneireman, and P. Winkler. "Random intervals." *The American mathematical monthly* 97.10 (1990): 881-889.

[K95] J. Krajicek *Bounded arithmetic, propositional logic and complexity theory*. Cambridge University Press, 1995.

[K04] Krajíček, J. (2004). Implicit proofs. *The Journal of Symbolic Logic*, *69*(02), 387-397.

[LMN93] Linial, Nathan, Yishay Mansour, and Noam Nisan. "Constant depth circuits, Fourier transform, and learnability." *Journal of the ACM* 40.3 (1993): 607-620.

[MM11] C. Moore, S. Mertens. *The nature of computation*. Oxford University Press, 2011.

[oD] R. O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.

[Pap94] Papadimitriou, C. H. (1994). On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and system Sciences*, *48*(3), 498-532.

[Rom15] D. Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2015.

[Rou14] T. Roughgarden. *Beyond Worst-Case Analysis (BWCA) Course, Stanford Lecture 13: Smoothed local search,2014.* Video available from the author's webpage and Youtube.

[RST15] Rossman, Benjamin, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for Boolean circuits." *56th FOCS, 2015.*

[Sch00] A. Schrijver, *A c*ombinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B* 80.2 (2000): 346-355.

[ST04] D. Spielman & S.H. Teng. *S*moothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, *51*(3), 385-463, 2004.

[TV07] T. Tao, V. Vu The condition number of a randomly perturbed matrix. In *STOC'07: Proceedings of the 39th annual ACM Symposium on the Theory of Computing*, 248-255, 2007.

[Wil14] R. Williams, Ryan. "Nonuniform ACC circuit lower bounds." *Journal of the ACM* 61.1 (2014): 2.