The complexity of deciding characteristic formulae in van Glabbeek's branching-time spectrum*

Luca Aceto, Antonis Achilleos, Aggeliki Chalki, and Anna Ingólfsdóttir

Department of Computer Science, Reykjavik University, Reykjavik, Iceland
`luca@ru.is`, `antonios@ru.is`, `angelikic@ru.is`. `annai@ru.is`

Introduction. In concurrency theory, *characteristic formulae* serve as a bridge between model checking and preorder or equivalence checking. At an intuitive level, a characteristic formula provides a complete logical characterization of the behaviour of a process with respect to some notion of behavioural equivalence or preorder. For example, consider the widely used bisimulation equivalence relation [1]; Hennessy and Milner have shown in [2] that, under a mild finiteness condition, two processes are bisimilar if and only if they satisfy the same Hennessy-Milner logic (**HML**) formulae. Apart from its intrinsic theoretical interest, this seminal logical characterization of bisimilarity means that, when two processes are *not* bisimilar, there is always a **HML** formula that distinguishes between them. However, using the Hennessy-Milner theorem to show that two processes are bisimilar would involve verifying that they satisfy the same **HML** formulae and there are infinitely many of those. This is where characteristic formulae come into play. A **HML** formula φ is characteristic for process p, if every process q satisfies φ iff p and q are bisimilar. As a consequence, one can decide bisimulation equivalence between p and q by finding the characteristic formula $\chi(p)$ for p and checking whether $q \models \chi(p)$, that is a model-checking problem. Characteristic formulae, thus, allow one to reduce bisimilarity checking to model checking. Conversely, Boudol and Larsen studied in [3] the problem of characterizing the collection of modal formulae for which model checking can be reduced to equivalence checking. See [4, 5, 6] for other contributions in that line of research. The aforementioned articles showed that characteristic formulae coincide with those that are consistent and prime. (A formula is prime if whenever it entails a disjunction $\varphi_1 \vee \varphi_2$, then it must entail φ_1 or φ_2.) Moreover, characteristic formulae with respect to the bisimulation relation coincide with the formulae that are consistent and complete, where a modal formula φ is complete, when for every modal formula ψ on the same propositional variables as φ, we can derive from φ either ψ or its negation. When one wants to reduce model checking to equivalence checking, the study of the complexity of identifying characteristic formulae modulo bisimilarity within (extensions of) **HML** is of relevance and has been addressed in [7, 8]. Typically, checking whether a formula is characteristic modulo bisimilarity has the same complexity as validity.

We described characteristic formulae using the example of bisimilarity, as it is the relation between processes that underlies the seminal Hennessy-Milner theorem and was used in much of the above-mentioned work. However there are a plethora of other preorder and equivalence relations that classify processes according to other possible behaviours; these and their logical characterizations have been extensively studied in concurrency theory—see e.g. [9, 10]. In this work, we address the complexity of deciding and finding characteristic formulae with respect to four different preorders in van Glabbeek's branching-time spectrum, namely *simulation* ($\leq_S$), *complete simulation* ($\leq_{CS}$), *ready simulation* ($\leq_{RS}$), and *trace simulation* ($\leq_{TS}$) [9]. A binary relation R over states in a labelled transition system is a simulation if it satisfies the following

*This work has been funded by the projects "Open Problems in the Equational Logic of Processes (OPEL)" (grant no 196050) and "Mode(l)s of Verification and Monitorability" (MoVeMent) (grant no 217987) of the Icelandic Research Fund.

condition: if pRq and $p \xrightarrow{a} p'$, then there is some q' such that $q \xrightarrow{a} q'$ and $p'Rq'$. Let p and q be states in a labelled transition system. We say that

1. p is simulated by q, written $p \leq_S q$, if there is a simulation relation R such that pRq;

2. p is complete simulated by q, written $p \leq_{CS} q$, if there is a simulation relation R that only relates deadlocked states with deadlocked states such that pRq. (A state is deadlocked if it does not have any outgoing transitions.)

3. p is ready simulated by q, written $p \leq_{RS} q$, if there is a simulation relation R that only relates states that have the same sets of initial actions such that pRq. (The set of initial actions of a state is the collection of actions that label its outgoing transitions.)

4. p is trace simulated by q, written $p \leq_{TS} q$, if there is a simulation relation R that only relates states that have the same sets of traces such that pRq. (The set of traces of p is the set of all possible sequences of actions that can be observed by executing p.)

It is well-known that simulation is the coarsest of the four preorders and $\leq_{TS} \subsetneq \leq_{RS} \subsetneq \leq_{CS} \subsetneq \leq_S$. We denote by $\mathcal{L}_S$, $\mathcal{L}_{CS}$, $\mathcal{L}_{RS}$, and $\mathcal{L}_{TS}$, respectively, the fragments of **HML** that characterize these four preorders—see [9, 6].

In the sequel, we first mention known complexity results on deciding preorders $\leq_S$, $\leq_{CS}$, and $\leq_{RS}$, respectively, and we present our results on the complexity of deciding $\leq_{TS}$. Then, we state propositions and theorems establishing the complexity of identifying and finding characteristic formulae for the aforementioned preorders.

The complexity of deciding simulation, complete simulation, ready simulation, and trace simulation. Let $\leq \in \{\leq_S, \leq_{CS}, \leq_{RS}\}$. Given two finite processes p and q, deciding whether $p \leq q$ can be done in polynomial time [9]. To the best of our knowledge, the complexity of deciding the trace simulation preorder has not been examined yet. In the following propositions, we show that the situation is different when we consider $\leq_{TS}$.

Proposition 1. *Deciding $\leq_{TS}$ on finite processes is* PSPACE*-complete under polynomial-time Turing reductions.*

Proposition 2. *Deciding $\leq_{TS}$ on loop-free finite processes is* coNP*-complete under polynomial-time Turing reductions.*

Note that we use polynomial-time oracle reductions instead of the more standard Karp reductions between decision problems. In both cases, hardness is established by showing that the trace equivalence of two processes can be decided by making two oracle calls to the problem of deciding the trace simulation preorder. Since deciding trace equivalence is PSPACE- and coNP-hard under Karp reductions on finite and loop-free finite processes, respectively [11, 12], we obtain our hardness results. Moreover, we provide a clear proof of the coNP-hardness of trace equivalence on loop-free finite processes, which results from the coNP-hardness of equivalence between star-free regular expressions that was shown in [12, Theorem 2.7(1)]. Membership in PSPACE can be easily proven for Proposition 1. To prove Proposition 2, we present an NP algorithm for deciding $\not\leq_{TS}$ on loop-free finite processes.

The complexity of deciding characteristic formulae modulo simulation, complete simulation, and ready simulation. We determine the complexity of deciding whether a formula is characteristic for the preorders $\leq_S$, $\leq_{CS}$, and $\leq_{RS}$ as stated in the following theorem.

Theorem 1. *Let $\mathbf{\Lambda}$ be one of the modal logics $\mathcal{L}_S$, $\mathcal{L}_{CS}$, $\mathcal{L}_{RS}$, which characterizes the respective* *preorder $\leq \in \{\leq_S, \leq_{CS}, \leq_{RS}\}$. Given a formula φ in $\mathbf{\Lambda}$, the problem of deciding whether φ is characteristic within $\mathbf{\Lambda}$ for a loop-free finite process can be solved in polynomial time.*

To prove Theorem 1, we make use of the fact that characteristic formulae within any of the logics $\mathcal{L}_S$, $\mathcal{L}_{CS}$, $\mathcal{L}_{RS}$, and $\mathcal{L}_{TS}$, are exactly the consistent and prime ones [6]. For each one of $\mathcal{L}_S, \mathcal{L}_{CS}$, and $\mathcal{L}_{RS}$, we will present a polynomial-time algorithm that, given a formula φ, constructs a labelled graph with both node- and edge-labels that represents φ. Let $|\varphi|$ denote the size of φ, i.e. the number of symbols that appear in φ. The constructed graph has a polynomial size in $|\varphi|$ and resembles term representation. By exploring this graph, we can efficiently decide whether formula φ is prime. In fact, we can also find a process p for which the input formula φ is characteristic within $\mathbf{\Lambda}$. Therefore, we also obtain the following corollaries.

Corollary 1. *Given $\varphi \in \mathbf{\Lambda}$, where $\mathbf{\Lambda} \in \{\mathcal{L}_S, \mathcal{L}_{CS}, \mathcal{L}_{RS}\}$, there is a polynomial-time algorithm that decides whether φ is characteristic for some p within $\mathbf{\Lambda}$, and in the case of a positive answer, the algorithm outputs p.*

Corollary 2. *Given $\varphi \in \mathbf{\Lambda}$, where $\mathbf{\Lambda} \in \{\mathcal{L}_S, \mathcal{L}_{CS}, \mathcal{L}_{RS}\}$, and a loop-free finite process p, the problem of verifying whether φ is characteristic for p can be solved in polynomial time.*

The complexity of finding characteristic formulae modulo simulation, complete simulation, ready simulation, and trace simulation. Given a process p, the problem of constructing the characteristic formula for p has been studied for a variety of preorders and equivalences [13, 4, 14, 15]. To resolve the complexity of the problem we consider two different ways of representing formulae and measuring their size. Given a formula φ, the first approach is to write φ explicitly and define its size to be equal to the number of symbols that appear in φ as above; the second one involves representing φ using recursive equations called declarations, and defining its declaration-size as the number of required declarations. We denote the former by $|\varphi|$ and the latter by $\mathrm{decl}(\varphi)$. For example, formula $\varphi = \langle a\rangle(\langle a\rangle\mathbf{tt} \wedge \langle b\rangle\mathbf{tt}) \wedge \langle b\rangle(\langle a\rangle\mathbf{tt} \wedge \langle b\rangle\mathbf{tt})$ has size $|\varphi| = 13$ and declaration-size $\mathrm{decl}(\varphi) = 2$, as it can be represented by the equations $\varphi = \langle a\rangle\varphi_1 \wedge \langle b\rangle\varphi_1$ and $\varphi_1 = \langle a\rangle\mathbf{tt} \wedge \langle b\rangle\mathbf{tt}$. We prove the following two propositions.

Proposition 3. *Let $\mathbf{\Lambda}$ be one of the modal logics $\mathcal{L}_S$, $\mathcal{L}_{CS}$, $\mathcal{L}_{RS}$. Given a loop-free finite process p, finding the characteristic formula $\chi(p)$ for p within $\mathbf{\Lambda}$ is:*

1. *in P, if $\chi(p)$ is given as a set of declarations, and*
2. *NP-hard under polynomial-time Turing reductions, if $\chi(p)$ is explicitly written.*

Proposition 4. *Given a loop-free finite process p, finding the characteristic formula $\chi(p)$ for p within $\mathcal{L}_{TS}$ is NP-hard under polynomial-time Turing reductions, regardless of the representation of $\chi(p)$.*

Thus, by assuming that $\chi(p)$ is given as a set of declarations, we isolate a sharp difference between the complexity of finding $\chi(p)$ within any $\mathbf{\Lambda} \in \{\mathcal{L}_S, \mathcal{L}_{CS}, \mathcal{L}_{RS}\}$, and finding $\chi(p)$ within $\mathcal{L}_{TS}$.

Conclusions. Finally, we mention some problems that still remain open and whose solutions we are currently pursuing. First, we haven't resolved the complexity of deciding characteristic formulae for the trace simulation preorder. Yet another relevant problem is the complexity of deciding whether a **HML** formula φ is logically equivalent to a formula φ' in $\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is one of $\mathcal{L}_S$, $\mathcal{L}_{CS}$, $\mathcal{L}_{RS}$, and $\mathcal{L}_{TS}$, so that we can answer whether φ is characteristic by applying our results on φ'. Moreover, we want to address all the aforementioned problems for other relations in van Glabbeek's spectrum and over finite processes with loops.

References

[1] R. Milner, *Communication and Concurrency.* Prentice Hall, 1989.

[2] M. Hennessy and R. Milner, "Algebraic laws for nondeterminism and concurrency," *J. ACM*, vol. 32, no. 1, pp. 137–161, 1985. [Online]. Available: https://doi.org/10.1145/2455.2460

[3] G. Boudol and K. G. Larsen, "Graphical versus logical specifications," *Theor. Comput. Sci.*, vol. 106, no. 1, pp. 3–20, 1992. [Online]. Available: https://doi.org/10.1016/0304-3975(92)90276-L

[4] B. Steffen and A. Ingólfsdóttir, "Characteristic formulae for processes with divergence," *Inf. Comput.*, vol. 110, no. 1, pp. 149–163, 1994. [Online]. Available: https://doi.org/10.1006/inco.1994.1028

[5] L. Aceto, I. Fábregas, D. de Frutos-Escrig, A. Ingólfsdóttir, and M. Palomino, "Graphical representation of covariant-contravariant modal formulae," in *Proc. of EXPRESS 2011*, ser. EPTCS, vol. 64, 2011, pp. 1–15. [Online]. Available: https://doi.org/10.4204/EPTCS.64.1

[6] L. Aceto, D. D. Monica, I. Fábregas, and A. Ingólfsdóttir, "When are prime formulae characteristic?" *Theor. Comput. Sci.*, vol. 777, pp. 3–31, 2019. [Online]. Available: https://doi.org/10.1016/j.tcs.2018.12.004

[7] A. Achilleos, "The completeness problem for modal logic," in *Proc. of LFCS 2018*, ser. Lecture Notes in Computer Science, vol. 10703. Springer, 2018, pp. 1–21. [Online]. Available: https://doi.org/10.1007/978-3-319-72056-2_1

[8] L. Aceto, A. Achilleos, A. Francalanza, and A. Ingólfsdóttir, "The complexity of identifying characteristic formulae," *J. Log. Algebraic Methods Program.*, vol. 112, p. 100529, 2020. [Online]. Available: https://doi.org/10.1016/j.jlamp.2020.100529

[9] R. J. van Glabbeek, "The linear time - branching time spectrum I," in *Handbook of Process Algebra.* North-Holland / Elsevier, 2001, pp. 3–99. [Online]. Available: https://doi.org/10.1016/b978-044482830-9/50019-9

[10] D. de Frutos-Escrig, C. Gregorio-Rodríguez, M. Palomino, and D. Romero-Hernández, "Unifying the linear time-branching time spectrum of process semantics," *Log. Methods Comput. Sci.*, vol. 9, no. 2, 2013. [Online]. Available: https://doi.org/10.2168/LMCS-9(2:11)2013

[11] P. C. Kanellakis and S. A. Smolka, "CCS expressions, finite state processes, and three problems of equivalence," *Inf. Comput.*, vol. 86, no. 1, pp. 43–68, 1990. [Online]. Available: https://doi.org/10.1016/0890-5401(90)90025-D

[12] H. B. H. III, D. J. Rosenkrantz, and T. G. Szymanski, "On the equivalence, containment, and covering problems for the regular and context-free languages," *J. Comput. Syst. Sci.*, vol. 12, no. 2, pp. 222–268, 1976. [Online]. Available: https://doi.org/10.1016/S0022-0000(76)80038-4

[13] S. Graf and J. Sifakis, "A modal characterization of observational congruence on finite terms of CCS," *Information and Control*, vol. 68, no. 1-3, pp. 125–145, 1986. [Online]. Available: https://doi.org/10.1016/S0019-9958(86)80031-6

[14] L. Aceto, A. Ingólfsdóttir, M. L. Pedersen, and J. Poulsen, "Characteristic formulae for timed automata," *RAIRO Theor. Informatics Appl.*, vol. 34, no. 6, pp. 565–584, 2000. [Online]. Available: https://doi.org/10.1051/ita:2000131

[15] L. Aceto, A. Ingólfsdóttir, P. B. Levy, and J. Sack, "Characteristic formulae for fixed-point semantics: a general framework," *Math. Struct. Comput. Sci.*, vol. 22, no. 2, pp. 125–173, 2012. [Online]. Available: https://doi.org/10.1017/S0960129511000375