Semi-quantitative Abstraction and Analysis of Chemical Reaction Networks (Extended Abstract)

Milan Češka[1(✉)] and Jan Křetínský[2]

[1] Brno University of Technology, FIT, IT4I Centre of Excellence, Brno, Czech Republic
ceskam@fit.vutbr.cz

[2] Technical University of Munich, Munich, Germany

Introduction. Chemical Reaction Networks (CRNs) are a versatile language widely used for *modelling and analysis* of biochemical systems [4] as well as for high-level *programming* of molecular devices [1,14]. Motivated by numerous potential applications ranging from system biology to synthetic biology, various techniques allowing simulation and formal analysis of CRNs have been proposed [2,7,10,13], and embodied in the design process of biochemical systems [6,11,12]. The time-evolution of CRNs is governed by the Chemical Master Equation (CME), which describes the probability of the molecular counts of each chemical species. Many important biochemical systems lead to complex dynamics that includes *state space explosion, stochasticity, stiffness, and multimodality* of the population distributions [9,15], and that fundamentally limits the class of systems the existing techniques can effectively handle. More importantly, biologist and engineers often seek for plausible explanations why the system under study has or has not the required behaviour. In many cases, a set of system simulations/trajectories or population distributions are not sufficient and the ability to provide an accurate explanation for the temporal or steady-state behaviour is another major challenge for the existing techniques.

In order to cope with the computational complexity of the analysis and in order to obtain explanations of the behaviour, we shift the focus from quantitatively precise results to a more qualitative analysis, closer to how a human would behold the system. Yet we insist on providing at least rough timing information on the behaviour as well as rough classification of probability of different behaviours at the extent of "very likely", "few percent", "barely possible", so that we can conclude on issues such as time to extinction or bimodality of behaviour. This gives rise to our *semi-quantitative* approach. We stipulate that analyses in this framework reflect quantities in orders of magnitude, both for time duration and probabilities, but not more than that. This paradigm shift is

This work has been accepted to the 31st International Conference on Computer-Aided Verification (CAV'19). The full version of the paper is available at [3]. The work has been supported by the Czech Science Foundation grant No. GA19-24397S, the IT4Innovations excellence in science project No. LQ1602, and the German Research Foundation (DFG) project KR 4890/2-1 "Statistical Unbounded Verification".

L. Bortolussi and G. Sanguinetti (Eds.): CMSB 2019, LNBI 11773, pp. 337–341, 2019.
https://doi.org/10.1007/978-3-030-31304-3_22

reflected on two levels: (1) We abstract systems into semi-quantitative models. (2) We analyse systems in a semi-quantitative way. While each of the two can be combined with a traditional abstraction/analysis, when combined together they provide powerful means to understand systems' behaviour with virtually no computational cost.

Semi-quantitative Models. The states of the models contain information on the current amount of objects of each species as an interval spanning often several orders of magnitude, unless instructed otherwise. For instance, if an amount of a certain species is to be closely monitored (as a part of the input specification/property of the system) then this abstraction can be finer. Similarly, whenever the analysis of a previous version of the abstraction points to the lack of precision in certain states, preventing us to conclude which of the possible behaviours is prevalent, the corresponding refinement can take place. Further, the rates of the transitions are also captured only with such imprecision. The crucial point allowing for existence of such models that are small, yet faithful, is our concept of *acceleration*. It captures certain *sequences* of transitions. It eliminates most of the non-determinism that paralyses other types of abstractions, which are too over-approximative, unable to conclude anything, but safety properties.

Semi-quantitative Analysis. Instead of performing exact transient or steady-state analysis, we can consider most probable transitions and then carefully lift this to most probable temporal behaviours. Technically, this is done by *alternating between transient and steady-state analysis* where only some rates and transitions are taken into account at different iterations. In order to further facilitate the resulting insight of the human on the result of the analysis, we provide an algorithm to perform this analysis with virtually no computation effort and thus possibly manually. The trivial computations immediately pinpoint why certain behaviours occur. Moreover, less likely behaviours can also be identified easily, to any desired degree of probability (dozens of percent, promilles etc.).

Summary. The first step of our approach yields tiny models, allowing for a synoptic observation of the model; due to their size these models can be either analysed easily using standard means, or can be subject to the second step. The second step provides an efficient approximative analysis, which is also very illustrative due to the limited use of quantities. It can be applied to any system; however, it is particularly interesting in connection with the models coming from the first step since (i) no extra effort (size, computation) is wasted on overly precise treatment that is ignored by the other step, and (ii) together they yield an understandable explanation of the behaviour. An entertaining feature of this paradigm is that the stiffer (with rates at hugely different time scales) the system is the easier it is to analyse.

To demonstrate the capabilities of our approach, we consider three challenging and biologically relevant case studies that have been used in literature to evaluate state-of-the-art methods for the CRN analysis. It has been shown that many approaches fail, either due to time-outs or incapability to capture

Table 1. Gene expression. The rates are in h^{-1}.

$D_{off} \xrightarrow{0.05} D_{on}$	$D_{on} \xrightarrow{0.05} D_{off}$	$D_{on} \xrightarrow{10} D_{on} + RNA$	$RNA \xrightarrow{1} \emptyset$
$RNA \xrightarrow{4} RNA + P$	$P \xrightarrow{1} \emptyset$	$P + D_{off} \xrightarrow{0.0015} P + D_{on}$	

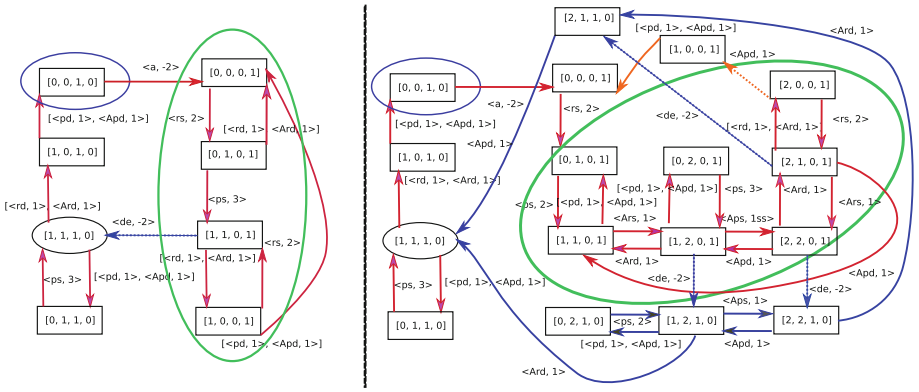

Fig. 1. Pruned abstraction for the gene expression model using the coarse discretisation (left) and after the refinement (right). The state vector is [P, RNA, D_{off}, D_{on}]. (Color figure online)

differences in behaviours, and some tailored ones require considerable computational effort, e.g. an hour of computation. Our experiments clearly show that the proposed approach can deliver results that yield qualitatively same information, more understanding and can be computed in minutes by hand (or within a fraction of a second by computer).

Demonstration: Analysis of Stochastic Gene Expression Model [8]**.** The CRN underlying the stochastic gene expression model is described in Table 1. As discussed in [5,10], the system oscillates between two phases characterised by the D_{on} state and the D_{off} state, respectively. Biologists are interested in how the distribution of the D_{on} and D_{off} states is aligned with the distribution of RNA and protein P.

In order to demonstrate the refinement step and its effect on the accuracy of the model, we start with a very coarse abstraction. It distinguishes only the zero population and the non-zero populations. The pruned abstract model obtained using our approach is depicted in Fig. 1 (left). The full one before pruning is shown in Fig. 6 [3, Appendix].

The proposed analysis of the model identifies the key trends in the system dynamic. The red transitions, representing iterations 1–3 of the semi-quantitative analysis, capture the most probable paths in the system. The green component includes states with DNA on where the system oscillates. The component is reached via the blue state with D_{off} and no RNAs/P. The blue state is promptly reached from the initial state and then the system waits (roughly 100 h according

our rate abstraction) for the next DNA activation. The component is left via a deactivation in the iteration 4 (the blue dotted transition). The estimation of the exit time is 100 h. The deactivation is then followed by fast red transitions leading to the blue state, where the system waits for the next activation. We obtain an oscillation between the blue state and the green component, representing the expected oscillation between the D_{on} and D_{off} states.

As expected, this abstraction does not clearly predict the bimodal distribution on the RNA/P populations – the green component includes states with both the zero and the non-zero population of the mRNA and the protein. In order to obtain a more accurate analysis of the system, we refine the population discretisation using a single level threshold for P and DNA, that is equal to 100 and 10, respectively (the rates in the CRN indicate that the population of P reaches higher values).

Figure 1 (right) depicts the pruned abstract model with the new discretisation (the full model is depicted in Fig. 7 [3, Appendix]. We again obtain the oscillation between the green component representing DNA_{on} states and the blue DNA_{off} state. The states in the green component more accurately predicts that in the DNA_{on} states the populations of RNA and P are high and drop to zero only for short time periods. The figure also shows orange transitions within the iteration 2 that extend the green component by two states. Note that the system promptly returns from these states back to the green component. After the deactivation in the iteration 4, the system takes (within the same iteration) the fast transitions (solid blue) leading to the blue component where system waits for another activation and where the mRNA/protein populations decrease. The expected time spent in states on blue solid transitions is small and thus we can reliably predict the bimodal distribution of the mRNA/P populations and its correlation with the DNA state. The refined abstraction also reveals that the switching time from the DNA_{on} mode to the DNA_{off} mode is lower. These predications are in accordance with the results obtained in [10]. See in Fig. 8 [3, Appendix] that is adopted from [10] and illustrates these results.

To conclude this case study, we observe a very aligned agreement between the results obtained using our approach with virtually no computational cost and results in [10] obtained via advanced and time consuming numerical methods.

References

1. Cardelli, L.: Two-domain DNA strand displacement. Math. Struct. Comput. Sci. **23**(02), 247–271 (2013)
2. Cardelli, L., Kwiatkowska, M., Laurenti, L.: A Stochastic hybrid approximation for chemical kinetics based on the linear noise approximation. In: Bartocci, E., Lio, P., Paoletti, N. (eds.) CMSB 2016. LNCS, vol. 9859, pp. 147–167. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45177-0_10
3. Češka, M., Křetínský, J.: Semi-quantitative abstraction and analysis of chemical reaction networks. Tech. Rep. abs/1905.09914 (2019)
4. Chellaboina, V., Bhat, S.P., Haddad, W.M., Bernstein, D.S.: Modeling and analysis of mass-action kinetics. IEEE Control Syst. Mag. **29**(4), 60–78 (2009)

5. Gandhi, S.J., Zenklusen, D., Lionnet, T., Singer, R.H.: Transcription of functionally related constitutive genes is not coordinated. Nat. Struct. Mol. Biol. **18**(1), 27 (2011)
6. Giacobbe, M., Guet, C.C., Gupta, A., Henzinger, T.A., Paixão, T., Petrov, T.: Model checking gene regulatory networks. In: Baier, C., Tinelli, C. (eds.) TACAS 2015. LNCS, vol. 9035, pp. 469–483. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46681-0_47
7. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**(25), 2340–2361 (1977)
8. Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C.: Real-time kinetics of gene activity in individual bacteria. Cell **123**(6), 1025–1036 (2005)
9. Goutsias, J.: Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. J. Chem. Phys. **122**(18), 184102 (2005)
10. Hasenauer, J., Wolf, V., Kazeroonian, A., Theis, F.: Method of conditional moments (MCM) for the chemical master equation. J. Math. Biol. **69**(3), 1–49 (2013)
11. Heath, J., Kwiatkowska, M., Norman, G., Parker, D., Tymchyshyn, O.: Probabilistic model checking of complex biological pathways. Theor. Comput. Sci. **391**(3), 239–257 (2008)
12. Lakin, M.R., Parker, D., Cardelli, L., Kwiatkowska, M., Phillips, A.: Design and analysis of DNA strand displacement devices using probabilistic model checking. J. R. Soc. Interface **9**(72), 1470–1485 (2012)
13. Salis, H., Kaznessis, Y.: Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J. Chem. Phys. **122**(5), 054103 (2005)
14. Soloveichik, D., Seelig, G., Winfree, E.: DNA as a universal substrate for chemical kinetics. Proc. Nat. Acad. Sci. U.S.A. **107**(12), 5393–5398 (2010)
15. Van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, vol. 1. Elsevier, Amsterdam (1992)