

Empirical Performance Investigation of a Büchi Complementation Construction

Daniel Weibel

June 19, 2015

Abstract

This will be the abstract.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Dr. Ulrich Ultes-Nitsche and Joel Allred to bring me to the topic of this thesis in the first place, and for the many enlightening discussions and explanations. A very special thank goes to Ming-Hsien Tsai, the main author of the GOAL tool, from the National Taiwan University. Without his very prompt and patient responses to my many questions regarding implementation details of GOAL, the plugin would not have been possible in its current form. I also would like to thank Michael Rolli and Nico Färber from the Grid Admin Team of the University of Bern for their very friendly and helpful responses to my questions about the Linux cluster where the experiments of this thesis have been executed.

Contents

| | | |
|----------|---|-----------|
| 1 | Performance Investigation of the Fribourg Construction | 2 |
| 1.1 | Implementation | 3 |
| 1.1.1 | GOAL | 3 |
| 1.1.2 | Implementation of the Construction | 5 |
| 1.1.3 | Verification of the Implementation | 7 |
| 1.2 | Test Data | 8 |
| 1.2.1 | GOAL Test Set | 8 |
| 1.2.2 | Michel Test Set | 11 |
| 1.3 | Experimental Setup | 12 |
| 1.3.1 | Internal Tests | 12 |
| 1.3.2 | External Tests | 14 |
| 1.3.3 | Time and Memory Limits | 14 |
| 1.3.4 | Execution Environment | 15 |
| 2 | Results and Discussion | 17 |
| 2.1 | Internal Tests | 18 |
| 2.1.1 | GOAL Test Set | 18 |
| 2.1.2 | Michel Test Set | 24 |
| 2.2 | External Tests | 26 |
| 2.2.1 | GOAL Test Set | 26 |
| 2.2.2 | Michel Automata | 29 |
| 2.3 | Discussion | 31 |
| 2.3.1 | Summary of the Results | 31 |
| 2.3.2 | Validity | 31 |
| A | Plugin Installation and Usage | 32 |
| B | Median Complement Sizes of the GOAL Test Set | 33 |
| C | Execution Times | 35 |

Chapter 1

Performance Investigation of the Fribourg Construction

Contents

| | | |
|------------|------------------------------------|-----------|
| 1.1 | Implementation | 3 |
| 1.1.1 | GOAL | 3 |
| 1.1.2 | Implementation of the Construction | 5 |
| 1.1.3 | Verification of the Implementation | 7 |
| 1.2 | Test Data | 8 |
| 1.2.1 | GOAL Test Set | 8 |
| 1.2.2 | Michel Test Set | 11 |
| 1.3 | Experimental Setup | 12 |
| 1.3.1 | Internal Tests | 12 |
| 1.3.2 | External Tests | 14 |
| 1.3.3 | Time and Memory Limits | 14 |
| 1.3.4 | Execution Environment | 15 |

In this chapter we come to the core of this thesis, namely the empirical performance investigation of the Fribourg construction. We are interested in two things. First, how the different versions of the Fribourg construction compare to each other. That is, which combination of optimisations makes the Fribourg construction most efficient? Second, we want to know how the Fribourg construction performs compared to other complementation constructions. Throughout this thesis, we will refer to the first question as the *internal* investigations, and to the second question as the *external* investigations.

To do an empirical performance investigation we need an implementation of the Fribourg construction. We decided to create this implementation as a part of an existing ω -automata tool called GOAL¹. This tool already contains implementations of various Büchi complementation constructions, so that once our own implementation of the Fribourg construction is in place, we can easily compare it to other constructions (that is, performing the external tests).

Next, we need test data. That is, a set of automata on which to test the different complementation constructions. We defined two, very different, test sets. One contains a large number of randomly generated automata and is referred to as the GOAL test set, and the other contains a very small number of very special automata and is referred to as the Michel test set.

Finally, the experiments need to be executed. Since the complementation of all the automata in our test set with many different constructions is a rather heavy computation task, we decided to make use of a professional high-performance computing (HPC) environment. All the experiments have been carried out on the computer cluster UBELIX at the University of Bern².

In this chapter, we are going to describe each of these points in separate sections. Section 1.3 also includes our concrete experimental setup, where we present which constructions we are executing for which test data, and what are the allocated resources and imposed constraints. The results of the experiments will finally be presented in Chapter 2.

1.1 Implementation

1.1.1 GOAL

GOAL stands for *Graphical Tool for Omega-Automata and Logics* and is being developed since 2007 at the National Taiwan University (NTU), Department of Information Management³. It has been presented in various scientific publications [15][16][17][14]. The tool is freely available on the website <http://goal.im.ntu.edu.tw>. It is a Java program, and thus works on all platforms that have a Java runtime environment.⁴

GOAL allows to graphically and interactively create and manipulate different types of ω -automata, including non-deterministic Büchi automata. The palette of provided manipulations is vast and ranges from input testing, over conversions to other types of non-deterministic ω -automata, to implementations of graph layout algorithms. Figure 1.1 shows a screenshot of GOAL's graphical user interface with an open menu showing the breadth of possible manipulations for ω -automata.

Relevant for our purposes are the complementation constructions for non-deterministic Büchi automata that GOAL provides. Also here, GOAL comes up with a comprehensive offer. In the 2014-11-17 version of GOAL, there are 10 pre-implemented complementation constructions for NBW. Table 1.1 summarises these constructions and indicates the authors and the publications on which the implementations are based.

The construction in Table 1.1 are sorted by the four fundamental complementation approaches. The first one, Ramsey, is the only construction belonging to the Ramsey-based complementation approach. The following four constructions, Safra, ModifiedSafra, Piterman, and MS, belong to the determinization-based approach. Rank, WAPA, and WAA are representants of the Rank-based approach. Finally, Slice and Slice+P belong to the slice-based approach.

¹<http://goal.im.ntu.edu.tw/>

²<http://ubelix.unibe.ch>

³<http://exp.management.ntu.edu.tw/en/IM>

⁴The required version of the runtime environment is 6 or higher (Java 1.6).

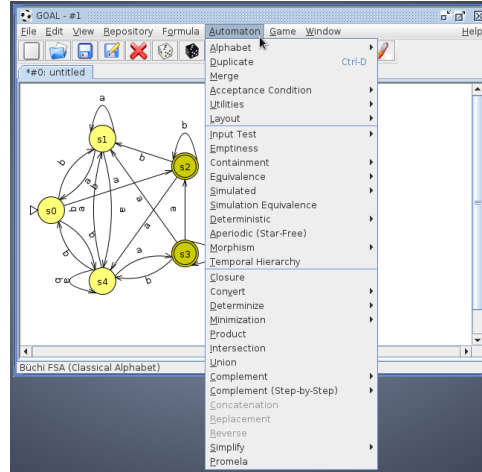


Figure 1.1: The graphical user interface of GOAL (here version 2014-11-17, which is, at the time of this writing the latest one) with a non-deterministic Buchi automaton in the editor window and an open menu showing the different types of manipulations that are provided for ω -automata.

| # | Identifier | Name/description | Authors (year) | Ref. |
|----|---------------|--------------------------------------|------------------------------|------|
| 1 | Ramsey | Ramsey-based construction | Sistla, Vardi, Wolper (1987) | [10] |
| 2 | Safra | Safra construction | Safra (1988) | [8] |
| 3 | ModifiedSafra | Modification by Althoff | Althoff (2006) | [1] |
| 4 | Piterman | Safra-Piterman construction | Piterman (2007) | [7] |
| 5 | MS | Muller-Schupp construction | Muller, Schupp (1995) | [6] |
| 6 | Rank | Rank-based construction | Schewe (2009) | [9] |
| 7 | WAPA | Via weak alternating parity automata | Thomas (1999) | [12] |
| 8 | WAA | Via weak alternating automata | Kupferman, Vardi (2001) | [4] |
| 9 | Slice+P | Slice-based construction (earlier) | Vardi, Wilke (2007) | [18] |
| 10 | Slice | Slice-based construction (later) | Kähler, Wilke (2008) | [3] |

Table 1.1: The pre-implemented NBW complementation constructions in GOAL (version 2014-11-17).

Slice and Slice+P are combined in a single construction in GOAL but the two specific constructions can be selected by the means of an option. If the P option of Slice is activated, then the Slice+P construction is used, and otherwise the Slice construction is used. In our experiments, we will use the Slice+P version, however, we will still refer to this construction as simply Slice in its short form.

Almost the entire functionality of GOAL is also available via a command line interface. This makes it suitable for automatic batch operation, as it is needed for our experiments. For storing automata, GOAL defines its own file format called GFF (GOAL file format) which is based on XML and is typically used with the filename extension `.gff`.

At this point, it is worth pointing at a related project of the same research group, called the Büchi Store. This is an online repository of classified and tagged ω -automata that can be downloaded in different formats (including GFF). The Büchi Store is located on <http://buchi.im.ntu.edu.tw/> and has also been described in a scientific publication [?]. Furthermore, there is a binding in GOAL to the Büchi Store, so that the contents of the store can be directly accessed from GOAL. For our project we did not make use the Büchi Store, but it is might be an interesting option for related projects.

1.1.2 Implementation of the Construction

The Extensible Plugin Architecture of GOAL

GOAL has been designed from the ground up to be modular and extensible. To this end, it has been created with the Java Plugin Framework (JPF)⁵. This framework allows to build applications whose functionality can be easily and seamlessly extended by writing additional plugins for it. These plugins can be installed in the main application without the need to recompile the whole application. Rather, the plugin is compiled separately and the resulting bytecode files are copied to the directory tree of the main application. It is not even necessary to know the source code of the main application in order to write a plugin. The interfaces of JPF itself, and the documentations of the relevant classes of the main application are all that a plugin writer needs to know.

In some more detail, JPF requires an application to define so called extension points. For a given extension point, multiple extensions can be provided. These extensions contain the actual functionality of the application. A JPF application basically consists of a groups of different types of extensions that are plugged in their corresponding extension points. A plugin, finally, is an arbitrary bundle of extensions and extension points. It is the basic unit of organisation in the Java Plugin Framework.

One of the extension points of GOAL is called **Complementation Construction**. In order to add a new complementation construction to GOAL, one has to write an extension for **Complementation Construction** that contains the implementation of the new construction. This extension can then be bundled as a plugin, and the plugin can be compiled and installed in the main application. This makes the new extension an integral part of the main application. That means that after starting GOAL with the installed plugin, the new complementation construction is included in GOAL in exactly the same way as all the other existing complementation constructions are included.

This is exactly what we did in order to add the Fribourg construction to GOAL. The name of our plugin is `ch.unifr.goal.complement`⁶ and it is publicly available. Appendix A contains instructions how the plugin can be obtained and installed.

There is more than just the extension point **Complementation Construction** that can be extended to add a new complementation construction to GOAL. There are separate extension points for including the construction in the command line interface of GOAL, in the menu structure, or for providing a step-by-step version of the construction. We aimed at making this integration as complete as possible so that the Fribourg construction provides the same facilities as the pre-implemented constructions. This includes, for example, comand line integration (of course required for our experiments), menu integration for setting default options or defining the Fribourg construction as the default complementation construction of GOAL, or an interactive step-by-step execution environment.

The Fribourg Construction Plugin

In our implementation of the Fribourg construction we also included the three optimisations, R2C, M1, and M2, described in Section ???. To leave the choice of which optimisations to use for any given complementation task to the user, we implemented these optimisation as *options*. Complementation construction options in the GOAL GUI are presented to the user as a list of checkboxes immediately before the start of the complementation. The user can then make his or her selection and launch the complementation process. In the command line interface, options are represented as flags that can be added to the corresponding command.

In addition to the three optimisations, we also added further options to our implementation. Table 1.2 shows a list of all the available options for the Fribourg construction in the current version of the plugin.

Table 1.2 defines an identifier for each option, and we will use these identifiers throughout the rest of this thesis for referring to the corresponding options. These identifiers also correspond to the names of the option flags in the command line interface (in lower case).

⁵<http://jpf.sourceforge.net/>

⁶By convention, JPF plugins are named after the base package name of their implementation files.

| Identifier | Description |
|------------|---|
| R2C | Deleting states with rightmost colour 2, if automaton is complete |
| M1 | Component merging optimisation |
| M2 | Single 2-coloured component optimisation |
| C | Make input automaton complete |
| R | Remove unreachable and dead states from output automaton |
| RR | Remove unreachable and dead states from input automaton |
| MACC | Maximise accepting states of input automaton |
| B | Use the “bracket notation” for state labels |

Table 1.2: The options of the Fribourg construction in GOAL.

The first three options in Table 1.2 represent the three optimisations from Section ???. The R2C optimisation is implemented so that it applies only to input automata that are complete. That is, selecting R2C for the complementation of an automaton that is not complete has no effect, and results in the same as not selecting R2C in the first place.

The options M1 and M2 implement the M1 and M2 optimisations. Since M2 is dependent on M1, we enforced that M2 can only be selected if M1 is also selected.

The C option is one of the option that preprocesses the input automaton before the actual complementation starts. This option checks if the input automaton is complete, that is, if every state has at least one outgoing transition for every symbol of the alphabet. If this is not the case, it makes the automaton complete by adding a sink state. A sink state is a non-accepting state that “aggregates” all the missing transitions of incomplete states, so that these states become complete. The sink state itself has transitions to itself for all symbols of the alphabet. The purpose of this option is to be used together with R2C. By making an automaton complete with C, we can ensure that the R2C optimisation will be applied. However, in this case, the size of the automaton to complement is increased by one what is in turn negative for the performance of the construction. This effects has been investigated by Göttel in previous work on the Fribourg construction [2], and we will re-investigate hem in this thesis as explained in Section 1.3.

The R option affects the complement automaton after the termination of the construction. Its effect is to remove all the unreachable and dead states from the complement. Unreachable states are states that cannot be reached from the initial state. Dead states are states from which it is not possible to reach an accepting state. These states can be removed from any automaton without changing the language of the automaton. The R option is common among the pre-implemented complementation constructions of GOAL. Ramsey, Piterman, Rank, and Slice, all have the R option. It allows for example to determine the number of unreachable and dead state that a complementation construction produces. To this end one can complement the same automaton with and without R and then calculate the difference of the sizes of the two output automata. Such investigations have been done in [13] for some of the pre-implemented complementation constructions in GOAL. In our experiments, we will also use the R option, however our investigations on the unreachable and dead states will not be as in-depth as in [13].

The RR option is similar to the R option, except that it removes the unreachable and dead states from the input automaton, before the start of the actual complementation construction, rather than from the output automaton. The pre-implemented constructions in GOAL do not have something like the RR option, and we added it to the Fribourg construction just for our own interest.

MACC is also a pre-processing option. Its effect is to maximise the accepting set of the input Büchi automaton. This means that as many states as possible are made accepting without changing the language of the automaton. This technique is introduced and described in [13]. In this paper Tsai et al. state that a higher ratio of accepting states generally results in smaller complements. A result that we can confirm with our own experimental results that we present in Chapter 2. The MACC option is available for the Ramsey, Piterman, Rank, and Slice constructions in GOAL. In our experiments we will however not make use of the MACC option.

Finally, the B option has no effect on the construction itself. It just activates the “bracket notation” for

marking the colours of the sets inside the produced states. This notation corresponds to the notation we used throughout Chapter ???. With the standard notation, sets are represented as 2-tuples, where the first element is the actual set of states and the second element is the colour of the set.

1.1.3 Verification of the Implementation

Having an implementation of a complementation construction, one also needs to be sure that this implementation is correct. That is, the language of an output automaton must really be the complement of the language of the input automaton. It is crucial that this is the case, especially when doing empirical performance investigations. A construction that produces wrong results would lead to a misinterpretation of the actual performance of the construction.

In order to verify the correctness of our implementation, we chose to compare the results of our construction to the results of the pre-implemented complementation constructions of GOAL. The existing complementation constructions in GOAL are thus taken as the “ground truth”. Concretely, the approach is as follows. We take an automaton A and complement it with, say, the Piterman construction in GOAL. This yields automaton A' . Then we complement the same automaton A with the Fribourg construction, what yields automaton A'' . By the means of GOAL’s equivalence operation, we then test whether A' and A'' accept the same language. We call this procedure *complement-equivalence test*.

Of course, with such an empirical approach it is not possible to conclusively verify that our implementation of the Fribourg construction is correct. This would require to do a complement-equivalence test for each of an infinite number of automata, which is not possible. With this approach we can only conclusively falsify the correctness of the implementation, as a single counterexample is enough to prove that the implementation produces wrong results. However, we think that the absence of counterexamples for a certain number of cases is still a good indicator that there are no errors in the implementation and thus confirms our hypothesis that the implementation is correct.

Concretely, we did 1,000 complement-equivalence tests for different versions of the Fribourg construction (resulting from different sets of activated options) with random automata. The automata we used for the test had four states, an alphabet size between 2 and 4, and were generated with the random automaton generation operation of GOAL. The concrete versions of the Fribourg construction that we tested are:

- Fribourg
- Fribourg+R2C+C
- Fribourg+M1
- Fribourg+M1+M2
- Fribourg+R
- Fribourg+RR
- Fribourg+MACC

For each of these versions, the 1,000 random test automata have been generated on-the-fly. The used “ground truth” construction was Piterman. All of the tests were successful, that is, there were no counterexamples.

The reason that we chose rather small test automata is to avoid practical problems regarding time and memory requirements during the computation. The equivalence operation in GOAL is implemented as a reciprocal containment test, which includes complementation. For a complement-equivalence test, we thus have to complement the complements of the initial automata. If the initial automata are too big, then their complements might be already so big that it becomes practically infeasible to complement them again.

These tests are certainly not extensive, and it would also be interesting to use larger and more complex automata as test cases (maybe with a different testing approach). However, given the results of the tests we made, we are still confident that our implementation is correct with a high probability.

With the above listing of Fribourg construction versions (for example, Fribourg+R2C+C) we also introduce a notation that we will use throughout the rest of this thesis. The first token is the base construction

and is one of the identifiers in Table ?? in Section 1.1.1. The option identifiers appended with plus signs indicate which options (if any) are “activated” in the given version of the construction.

1.2 Test Data

The core of this thesis is to empirically investigate the performance of the Fribourg construction. To this end, we “feed” concrete automata to the Fribourg construction and analyse the results. This set of automata is our test data and is of crucial importance for the validity of the performance investigation. Test data should not be biased in favour or unfavour of one of the tested algorithms. Ideally, publicly available, or even standardised, test data is used, that has been previously used in other experiments. Furthermore, test data should cover interesting cases that reveal important aspects about the algorithm.

For our investigation we chose two test sets that are publicly known or available and cover very different scenarios. While one of them consists of 11,000 automata and tries to cover a broad range of “everyday” complementation problems, the other one consists of just 4 automata, that are however very special. We call the former one the GOAL test set and the latter one the Michel test set, and we are going to describe them in the following two sections.

1.2.1 GOAL Test Set

Introduction

The GOAL test set has been created by Tsai et al. for the experiments described in the paper *State of Büchi Complementation* [?]. In this paper, the authors carry out empirical performance investigations of Büchi complementation constructions that are very similar to our own investigations. The experiments were run with the GOAL tool, hence the name GOAL test set. The idea leading the design of this test set was to generate a large set of complementation problems with various difficulty levels.

The test set consists of 11,000 automata of size 15, and with an alphabet size of 2. Within the given constraints, the 11,000 automata were generated at random. The test set is furthermore structured into 110 classes that are combinations of two parameters, the so called transition density and the acceptance density. Simply speaking, the transition density determines the number of transitions in an automaton, and the acceptance density determines the number of accepting states.

More precisely, the transition density t is defined as follows. Let n be the number of states of automaton A , and t its transition density. Then A contains tn transitions for each symbol of the alphabet⁷. That is, if one of our test set automata with 15 states and the alphabet $0, 1$ has a transition density of 2, then it contains exactly 30 transitions for symbol 0 and 30 transitions for symbol 1. Consequently, each state has on average two outgoing and two incoming transitions for each symbol. Therefore, the transition density can also be seen as the average number of outgoing and incoming transitions per alphabet symbol that a state has.

The acceptance density a is defined as the ratio of accepting states to non-accepting states in the automaton. That is, if automaton A has n states and an acceptance density of a , then it contains an accepting states⁸. The acceptance density is thus bound to a number between 0 and 1. For example, if an automaton of our test set has an acceptance density of 0.1, then it has 2 accepting states, and if an automaton has an accepting density of 1, then all of its states are accepting states.

In the GOAL test set there are 11 transition densities and 10 acceptance densities. The transition densities range from 1 to 3 in steps of 0.2, and the acceptance densities range from 0.1 to 1 in steps of 0.1. Thus, we have

$$\begin{aligned} t' &= \{1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0\} \\ a' &= \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\} \end{aligned}$$

⁷In the case tn is not an integer, it is rounded up to the next integer.

⁸Again, if an is not an integer, it is rounded up to the next integer.

The 110 classes result from the Cartesian product of t' and a' . Each class consequently contains 100 automata. Throughout this thesis, we will often present these classes as a 11×10 matrix where the rows represent the transition densities t' , and the columns represent the acceptance densities a' .

There is a second GOAL test set that is similar to the one that we just described, except that the automata sizes are 20 instead of 15. Tsai et al. [?] did their experiments with both, the size-15 (as the primary), and the size-20 (as the secondary) test set. For our own experiments we are not using the size-20 test set. First, we think that repeating the experiments with the size-20 test set would add only a limited amount of insight. It might be more beneficial to do further tests with a more diverse type of test data, like for example the Michel test set that we describe in the next section. Furthermore, the bigger automata sizes in the size-20 test set are likely to cause practical problems in terms of computing and time resources (see Section 1.3.4). We decided therefore to restrict ourselves to the size-15 test set, and when we refer to *the* GOAL test set throughout the rest of this thesis, we specifically mean the size-15 test set.

The GOAL test sets (both, size-15 and size-20) are publicly available on the GOAL website via the following link: http://goal.im.ntu.edu.tw/wiki/lib/exe/fetch.php?media=goal:ciaa2010_automata.tar.gz.

Analysis

We analysed some properties of the GOAL test set that may be of importance for the interpretation of the results of our experiments. In particular, we tested how many automata of the GOAL test set are *complete*, *universal*, and *empty*.

An automaton is complete if every state has at least one outgoing transition for every symbol of the alphabet. Universality means that an automaton accepts *every* word that can be generated from its alphabet. Emptiness is the contrary of universality and means that an automaton does not accept *any* word (except the empty word ϵ).

To know how many automata are complete is relevant because the R2C optimisation of the Fribourg construction only applies to complete automata. Universality and emptiness are interesting because their respective complement sizes have a lower bound of 1. The complement of a universal automaton is an empty automaton, which can be represented by a single state, and the complement of an empty automaton is a universal automaton which can be represented by a single state as well. In this way we can get an idea of how many unreachable and dead states the Fribourg construction produces. If the produced complement of a universal automata has, say, 2,000 states then 1,999 states of this automata are actually unnecessary as the same automaton could be represented by a single state. We will investigate this aspect to some extent with the R option for the Fribourg construction (see Section 1.3).

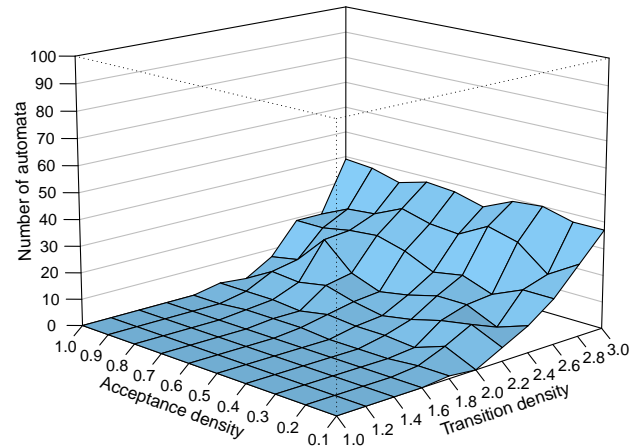
GOAL provides a command for testing emptiness of an automaton. However, it does not provide commands for testing completeness and universality. We therefore implemented these commands on our own as a separate GOAL plugin (`ch.unifr.goal.util`, also available as described in Appendix A). The tests were executed in the execution environment as described in Section 1.3.4. The overall results are as follows.

- 990 of the 11,000 automata are complete (9%)
- 6,796 of the 11,000 automata are universal (61.8%)
- 63 of the 11,000 automata are empty (0.6%)

In addition to these overall results, we wanted to know how these properties are distributed over the 110 transition/acceptance density classes. Therefore, we counted the number of complete, universal, and empty automata for each of these 110 classes. This naturally results in three-dimensional data where two dimensions are constituted by the transition density and acceptance density (to form the 110 classes), and the third dimension is the number of automata having a specific property. There are several ways to represent such data, and two of them are matrices and so called perspective plots. In Figure 1.2 we present our results in both of these ways.

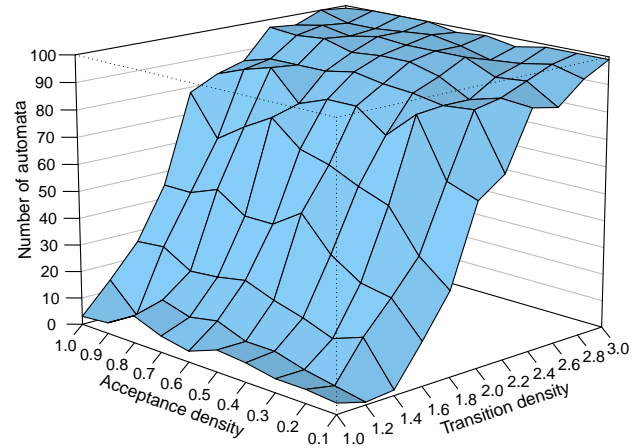
The matrices have 11 rows and 10 columns. The rows represent the transition densities and the columns represent the acceptance densities. The cells represent the measured values for the class constituted by the transition density of its row and the acceptance density of its column. We will use this matrix representation again throughout the rest of this thesis. The perspective plots represent the transition

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1.4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1.8 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| 2.0 | 0 | 5 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 2 |
| 2.2 | 5 | 10 | 8 | 5 | 3 | 5 | 8 | 6 | 7 | 1 |
| 2.4 | 10 | 6 | 11 | 11 | 8 | 6 | 10 | 20 | 9 | 7 |
| 2.6 | 17 | 17 | 12 | 16 | 14 | 19 | 22 | 21 | 19 | 19 |
| 2.8 | 27 | 20 | 29 | 32 | 26 | 27 | 30 | 25 | 24 | 19 |
| 3.0 | 37 | 37 | 40 | 39 | 34 | 37 | 38 | 35 | 38 | 39 |



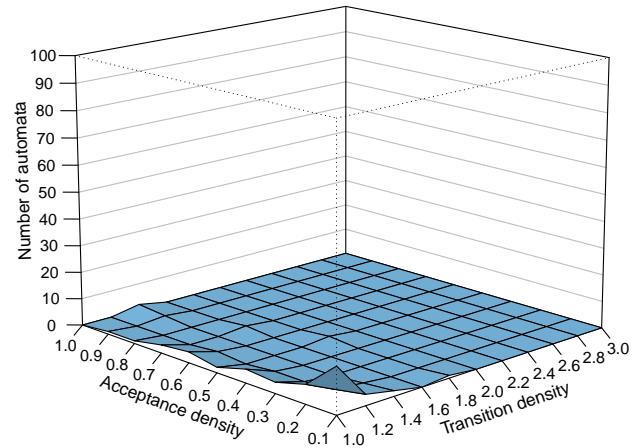
(a) Completeness

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.0 | 4 | 5 | 5 | 7 | 8 | 4 | 6 | 10 | 4 | 3 |
| 1.2 | 1 | 3 | 5 | 8 | 8 | 12 | 10 | 13 | 4 | 14 |
| 1.4 | 2 | 17 | 13 | 17 | 20 | 24 | 22 | 21 | 27 | 26 |
| 1.6 | 16 | 28 | 30 | 37 | 49 | 42 | 42 | 49 | 45 | 45 |
| 1.8 | 31 | 40 | 55 | 59 | 64 | 67 | 76 | 70 | 63 | 78 |
| 2.0 | 60 | 64 | 85 | 75 | 83 | 83 | 79 | 90 | 87 | 83 |
| 2.2 | 67 | 87 | 86 | 88 | 89 | 91 | 89 | 89 | 89 | 86 |
| 2.4 | 88 | 89 | 86 | 92 | 95 | 95 | 94 | 97 | 96 | 97 |
| 2.6 | 86 | 93 | 92 | 97 | 97 | 97 | 98 | 96 | 98 | 96 |
| 2.8 | 94 | 97 | 95 | 94 | 97 | 99 | 98 | 97 | 97 | 100 |
| 3.0 | 99 | 99 | 99 | 97 | 99 | 98 | 100 | 100 | 100 | 99 |



(b) Universality

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.0 | 17 | 7 | 4 | 5 | 2 | 4 | 3 | 1 | 1 | 0 |
| 1.2 | 4 | 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1.4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| 1.6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1.8 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |



(c) Emptiness

Figure 1.2: Completeness and universality in the GOAL test set.

densities, acceptance densities, and measured values along three axes, x , y , and z . The individual data points are connected by lines so that a grid emerges that can be seen to form a three-dimensional surface. Perspective plots can be seen as three-dimensional visualisations of matrices where the cell values extend as “heights” along the z -axis in the three-dimensional space. However, perspective plots are typically rotated, and this rotation is chosen for making the presented data optimally visible. In Figure 1.2, looking

at the perspective plots is like looking at the corresponding matrices from the upper left corner. We will use perspective plots throughout the rest of this thesis. While their relation to matrices is the same, their rotation will be different. For example, in Chapter 2, the viewpoint of the perspective plots will correspond to looking from the bottom right corner at the corresponding matrices.

Regarding the data itself, we can identify the following points. With 9%, a relatively small number of the automata is complete. This affects the R2C optimisation, which therefore makes a difference to only 9% of the test data, while leaving 91% unaffected. We can see that there are no complete automata in classes with transition densities up to 1.6, and then starting from 1.8 the number starts to increase up to a ratio of 34–40% at a transition density of 3.0. Naturally, a higher number of transitions in the automata increases the probability that each state has at least one outgoing transition for every symbol of the alphabet. Considering that a transition density of, for example, 3.0 means that each state has on average three outgoing transitions for every symbol of the alphabet, this ratio might seem rather small. However, a single incomplete state is enough to make the whole automaton incomplete, even though all the other states would be complete. With a total number of 15 states, the probability of an incomplete state is apparently still quite high, even for a transition density of 3.0, as our empirical data suggests. Naturally, the acceptance density does not influence the ratio of complete automata.

The ratios look very different for universality. With 61.8%, a very high number of automata in the test set is universal. This high number might result from the small alphabet of just two symbols of the test automata. With a smaller alphabet, there are fewer possible words, and thus a higher probability that an automaton accepts all the possible words. It would be interesting to investigate the universality ratio of similar automata with larger alphabet sizes. As for completeness, the ratio of universal automata also increases with the transition density, although much faster up to a ratio of 100%. In addition, we can see a slightly higher number of universal automata in classes with higher acceptance densities (although the pattern is clearly dominated by the transition density). We can summarise this, by saying that more transitions and more accepting states increase the probability that a given word is accepted, and thus the probability that an automaton accepts all the possible words.

The other extreme is the percentage of empty automata, which is just 0.6%, or 63 instances. Apparently the probability that an automaton of size 15 and with 15 to 45 transitions per alphabet symbol accepts no words at all is very small. The probability is highest in automata with low transition densities and acceptance densities.

1.2.2 Michel Test Set

Our second test set is very different from the first one. It consists of few well-defined, rather than a large number of random automata. These automata are the Michel automata with $m = \{1, \dots, 4\}$, that we call Michel 1, Michel 2, Michel 3, and Michel 4, respectively.

Michel automata have been introduced in 1988 by Max Michel in order to prove a lower bound for the state growth of Büchi complementation of $(n - 2)!$, where n is the number of states of the input automaton [5][11]. Michel constructed a family of automata, characterised by the parameter m , that have $m + 1$ alphabet symbols, and $m + 2$ states. He proved that the complements of these automata cannot have less than $m!$ states. Since the number of states of the input automata is $n = m + 2$, the state growth in terms of input and output states is $(n - 2)!$, which is around $(0.36n)^n$.

Michel automata are thus very hard automata, and that is why we chose them as our second test set. By testing our Fribourg construction on very hard automata, we can empirically determine “lower bounds for the lower bounds” for the state complexity of our construction. For example, if in our experiments we observe a state growth of a Michel automaton of, say, $(0.99n)^n$, then the real (theoretical) lower bound of the construction is either $(0.99n)^n$ or higher. But it cannot be less than $(0.99n)^n$, because with our experiment we have a proof that there exists an automaton on which our construction produces a state growth of $(0.99n)^n$.

Michel automata are very hard to complement, but of course we cannot prove that they are the *hardest* ones (of their respective size) for our construction. If for example, the Michel automaton with six states (Michel 4), produces a state growth of $(0.99n)^n$, then we don’t know if this is the hardest automaton of size 6 for our construction (in which case the real state complexity of our construction would be $(0.99n)^n$),

or if there is another automaton of size 6 that produces an even larger complement (in which case the real state complexity of our construction would be higher than $(0.99n)^n$). However, since Michel automata are so hard, we think that they give a useful approximation of a state complexity lower bound of our construction.

In Figure 1.3 we present the four Michel automata with $m = \{1, \dots, 4\}$, as they are defined in [5] and [11]. They are in particular:

- Michel 1: 3 states, 2 symbols, 7 transitions
- Michel 2: 4 states, 3 symbols, 14 transitions
- Michel 3: 5 states, 4 symbols, 23 transitions
- Michel 4: 6 states, 5 symbols, 34 transitions

All Michel automata have a single accepting state. The reason that we included only the first four Michel automata are of practical nature. Michel automata are so hard, that starting from Michel 5, the required computing and time resources are so high, that it would become practically infeasible for us to carry out the complementation process. We present some extrapolations on this topic when we discuss the results of the Michel experiments in Chapter 2.

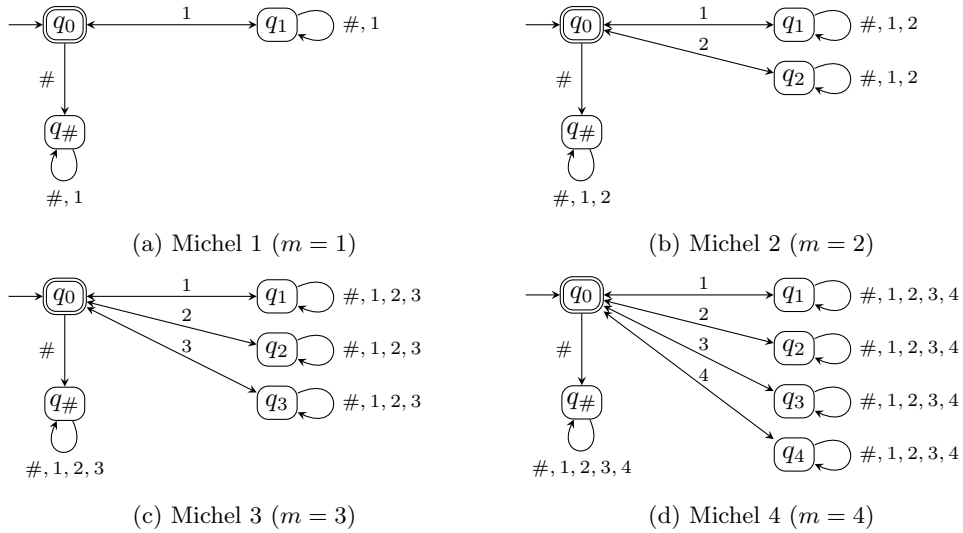


Figure 1.3: The Michel automata with $m = \{1, \dots, 4\}$, an alphabet size of $m + 1$, and $m + 2$ states.

1.3 Experimental Setup

1.3.1 Internal Tests

In the internal tests our aim is to compare different versions of the Fribourg construction. A specific version of the Fribourg construction is composed of a combination of options. Options can be the

As presented in Section ??, there are three optimisations to the Fribourg construction:

- R2C: if the input automaton is complete, remove all states whose rightmost colour is 2
- M1: merge certain adjacent sets within a state
- M2: reduce 2-coloured sets (requires M1)

Furthermore, our GOAL plugin includes, among others, the following options:

- C: make the input automaton complete (by adding a sink state)
- R: remove unreachable and dead states from the output automaton

The versions of the Fribourg construction that we chose for our internal tests consist of combinations of these five options. According to the nature of the two test sets (the GOAL test set and the Michel automata), we chose two different sets of versions for the two test sets. We are now first going to describe the setup for the GOAL test set and then the one for the Michel test set.

GOAL Test Set

For the tests on the GOAL test set, we chose the following eight versions of the Fribourg construction:

1. Fribourg
2. Fribourg+R2C
3. Fribourg+R2C+C
4. Fribourg+M1
5. Fribourg+M1+M2
6. Fribourg+M1+R2C
7. Fribourg+M1+R2C+C
8. Fribourg+R

The first version is the plain Fribourg construction without any optimisations or options. The next two versions are devoted to investigate the R2C optimisation. Version 2 applies the optimisation only to the automata which happen to be complete (and as we have seen in Section ?? these are 9%). Version 3, on the other hand, makes all input automata preliminarily complete by adding a sink state, so that the R2C optimisation can be applied to *all* automata. The question here is, does it pay off to increase the size of the automata by one (for adding the sink state) but then being able to apply R2C, or is it better to not add the extra state but then not applying R2C neither?

Similar investigations about the R2C optimisation of the Fribourg construction have been made by Göttel [2]. In particular, he compared Version 1 in our above listing with Version 3. His results were that the mean complement sizes of Version 3 are higher than in Version 1. He evaluated, however, only the mean values, but looking closely at the results suggests that the median values could be in favour of Version 3. Therefore, we decided to reinvestigate this question. Indeed, in our own results the median complement sizes of Version 3 are considerably lower than the ones of Version 1 and Version 2. We will further elaborate on this point in Chapter 2.

Versions 4 and 5 in our above listing are for investigating the M1 and M2 optimisations. As M2 requires M1, there are only these two possible combinations. Versions 6 and 7 then enhance the “better” one of Version 4 and 5 with R2C and its alternative R2C+C. As we will see in Chapter 2, the better one of Version 4 and 5 in terms of median complement sizes is Version 4. That is, the application of M2 results in a decline, rather than a gain, in performance compared to the application of M1 alone. We have to note at this point that such results are always specific to the used test set, and not universally valid. With a different test set, Version 5 might indeed be better than Version 4. As we will see in the next section, this is the case for our alternative test set consisting of the first four Michel automata.

Version 8, finally, is again the plain Fribourg construction, but this time the output automata are reduced by removing their unreachable and dead states. Comparing the results of Version 8 with Version 1 gives an idea of how many unreachable and dead states the Fribourg construction produces. This is inspired by the paper of the GOAL authors [13] in which the number of unreachable and dead states is one of the main metrics for assessing the performance of a construction.

Michel Test Set

The versions we tested for the Michel test set are the following:

1. Fribourg
2. Fribourg+R2C
3. Fribourg+M1

4. Fribourg+M1+M2
5. Fribourg+M1+M2+R2C
6. Fribourg+R

The rationale for choosing these versions is basically the same as for the GOAL test set with the following differences. First, Michel automata are complete, thus it is not necessary to apply the C option. Rather, R2C will automatically apply to all automata. Second, the aim of Version 5 is again to enhance the better one of Versions 3 and 4 with R2C. However, contrarily to the GOAL test set Fribourg+M1+M2 is better than Fribourg+M1 for the Michel automata. That is why Version 5 is Fribourg+M1+M2+R2C.

1.3.2 External Tests

In the so called external tests we compare the best version of the Fribourg construction with different complementation constructions, again for both the GOAL test set and the Michel test set. The concrete constructions we compared for the external tests are the following.

1. Piterman+EQ+RO
2. Slice+P+RO+MADJ+EG
3. Rank+TR+RO
4. Fribourg+M1+R2C (GOAL test set) and Fribourg+M1+M2+R2C (Michel test set)

For the alternative constructions, we chose the Piterman, Slice, and Rank construction. These constructions are representative for three of the four main complementation approaches, determinization-based, slice-based, and rank-based. The fourth complementation approach is Ramsey-based and there is an implementation of the Ramsey construction in GOAL. However, in preliminary tests, we realised that this construction is not performant enough to be used on our test sets within our time and memory constraints. The authors of GOAL came to a similar conclusion when they made similar experiments on the GOAL test set [13]. In their case, the Ramsey construction could not complete any of the 11,000 automata within the set time and memory constraints, and they went on to do the result analysis without the Ramsey construction.

The same applies to all the other constructions that are implemented in GOAL. We did preliminary tests with all of them and saw that only the three mentioned constructions, Piterman, Slice, and Rank, can be reasonably used on the GOAL test set⁹.

The three chosen constructions also have optimisations in their implementations in GOAL. To have a fair comparison to the best version of the Fribourg construction, which also uses optimisations, we activated the optimisation of these constructions as well. In particular, we chose those optimisations which are set as default in the GOAL GUI for each construction.

We use Fribourg+M1+R2C for the GOAL test set and Fribourg+M1+M2+R2C for the Michel test set. This is because these two versions are the most performant ones for the respective test sets.

1.3.3 Time and Memory Limits

We defined a time limit of 600 seconds CPU time and a memory limit of 1 GB per complementation task in the GOAL test set. That means, if the complementation of a single automaton is not finished after 600 seconds CPU time or uses more than 1 GB memory, the task is aborted and marked as a *timeout* or *memory excess*.

These limits correspond to the ones used by the experiments of the GOAL authors [13]. However, from their paper it is not clear if their time limit is in CPU time or wallclock time. But since they used different computing nodes, our results regarding the number of timeouts will anyway differ from theirs.

The reason for these limits is simply the restricted amount of time and memory resources that we have available for the experiments. In an ideal world, we would let every complementation task run to its end,

⁹The Safra construction would also have been possible, but the Safra construction is similar to the Piterman construction.

no matter how long it takes and how much memory it uses. This would give a perfectly unbiased picture of the results. By setting time and memory limits, we basically exclude the most difficult automata from the experiment. However, as mentioned, the practical reasons of limited time and computing power force us to make this compromise.

The timeout and memory limit of 1 GB apply just to the automata of the GOAL test set. For the Michel test set we did not set a timeout because we wanted each one of the four automata to finish. The longest complementation task of a Michel automaton consequently lasted 109,810 seconds which is about 28 hours. We set a very high memory limit of 14 GB for the Michel test set to avoid memory excesses as we wanted each automaton to successfully complete. The number of 14 GB is determined by the physically available memory on the used computing nodes. All Michel automata successfully completed with this amount of memory.

The timeout was implemented by the means of the `ulimit` Bash builtin¹⁰, which allows to set a maximum time after which running processes are killed. The memory limit was implemented by setting the maximum size of the Java heap, which can be done by the `-Xmx` option to the Java Virtual Machine (JVM). The heap is the main memory area of Java and the place where all the objects reside. Note that since our memory limit defines actually the size of the Java heap, the total amount of memory used by the process is higher than our limit, as Java has some other memory areas, for example for the JVM itself. However, this is a rather constant amount of memory and independent from the current automaton, so it does not disturb the relative comparisons of the results.

The presence of aborted complementation tasks requires the consideration of the so called effective samples in the result analysis, as introduced in the experiment paper of the GOAL authors. The effective samples are those automata which have been successfully completed by *all* constructions that are to be compared to each other. Imagine two constructions *A* and *B* where *A* is successful in complementing all the automata, whereas *B* has timeouts or memory excesses at 100 of the automata. If we would now take, for example, the median complement sizes of the two result sets without first extracting the effective samples, then *B* is likely to be assessed as too good relative to *A*, because *B*'s results do not include the 100 automata at which it failed, and which are thus likely to have large complement sizes with *B*. The same 100 automata would however be included in the results of *A*. Therefore, all the result analysis of the experiments with the GOAL test sets, that we present in Chapter 2, are based on the effective samples of the result sets.

1.3.4 Execution Environment

We executed the experiments on a high performance computing (HPC) computer cluster called UBELIX at the University of Bern¹¹. This cluster consists of different types of Linux-driven HPC computing nodes, and is managed by Oracle Grid Engine¹² (formerly known as Sun Grid Engine) version 6.2. Oracle Grid Engine is a so called load scheduler that is responsible for automatically distributing computing tasks to computing nodes.

The basic workflow of working on the cluster is to prepare a so called job and specify the resources that the job requires for running (time, memory, number of CPU cores, and so on). Then the job can be submitted to the grid engine. The grid engine first puts incoming jobs in a queue and then automatically dispatches them to suitable computing nodes as soon as the required capacity is available.

A job in our case is the execution of a construction (or version of the Fribourg construction) over an entire test set. Thus, we were running 22 jobs in total. We arranged that all jobs were run on identical nodes with the following specifications:

- Processor: Intel Xeon E5-2665 2.40GHz (64 bit)
- CPU cores: 16
- Operating System: Red Hat Enterprise Linux 6.6

Since we use the execution time (CPU time) as a secondary metrics, next to the complement sizes, it is important that all complementation tasks are executed on the same type of hardware. GOAL is

¹⁰<http://linux.die.net/man/1/bash>

¹¹<http://ubelix.unibe.ch>

¹²<http://www.oracle.com/us/products/tools/oracle-grid-engine-075549.html>

multithreaded and thus the jobs may use multiple CPU cores. The number of CPU cores a job may use is not restricted (up to the number of available cores on the node, in our case 16), but our observation is that the jobs use a rather small number of cores (2–3). Note that the measurement of the execution time is not affected by the number of cores a process uses, as the CPU times are measured separately on each core and then added together.

Chapter 2

Results and Discussion

Contents

| | | |
|------------|----------------------------------|-----------|
| 2.1 | Internal Tests | 18 |
| 2.1.1 | GOAL Test Set | 18 |
| 2.1.2 | Michel Test Set | 24 |
| 2.2 | External Tests | 26 |
| 2.2.1 | GOAL Test Set | 26 |
| 2.2.2 | Michel Automata | 29 |
| 2.3 | Discussion | 31 |
| 2.3.1 | Summary of the Results | 31 |
| 2.3.2 | Validity | 31 |

2.1 Internal Tests

2.1.1 GOAL Test Set

First of all, let us see how many unsuccessful complementation tasks there were with the internal tests on the GOAL test set. Table 2.1 shows the number of timeouts and memory excesses for each of the eight tested versions of the Fribourg construction.

| Construction | Timeouts | Memory excesses |
|-------------------|----------|-----------------|
| Fribourg | 48 | 0 |
| Fribourg+R2C | 30 | 0 |
| Fribourg+R2C+C | 54 | 0 |
| Fribourg+M1 | 2 | 0 |
| Fribourg+M1+M2 | 1 | 0 |
| Fribourg+M1+R2C | 1 | 0 |
| Fribourg+M1+R2C+C | 8 | 0 |
| Fribourg+R | 48 | 0 |

Table 2.1: Number of timeouts and memory excesses in the internal tests on the GOAL test set.

As we can see in the table, there were no memory excesses, that is, none of the 11,000 complementation tasks required more than 1 GB memory (actually, Java heap size). On the other hand, there is quite a number of timeouts. If we extract the effective samples from these result sets, we get a set of 10,939 automata. That means that these 10,939 automata have been successfully completed by all the versions of the Fribourg construction, whereas the remaining 61 automata (0.55%) provoked a timeout in at least one of the versions.

Our main analysis is now about the sizes of the complements of these 10,939 automata. With size we mean the number of states of an automaton. A stripchart as the one in Figure 2.1 is a good way to get a first glance of the complement sizes. Each horizontal strip contains 10,939 dots, each one corresponding to a produced complement automaton. The x-position of a dot indicates the size of the corresponding complement automaton.

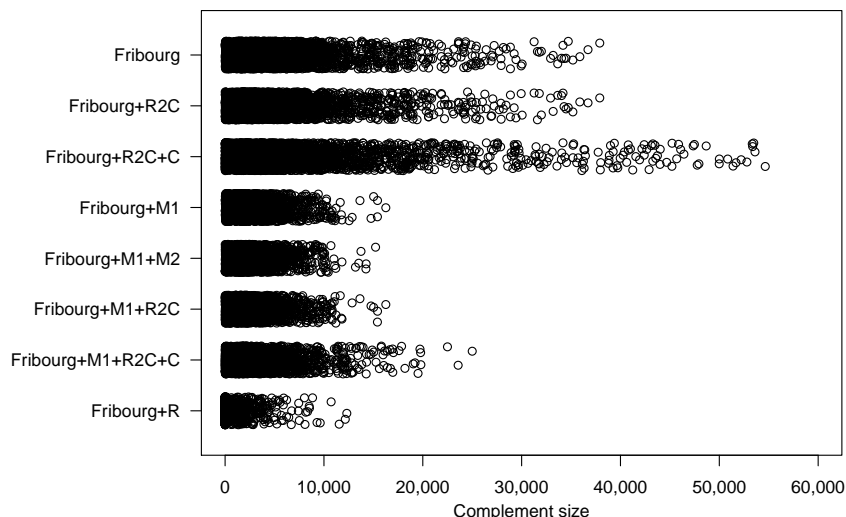


Figure 2.1: Stripchart with the complement sizes of the 10,939 effective samples of the GOAL test set.

The first thing to note is that the distribution of complement sizes is extremely right-skewed (also known as positive-skewed). The peak close to the left end of the x-axis and there is a long tail toward the right. This means that most of the complements are very small and the frequency decreases with increasing

complement size. Finally, there are some few complements which are very large. This distribution implies that the mean is generally higher than the median, because the mean is “dragged” to the right by the few very large complements.

Next, we can compare the distributions of the individual version to each other. While Fribourg and Fribourg+R2C have a similarly long tail, Fribourg+R2C+C has a considerably longer one. This clearly is the effect of increasing the size of 91% of the automata by adding a sink state in order to make them complete. Next, Fribourg+M1, Fribourg+M1+M2, and Fribourg+M1+R2C have similarly long tails that are significantly shorter than the ones of the previous versions. This indicates that the M1 optimisation is effective in reducing the size of otherwise large complements. Fribourg+M1+R2C+C, as expected, again increases the size of the tail due to the C option. Finally, Fribourg+R has a very short tail and an even higher concentration of very small complements. This is because the complements are pruned by removing their unreachable and dead states, and thus at least the complements of the 61.8% universal input automata are reduced to empty automata of size 1.

Having a first impression of the distribution of complement sizes, we can look at some statistics characterising this distribution. Table 2.2 shows the mean complement size along with the classic five-number summary consisting of minimum value, 25th percentile, median, 75th percentile and maximum value for each version of the Fribourg construction. As expected, the mean is always higher than the median. Generally, the median is a more robust metrics than the mean, because it is not affected by the value of outliers. This applies specifically to our distribution where some few very large complements might significantly increase the man whereas they leave the median unaffected. Therefore, the median will be our main metric, and the subsequent analyses will be based on the median.

| Construction | Mean | Min. | P25 | Median | P75 | Max. |
|-------------------|---------|------|-------|--------|---------|--------|
| Fribourg | 2,004.6 | 2 | 222.0 | 761.0 | 2,175.0 | 37,904 |
| Fribourg+R2C | 1,955.9 | 2 | 180.0 | 689.0 | 2,127.5 | 37,904 |
| Fribourg+R2C+C | 2,424.6 | 2 | 85.0 | 451.0 | 2,329.0 | 54,648 |
| Fribourg+M1 | 963.2 | 2 | 177.0 | 482.0 | 1,138.0 | 16,260 |
| Fribourg+M1+M2 | 958.0 | 2 | 181.0 | 496.0 | 1,156.5 | 15,223 |
| Fribourg+M1+R2C | 937.7 | 2 | 152.0 | 447.0 | 1,118.0 | 16,260 |
| Fribourg+M1+R2C+C | 1,062.6 | 2 | 83.0 | 331.0 | 1,208.5 | 25,002 |
| Fribourg+R | 136.3 | 1 | 1.0 | 1.0 | 21.0 | 12,312 |

Table 2.2: Statistics of the complement sizes of the 10,939 effective samples of the GOAL test set.

Going through the median values in Table 2.2 we encounter some surprises. To begin with, as expected, there is a decrease from Fribourg (761) to Fribourg+R2C (689). Then, however, there is a significant drop to 451 with Fribourg+R2C+C. This is a surprise insofar as by looking at Figure 2.1, Fribourg+R2C+C seems to have the worst performance at a first glance. Indeed it also has the highest mean, which is due to the group of extremely large complements. The median, however, is very low, even lower than the one of Fribourg+M1 with its significantly shorter tail in Figure 2.1. Also the 25th percentile of Fribourg+R2C+C is with 85 one of the lowest. Going to the other side of the median, however, the 75th percentile (2,329) is the largest of all versions. A possible characterisation of this phenomenon is that the C option (together with R2C) makes small complements smaller, and large complements larger. The diminishment of small complements is far-reaching enough that the median is affected by it and decreased significantly.

The next thing we see in Table 2.2 is that the median of Fribourg+M1 (482) is slightly lower than the median of Fribourg+M1+M2 (496). The same applies to the 25th and 75th percentile. This backs up our statement from Section 1.3.1 that Fribourg+M1 performs better on the GOAL test set than Fribourg+M1+M2. The difference is rather small (the median increase is by 2.9%), but it is enough to consider Fribourg+M1 as the better of the two versions, and to combine it therefore with R2C and R2C+C.

Fribourg+M1+R2C brings down the median from 482 to 447, with respect to Fribourg+M1. Also the 25th and 75th percentile are decreased. Adding the C option to Fribourg+M1+R2C, again causes the median to drop dramatically, from 447 to 331. The 25th percentile decreases from 152 to 83. The 75th

percentile however increases from 1,118 to 1,208.5. Here we have again the same picture of the effect of adding the C option that we had before. Namely that small complements are made smaller, and large complements are made larger.

Finally, the last row in Table 2.2 with Fribourg+R shows the extent of unreachable and dead states that the Fribourg construction produces. The median is 1, and a further analysis reveals that the single-state complements go up to the 61st percentile. That is, 61% of the complements have a size of 1. This is likely to correspond to the 61.8% of universal automata in the GOAL test set whose complements can be reduced to empty automata with a single state.

Up to now, we only looked at statistics aggregated over the entire test set. But as we know from Section 1.2.1, the 11,000 automata of the GOAL test set are divided into 110 classes which consist of combinations of 11 transition densities and 10 acceptance densities. The transition densities range from 1 to 3 in steps of 0.2, and the acceptance densities range from 0.1 to 1 in steps of 0.1. In each class there are 100 automata. It is assumable that the median complement sizes are different for different classes. In the next step of the analysis, we want to analyse exactly this point. How do the median complement sizes of the different classes compare to each other, and can we identify a pattern of the classes that result in large and small complements?

This type of analysis naturally results in three-dimensional data. The classes themselves have two dimensions (the transition densities and the acceptance densities), and each of them has a third dimension consisting of the median complement size of this class. One way to present such data is by, for example, a 11x10 matrix with the rows being the transition densities, the columns the acceptance densities, and the cells the median values. Another way is by so called perspective plots. We used both of these ways in Section 1.2.1 when we analysed the number of complete and universal automata in each class of the GOAL test set. In the present analysis, we will use perspective plots and omit the matrices for space reasons. However, we present the corresponding matrices in Appendix B.2. In this way, the interested reader may consult the exact values of the median complement sizes for each class. In the same way, we restrict ourselves to the median complement sizes as the third dimension. The same analysis could also be done, for example, for the mean, the 25th percentile, or the 75th percentile of complement sizes. However, as mentioned, we consider the median to be the most meaningful statistics, and thus exclusively focus on it for the matter of conciseness.

Figures ?? and ?? show the perspective plots of the median complement sizes for the 110 classes of the GOAL test set. Each crossing of two lines on the xy-plane represents a class, and the z-value (height) of this crossing indicates the median complement size of this class. As a help for the orientation, if one imagines the corresponding matrices, as they are presented in Appendix B.2, then in the perspective plots we are looking at these matrices from the bottom-right corner. That is, the class that is in the top-left corner of a matrix (transition density 1.0 and acceptance density 0.1) is the one that is the farthest away from the viewer in the corresponding perspective plot. This orientation holds for all the remaining perspective plots in this thesis. The colours of the squares in the perspective plot (called facets) are a function of the z-values of their four corners and are selected to draw an analogy with topographical terrain maps.

The first thing we notice when looking at the plots in Figure 2.2 and ?? is that there are indeed large differences in the median complement sizes across the transition/acceptance density classes. Figuratively speaking, there is an oblong mountain (or hill) in the northern part of the area in east-west orientation, whose ridge increases in height from east to west and stays high until the western edge of the area. The mountain is located roughly in the area of transition densities between 1.2 and 2.2 and acceptance densities from 0.1 up to 0.9. This means that the automata of these classes are apparently harder to complement than the automata of other classes, as they result more frequently in larger complements.

We will further elaborate on the reasons of these different complement sizes across classes, and also try to characterise easy, medium, and hard automata for the Fribourg construction, at the end of this subsection. For now, we focus on the relative differences of median complement sizes between the different versions of the Fribourg construction.

Looking at Figure 2.2, the perspective plots for Fribourg and Fribourg+R2C are rather similar. The top of the mountain ridge is at between 3,500 and 4,000 states with a single peak of around 4900 states in the class with transition density 1.6 and acceptance density 0.3. From Table 2.2 we can learn that the

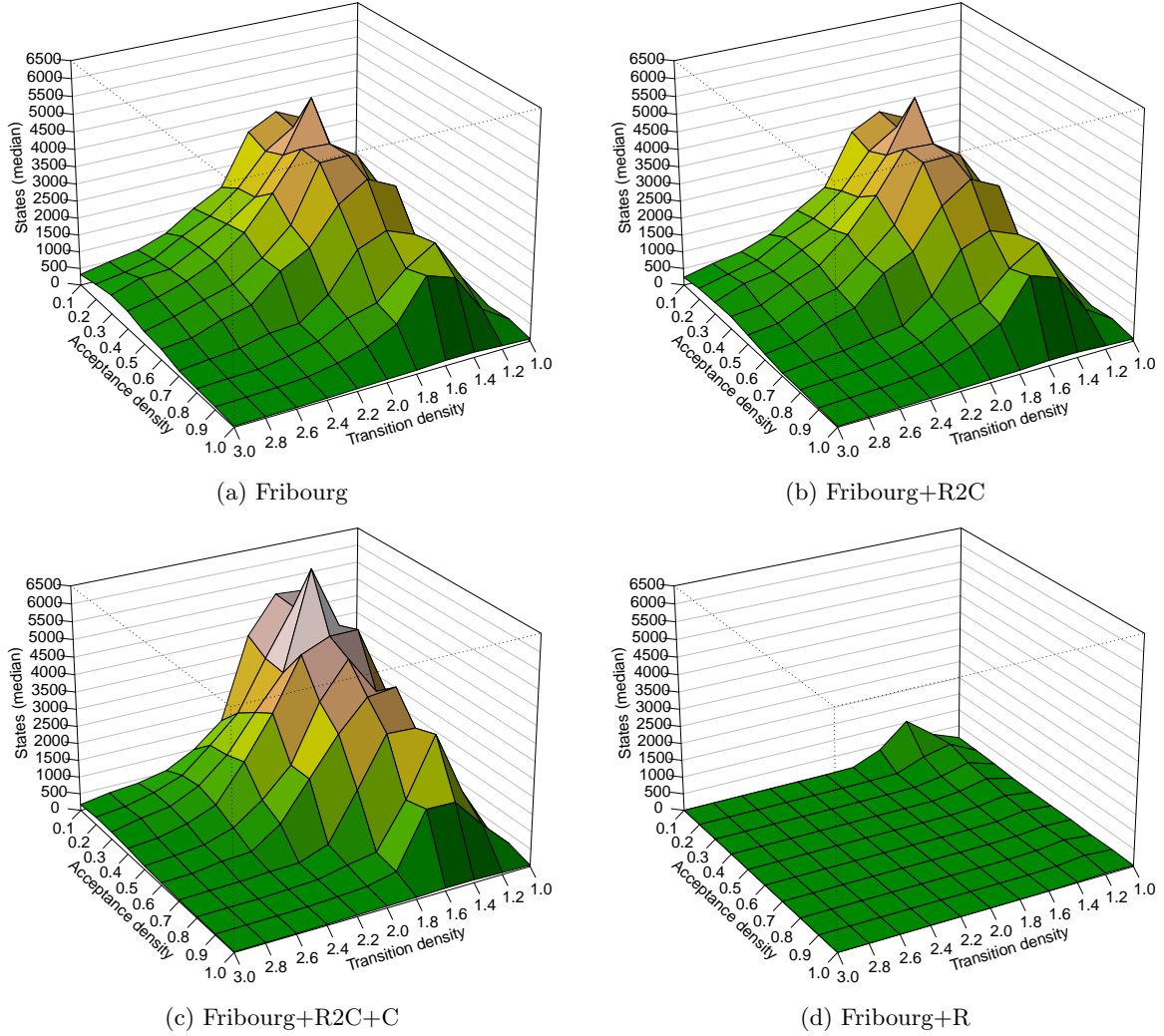


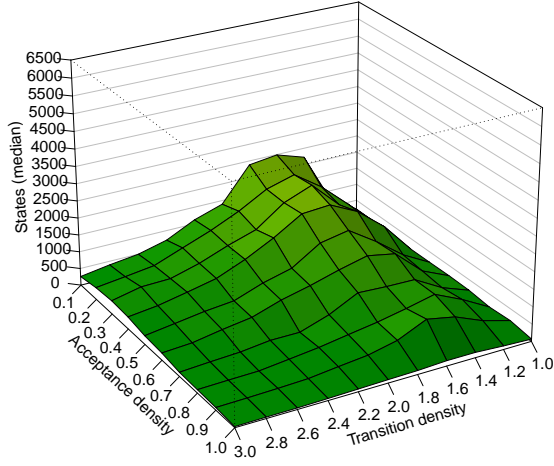
Figure 2.2: Median complement sizes of the 110 transition density/acceptance density classes of the 10,939 effective samples of the GOAL test set.

overall median complement size is 761 for Fribourg and 689 for Fribourg+R2C. These low values might surprise at first as the mountain, which is much higher, seems to dominate. However, by taking a closer look, it becomes apparent that around half of the classes are in rather low terrain (less than 1,000 states). Furthermore, the heights of the mountain peak do not allow to deduce anything about the overall median, because the median is not affected by the actual values of the data points which are greater than the median. This is the main characteristics of the median and applies to all the subsequent perspective plots in this chapter. The means in turn are 2,004.6 for Fribourg and 1955.9 for Fribourg+R2C.

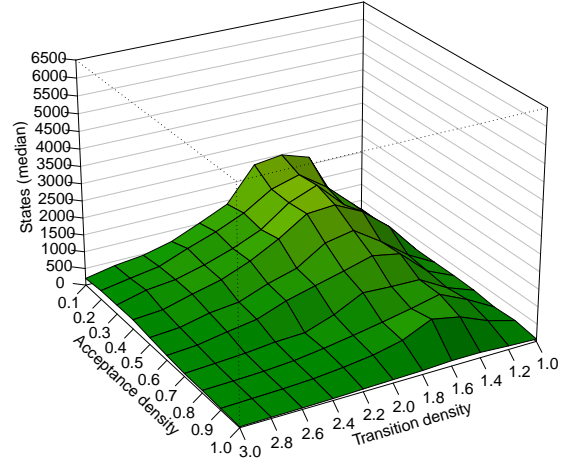
Fribourg+R2C+C in Figure 2.2 (c) has an even considerably higher mountain than the previous two versions. The top of the ridge is at around 5,000 states and the peak at the class with transition density 1.6 and acceptance density 0.3 has close to 6,500 states. This increase in height is however not reflected in the overall median of Fribourg+R2C+C relative to Fribourg+R2C. As we have seen in Table 2.2, the median of Fribourg+R2C+C is 451, and thus 34.5% lower than the median of Fribourg+R2C. By taking a closer look at the perspective plots, the reason for this can be seen, as the low areas of Fribourg+R2C+C are indeed slightly lower than the low areas of Fribourg+R2C. The significantly higher high areas of Fribourg+R2C+C do not have an effect on the overall median. However, they do have an effect on the overall mean which with 2,424.6 is indeed 24% higher than the one of Fribourg+R2C (1,955.9).

Going from the first three perspective plot in Figure 2.2 to the perspective plot of Fribourg+R is like going from the Swiss Alps to a Dutch polder. The mountain shrinks to a small hillock and the rest of

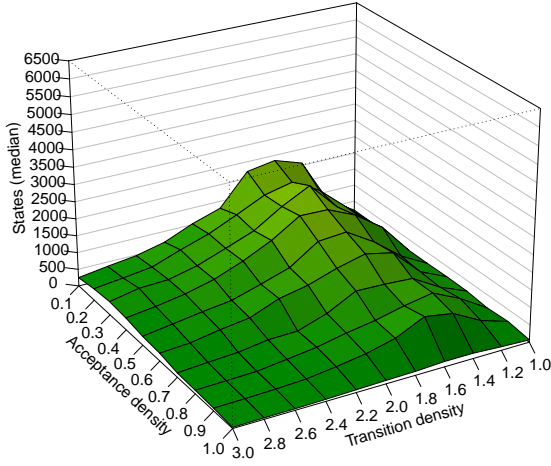
the terrain is low and flat. This is because so many complements of the Fribourg construction can be reduced to very small sizes by removing their unreachable and dead states. A look at the corresponding matrix in Appendix B.2 reveals that 68 of the 110 classes have a median complement size of 1. If we further compare this matrix to the matrix with the number of universal automata in Figure ?? (b), we see that all the classes with a median of 1 contain more than 50 universal automata, and the classes with a median greater than 1 contain less than 50 universal automata. There is a total of 100 automata per class. This makes sense as the complements of universal automata are empty automata, and every empty automaton can be reduced to an automaton with a single non-accepting state. Looking at the classes with a median greater than 1, we see that their values are still considerably lower than the ones of the plain Fribourg construction, which indicates that the Fribourg construction generates a large number of unreachable and dead states. However, as already mentioned, this thread of analysis is not the focus of the present thesis, and we leave it for future work.



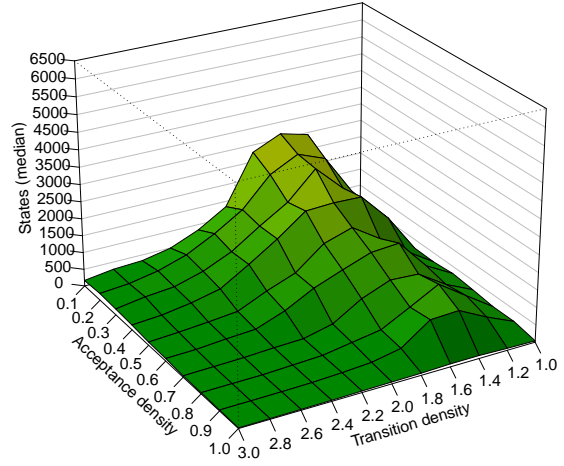
(a) Fribourg+M1



(b) Fribourg+M1+R2C



(c) Fribourg+M1+M2



(d) Fribourg+M1+R2C+C

Figure ?? shows the perspective plots of the remaining four versions of the Fribourg construction, all of which include the M1 optimisation. The first thing we note is that, while the mountain is still there, its height shrinks significantly. For Fribourg+M1, and Fribourg+M1+R2C, the top of the ridge is between is around 2,500 states. This is reflected by the overall means of these two versions compared to their counterparts without the M1 optimisation, Fribourg, and Fribourg+R2C. The decrease of the overall mean from Fribourg to Fribourg+M1 is by 52% (from 2004.6 to 963.2) and from Fribourg+R2C to Fribourg+M1+R2C by 52.1% (from 1955.9 to 937.7). The decreases of the overall medians are by 36.6% (from 761 to 482), and 35.1% (from 689 to 447) for the same two pairs of versions. With this we can confirm that the M1 optimisation brings a significant performance gain.

Regarding the M2 optimisation, we can see that the mountain ridge in the Fribourg+M1+M2 perspective

plot is slightly lower than the one in the Fribourg+M1 perspective plot. The flatland regions, however, seem to not change much. This is reflected by the overall mean of Fribourg+M1+M2 which is slightly lower than in Fribourg+M1 (958.9 opposed to 963.2). The overall median, on the other hand, is higher for Fribourg+M1+M2 than for Fribourg+M1 (496 opposed to 482). An interpretation of this behaviour is that the application of the M2 optimisation results in smaller complements for *some* input automata. These automata are especially the “hard” ones that produce large complements. This positive effect of M2 does however not affect enough input automata, especially not the “easy” automata, as to improve the overall performance of the construction in terms of median complement size. As already stated previously, we consider therefore Fribourg+M1 as the better construction on the GOAL test set than Fribourg+M1+M2.

Finally, Fribourg+M1+R2C+C differs from Fribourg+M1+R2C in a similar way Fribourg+R2C+C differs from Fribourg+R2C. The higher regions get higher and the lower regions get lower, that is, a performance decline on “hard” automata, but a performance gain on easy automata. The performance gain on the easy automata is however effective enough to decrease the overall median from 447 to 331, which is minus 26%.

With 331 states, Fribourg+M1+R2C+C has the lowest median of all the versions (except Fribourg+R which is a special case because it modifies the output of the construction). However, we still declare Fribourg+M1+R2C as the winner on the GOAL test set, mainly for two reasons. First, while Fribourg+M1+R2C+C has a lower median, the mean is still higher (1062.6 to 937.7 which is a plus of 13.3%). This results from the complements of the hard automata, which are larger than with Fribourg+M1+R2C. From a practical point of view, the mean might be relevant, because it relates more directly to the required computing resources than the median. For example the execution CPU time per complementation task, that we also measured along with our experiments, is 25.4% higher for Fribourg+M1+R2C+C than for Fribourg+M1+R2C. The increase in the average execution time per automaton is from 4.44 to 5.57 seconds and in the total execution time from 48,572 seconds (\approx 135 hours) to 60,919 seconds (\approx 169 hours). Fribourg+M1+R2C, on the other hand, has the lowest mean of all versions. The second reason that we choose Fribourg+M1+R2C as the winner and not Fribourg+M1+R2C+C is that the C option is not a real part of the construction. It actually modifies the input automata before the construction starts in order to make them better suited for the construction. Fribourg+M1+R2C, on the other hand, includes only native options, and input and output of the construction are not modified. This should also make it fairer to compare the Fribourg construction to other constructions in the external tests (Section ??).

As we have seen, there are big difference in the complement sizes across the different classes of the GOAL test set. Furthermore, there is a certain pattern, namely the mountain in the north-western region of the class matrix. Automata of classes that are in the mountain region seem to be harder to complement than automata from the classes in the flatland region. We attempted to categorise the classes of the GOAL test set into the three groups “easy”, “medium”, and “hard”. To do so, we first averaged the matrices with the median complement sizes of all the eight versions of the Fribourg construction. In this way, we have a mean median complement size for each class. Then we defined two breakpoints that divide the classes into easy, medium, and hard groups. The breakpoints 500 and 1,600 result in an appropriate groups that seem to capture the reality well. The result with these breakpoints can be seen in Figure ??.

As can be seen in Figure ??, there are 53 easy, 36 medium, and 21 hard classes. The easy classes are mainly those with extreme values. In particular, all the classes with a low or high transition density of 0.1, or 2.8 and 3.0, and a high acceptance density of 1.0 are easy. Furthermore, there is a “triangle” of easy classes between transition densities 2.0 and 2.6. and acceptance densities 0.5 and 0.9. The higher the transition density, the lower the acceptance density may be so that the corresponding class is still easy. The hard classes are roughly those with a transition density between 1.4 and 1.8 and an acceptance density between 0.1 and 0.6. The medium classes finally are situated as a “belt” around the hard classes.

It is interesting that the extreme values of transition density and acceptance density result in easy automata. With a transition density of 1.0 and an alphabet size of 2, each of the 15 states has on average two outgoing and two incoming transitions¹. With a transition density of 3.0, each state has on average 6 outgoing and 6 incoming transitions. These low or high connectivity seems to considerably simplify the

¹The transition density multiplied with the number of states defines the number of transition for each symbol of the alphabet in the automaton, see Section 1.2.1.

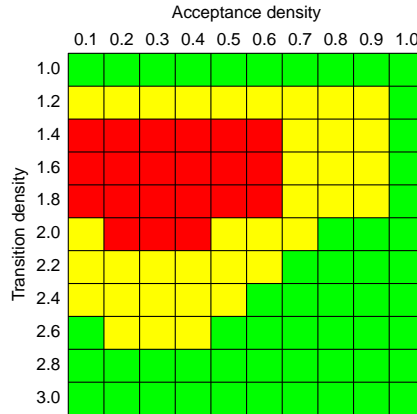


Figure 2.4: Difficulty categories of the transition/acceptance density classes of the GOAL test set. Green: easy; yellow: medium; red: hard.

complementation task. The same applies to a high acceptance density of 1.0, which means that every state is an accepting state². Generally, we can say that automata with high acceptance densities are easier to complement than automata with lower acceptance densities. This also means that the pattern of easy automata at the extreme values of transition and acceptance density, does not apply to the lower extreme of the acceptance density. Automata with a very low acceptance density of 0.1 are hard to complement—unless they are made easy by a low or high transition density.

Another interesting point is that the hard automata have transition densities between 1.4 and 1.8. It seems that this range of transition densities is the crucial factor in the hardness of a complementation task, and that it is only alleviated by a growing acceptance density. This explains the decline of the mountain ridge from west to east.

Summarising we can say that transition densities between 1.4 and 1.8 produce the hardest complementation tasks, and that to the both sides the difficulty steadily decreases with declining or growing transition density. Furthermore, a growing acceptance density generally implies easier complementation tasks.

2.1.2 Michel Test Set

Our second test set consists of the four Michel automata with $m = \{1, \dots, 4\}$ that are listed in Figure 1.3. They have 3, 4, 5, and 6 states, respectively. Our main motivation to use these automata is to test our complementation construction on very difficult automata. In this way we can get an idea about the worst-case state complexity of our construction.

We complemented the four Michel automata with the six versions of the Fribourg construction which are listed in Section 1.3.1. We did not impose a time or memory limitation for the individual complementation tasks, as we wanted every task to finish. The resulting complement sizes are listed in Table 2.3.

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|----------|--------------|------------|
| Fribourg | 57 | 843 | 14,535 | 287,907 | $(1.35n)^n$ | 0.01% |
| Fribourg+R2C | 33 | 467 | 8,271 | 168,291 | $(1.24n)^n$ | 0.06% |
| Fribourg+M1 | 44 | 448 | 5,506 | 81,765 | $(1.10n)^n$ | 0.07% |
| Fribourg+M1+M2 | 42 | 402 | 4,404 | 57,116 | $(1.03n)^n$ | 0.12% |
| Fribourg+M1+M2+R2C | 28 | 269 | 3,168 | 43,957 | $(0.99n)^n$ | 0.04% |
| Fribourg+R | 18 | 95 | 528 | 3,315 | $(0.64n)^n$ | 0.35% |

Table 2.3: Complement sizes of the Michel automata with $m = \{1, \dots, 4\}$ and 3, 4, 5, and 6 states, respectively.

²The acceptance density defines the percentage of states that are accepting, see Section 1.2.1 as well.

In the second-last column “Fitted curve” of Table 2.3, we fitted a function of the form $(an)^n$ to the measured data points of the previous four columns. These data points consist of the sizes of the four Michel automata (3, 4, 5, and 6) as. The fitted function $(an)^n$ can be seen as an approximated generalisation of the state growth, where n is the size of the input automaton. The last column “Std. error” shows the standard error that resulted from the fit of the previous column.

We can see in Table 2.3 that the state growths are indeed very large. For example, complementing Michel 4, which has six states, with the plain Fribourg construction results in a complement of 287,907 states. However, the optimisations R2C, M1, and M2 have a large influence on the complement sizes. If we consider Michel 4, then the R2C optimisation alone reduces the complement size from 287,907 to 168,291 which is a reduction of 51.5%. The M1 optimisation has an even larger influence as it reduces the complement size from 287,907 to 81,765 which is a reduction of 71.6%. Adding M2 to M1 further reduces the complement size of Fribourg+M1 by 30.1% (from 81,765 to 57,116). Finally, adding R2C on top of M1 and M2 brings a further reduction of 23% (from 57,116 to 43,957). If we compare the most efficient version (Fribourg+M1+M2+R2C) with the least efficient one (Fribourg), then the three optimisations reduce the complement size by 85.7%, or in other words, the complement size of Fribourg+M1+M2+R2C is 15.3% of the complement size of Fribourg.

It is interesting to see that for the Michel automata Fribourg+M1+M2 is more efficient than Fribourg+M1. For the GOAL test set, Fribourg+M1+M2 had a slightly higher overall median than Fribourg+M1 although it had a slightly lower overall mean. We identified in Section 2.1.1 that the M2 optimisation has a positive effect only on some automata, and that these are mostly the hard automata. Michel automata are very hard automata, and indeed the M2 optimisation has a considerably positive effect. These results support thus the observation we made in Section 2.1.1.

The special version Fribourg+R yields very small complements compared to the other versions. This tells us that the complements of the other versions contain a large number of unreachable and dead states. For example, the complement of Michel 4 of Fribourg+R (3,315 states) is 1.2% of the size of the complement of Fribourg. This means that 98.8% of the 287,907 states of the complement of Fribourg are unreachable and dead states. This is actually not surprising, because, following the proof of Michel [5][11], the smallest possible complement of Michel 4 has 24 states. This is because Michel 4 has $m = 4$ and Michel proved that the complement has at least size $m!$. This means that the Fribourg construction is very far from reaching these optimal complement sizes.

Up to now we just looked at the specific results of Michel 4. The fitted functions of the form $(an)^n$ summarise the results of all the four Michel automata. These functions give us reference points for the worst-case state complexities of the different versions of the Fribourg construction. For example, for the plain Fribourg construction with its fitted function of $(1.35n)^n$, we know now empirically that this construction produces complements of size $(1.35n)^n$, where n is the size of the input automaton. This means that the real (theoretical) worst-case complexity cannot be lower than $(1.35n)^n$ (but it can still be higher). This bound decreases for the different versions of the Fribourg construction until $(0.99n)^n$ for Fribourg+M1+M2+R2C. This value is still greater than the lower bound of $(0.76n)^n$ for Büchi complementation determined by Yan [?].

We also measured the execution times for the individual complementation tasks. This time was measured in CPU time, that is, the time the task is running on the CPU. Table 2.4 shows the measured values in seconds. We can see that the difference between the least and most efficient version is bigger than for the complement sizes. For example for Michel 4, Fribourg+M1+M2+R2C is more than 43 times faster than Fribourg (2,332.6 seconds compared to 100,976 seconds). In more familiar unities, this corresponds to approximately 39 minutes for Fribourg+M1+M2+R2C against 28 hours for Fribourg. We also fitted functions of the form $(an)^n$ to the measured execution times where n is the number of states of the input automaton, and the value of the function is the execution time of the task in CPU time seconds.

The fitted functions that we calculated for the measured complement sizes and execution times are based on four data points and basically only valid for the first four Michel automata. They can thus not be used to reliably extrapolate values for larger Michel automata. However, it is still interesting to do such an extrapolation in order to see the involved complexity and to show why we were restricted to include only the first four Michel automata in the test set. In Table 2.5, we show extrapolated values for the complement sizes and execution times for the plain Fribourg construction (the least efficient one), based on the corresponding fitted functions. The table includes the values for the Michel automata 5 to 8, which

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|-----------|--------------|------------|
| Fribourg | 2.3 | 4.0 | 88.8 | 100,976.0 | $(1.14n)^n$ | 0.64% |
| Fribourg+R2C | 2.3 | 3.4 | 27.4 | 27,938.3 | $(0.92n)^n$ | 0.64% |
| Fribourg+M1 | 2.2 | 3.6 | 17.9 | 6,508.4 | $(0.72n)^n$ | 0.63% |
| Fribourg+M1+M2 | 2.3 | 3.5 | 13.8 | 2,707.4 | $(0.62n)^n$ | 0.62% |
| Fribourg+M1+M2+R2C | 2.5 | 3.5 | 10.8 | 2,332.6 | $(0.61n)^n$ | 0.62% |
| Fribourg+R | 2.4 | 3.7 | 86.0 | 101,809.6 | $(1.14n)^n$ | 0.64% |

Table 2.4: Execution times.

have 7 to 10 states.

| Automaton | Compl. size $(1.35n)^n$ | Exec. time $(1.14n)^n$ | \approx days/months/years |
|-----------|-------------------------|------------------------|-----------------------------|
| Michel 5 | 6,882,980 | 2,020,385 | 23 days |
| Michel 6 | 189,905,394 | 46,789,245 | 18 months |
| Michel 7 | 5,939,189,262 | 1,228,250,634 | 39 years |
| Michel 8 | 207,621,228,081 | 36,039,825,529 | 1,142 years |

Table 2.5: Extrapolated values for the complement sizes and execution times for the Michel automata with $m = \{5, \dots, 10\}$ with the plain Fribourg construction.

According to the fitted state growth function, the complement of Michel 5 would have nearly 7 million states, and the complement of Michel 8 even more than 207 billions. Already the computation of the 7 million states of Michel 5 would most probably exceed the available memory resources in our computing environment. Regarding the extrapolated execution times, the complementation of Michel 5 would take 23 days. This would similarly exceed the maximum running time of a job on the computer cluster on which we executed the computations. And even without these restriction, at least starting from Michel 6, the execution times between 18 months and 1,142 years are clearly too long to be included in a master's thesis.

2.2 External Tests

In the external tests we compare the results of the most efficient version of the Fribourg construction to the results of three pre-implemented constructions in GOAL. These constructions are Piterman+EQ+RO, Slice+P+RO+MADJ+EG, and Rank+TR+RO. The most efficient version of the Fribourg construction differs for the two test sets. For the GOAL test set it is Fribourg+M1+R2C, and for the Michel test set it is Fribourg+M1+M2+R2C. In the following two sections we present the results of all these constructions on the two test sets.

2.2.1 GOAL Test Set

For the GOAL test set we compared Fribourg+M1+R2C with Piterman+EQ+RO, Slice+P+RO+MADJ+EG, and Rank+TR+RO. As for the internal tests, we set a time limit of 600 seconds CPU time, and a memory limit determined by a Java heap size of 1 GB per complementation task. If a task does not finish within these limits, it is aborted and marked as either a timed out or memory exceeded. Table 2.6 shows the number of timeouts and memory excesses that we observed for the four constructions.

Most striking in Table 2.6 is the high number of unsuccessful tasks for the Rank construction. 3,317 of the 11,000 automata (33.8%) were aborted due to a timeout, and further 83 (0.8%) due to a memory excess. Regarding the other constructions, there are just two timeouts for the Piterman, and one timeout for the Fribourg construction.

Determining the effective samples of these runs gives a number of 7,204 automata (65.5%). In Figure 2.5 we present the complement sizes of these 7,204 effective samples as a stripchart. Looking at this picture

| Construction | Timeouts | Memory excesses |
|--------------------|----------|-----------------|
| Piterman+EQ+RO | 2 | 0 |
| Slice+P+RO+MADJ+EG | 0 | 0 |
| Rank+TR+RO | 3,713 | 83 |
| Fribourg+M1+R2C | 1 | 0 |

Table 2.6: Number of timeouts and memory excesses.

makes the reason for the high number of aborted tasks of the Rank construction apparent. Rank+TR+RO produces a very big number of very large complements compared to the other constructions.

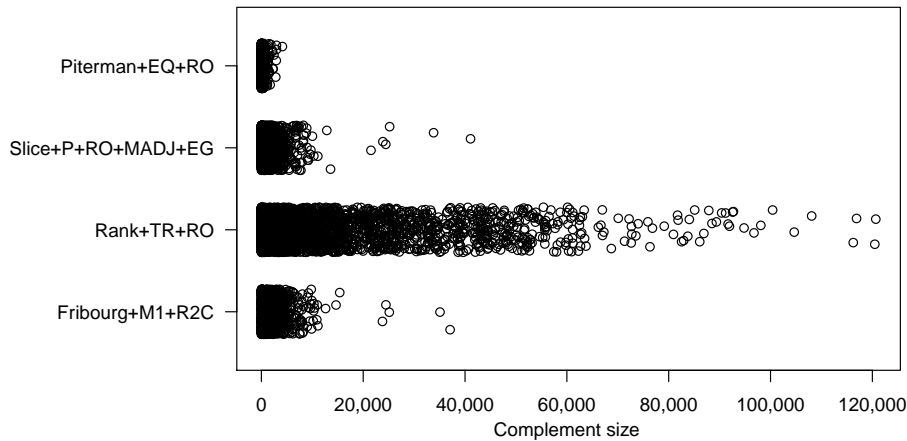


Figure 2.5: Complement sizes of the 7,204 effective samples.

But from the stripchart in Figure 2.6 alone we cannot yet tell whether the Rank construction *generally* produces larger complements than the other constructions, or if this holds just for *some* automata. Therefore, we calculated further statistics of the complement sizes of the 7,204 effective samples in Table 2.7.

| Construction | Mean | Min. | P25 | Median | P75 | Max. |
|--------------------|---------|------|-------|--------|---------|---------|
| Piterman+EQ+RO | 106.0 | 1 | 29.0 | 58.0 | 121.0 | 4,126 |
| Slice+P+RO+MADJ+EG | 555.4 | 2 | 70.0 | 202.0 | 596.0 | 41,081 |
| Rank+TR+RO | 5,255.6 | 2 | 81.0 | 254.5 | 3,178.2 | 120,674 |
| Fribourg+M1+R2C | 662.9 | 2 | 101.0 | 269.0 | 754.5 | 37,068 |

Table 2.7: Aggregated statistics of complement sizes of the 7,204 effective samples.

And indeed, the 25th percentile and the median of Rank are higher than for Piterman and Slice, but still lower than for our Fribourg construction. However, the picture changes dramatically for the 75th percentile where the value of Rank is more than four times higher than the value for Fribourg. Also the mean of Rank is many times higher than the means of all the other constructions. A possible explanation for this is that the Rank construction performs comparably with the other constructions on easy automata. For the one half of easier automata of the 7,204 effective samples, Rank performs even better than our Fribourg construction, as shown by the median values. For harder automata, however, the performance of Rank is decreased tremendously, and the construction results in very large complements that are beyond any comparison with the complements of the other constructions. In addition, the automata that are hardest for Rank are not even included in this analysis. These are the 34.5% of the test set that Rank could not complete within the given resource restrictions. If we would have these results and include them in the analysis, then the picture would probably look even much more distorted.

What we cannot tell is whether the automata which are hard for Rank are the same that are hard for the other constructions. However, as we will see later, we think that this is not necessarily the case.

Given the large number of unsuccessful complementation tasks of Rank we decided to do the main analysis and comparison of the results without the Rank construction. Because with the Rank construction would basically exclude more than one third of the tasks that have been successfully completed by the other constructions from the result analysis. Our main interest is to compare the performance of the Fribourg construction to the other constructions. Above we have already compared it to the special behaviour of the Rank construction, and by excluding Rank from the further analysis, we can compare the Fribourg constructions in full detail with the Piterman and Slice constructions.

Without Rank there are 10,998 effective samples, that is, there are only two automata that have not been completed by all the three of Piterman, Slice, and Fribourg. In Figure 2.6 we display the complement sizes of these 10,998 effective samples as a stripchart.

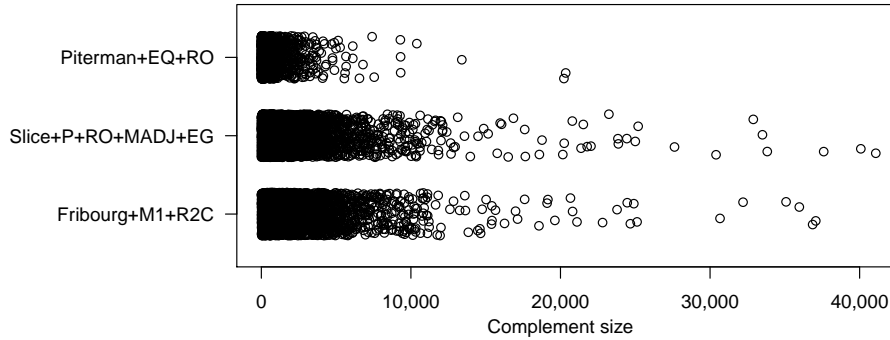


Figure 2.6: Complement sizes of the 10,998 effective samples.

From the stripchart we can see that Fribourg and Slice have a comparable distribution of complement sizes, whereas Piterman has a considerably higher concentration of small complement sizes. We can say that Piterman generally produces smaller complement than Fribourg and Slice.

We present the statistics of these distributions in Table 2.8. Indeed, for all statistics Piterman has values that are in the order of magnitudes lower than the ones of Fribourg and Slice. It is interesting that this order of magnitude for the mean, 25th percentile, median, and 75th percentile is very roughly five. It seems that Piterman produces throughout complements that are around five times smaller than the complements of Fribourg and Slice.

| Construction | Mean | Min. | P25 | Median | P75 | Max. |
|--------------------|---------|------|-------|--------|---------|--------|
| Piterman+EQ+RO | 209.6 | 1 | 38.0 | 80.0 | 183.0 | 20,349 |
| Slice+P+RO+MADJ+EG | 949.4 | 2 | 120.0 | 396.0 | 1,003.0 | 41,081 |
| Fribourg+M1+R2C | 1,017.3 | 2 | 153.0 | 452.0 | 1,134.0 | 37,068 |

Table 2.8: Aggregated statistics of complement sizes of the 10,998 effective samples without Rank.

Comparing Fribourg and Slice, there is a slight favour for Slice. Mean, 25th percentile, median, and 75th percentile are lower for Slice than for Fribourg by 6.7%, 21.6%, 12.4%, and 11.6%, respectively. We have to conclude that from an overall point of view, the Fribourg construction has the second-worst performance for the GOAL test set after Piterman and Slice, and before Rank.

But now let us look at the results for the 110 classes of transition density and acceptance density classes of the GOAL test set. As already for the internal tests, we take the median complement size as our statistic of interest and calculate it for each of the 110 classes. The results as perspective plots are shown in Figure 2.7. The same data in matrix form can be found in Appendix B.2.

First of all, note that the data displayed in the perspective plot of Fribourg+M1+R2C in Figure 2.7 (c) is the same as the one from the internal tests in in Figure ?? (b). The only difference is the range of the z -axis, which in Figure ?? ranges from 0 to 6,500, and in Figure 2.7 from 0 to only 3,000.

We can see that the pattern for Fribourg and Slice are very similar. The median complement sizes in the individual classes do not differ a lot, both relatively and absolutely. However, the medians of Fribourg

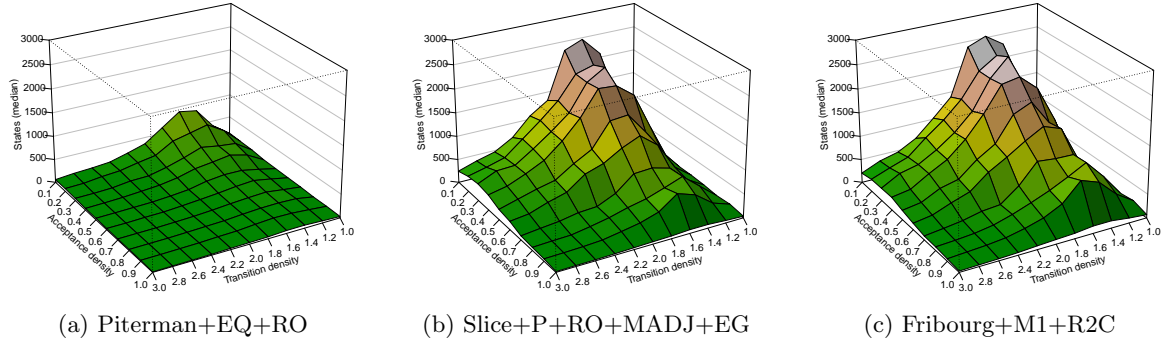


Figure 2.7: Median complement sizes (10,998 samples)

seem to be throughout (with some exceptions) slightly higher than the ones of Slice. This means that Fribourg and Slice seem to have similar strengths and weaknesses, but Slice is slightly more efficient on the tested automata.

Piterman, as expected, has medians that are throughout multiple times lower than the corresponding medians of Fribourg and Slice. The basic pattern, however, is still similar. There is a mountain ridge along the classes with a transition density of 1.6 with its top in the class with transition density 1.6 and acceptance density 0.1. Thus, this supports our claim from before that throughout all of the tested automata, Piterman produces complements that are multiple times smaller than the corresponding ones of Fribourg and Slice.

2.2.2 Michel Automata

We tested the same three third-party constructions, Piterman+EQ+RO, Slice+P+RO+MADJ+EG, and Rank+TR+RO on the Michel automata 1 to 4 and compared the results with the version of the Fribourg construction that was most efficient on the Michel automata in the internal tests. This version is Fribourg+M1+M2+R2C. As for the internal tests, we did not set a time or memory limitation as we wanted every task to successfully finish.

The resulting complement sizes are shown in Table 2.8. Again, we fitted a function of the form $(an)^n$ to the four measured data points of each construction and calculated the standard error of this fit.

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|----------|--------------|------------|
| Piterman+EQ+RO | 23 | 251 | 5,167 | 175,041 | $(1.25n)^n$ | 0.29% |
| Slice+P+RO+MADJ+EG | 35 | 431 | 6,786 | 123,180 | $(1.18n)^n$ | 0.02% |
| Rank+TR+RO | 23 | 181 | 1,884 | 25,985 | $(0.91n)^n$ | 0.01% |
| Fribourg+M1+M2+R2C | 28 | 269 | 3,168 | 43,957 | $(0.99n)^n$ | 0.04% |

Figure 2.8: Complement sizes of the first four Michel automata.

Considering the results of the GOAL test set, the results in Table 2.8 are surprising. Rank is the most efficient construction. It produces the smallest complements for all Michel automata, and with $(0.91n)^n$ it has the flattest fitted curve of all constructions. This is surprising because for the GOAL test set, Rank produced by far the largest complements, and 34.5% of the test data could not even be completed within the given time and memory limits. With the Michel automata, however, the case seems to be reversed and Rank produces by far the smallest complements.

Rank is followed by the Fribourg construction, which has the second-smallest complements for Michel 3 and 4, and with $(0.99n)^n$ the second-flattest fitted curve. The complements of Michel 2, 3, and 4 of the Fribourg construction are bigger than the ones of the Rank construction by 48.6%, 68.2%, and 69.2%, respectively. With this, our Fribourg construction is relatively close to the clear winner, which is the Rank construction.

The next in the ranking is the Slice construction with a fitted curve of $(1.18n)^n$. for Michel 1 to 3, this is actually the worst construction, but then for Michel 4, the complement is smaller than the one of Piterman what results in the flatter fitted curve. The gap to the second-ranked construction, Fribourg, is big. The complement sizes of Michel 2 to 4 exceed the ones of Fribourg by 60.2%, 114.2%, and 180.2%, respectively. This is also a remarkable point, because for the GOAL test set, Fribourg and Slice showed a very similar performance.

The last in the ranking is Piterman with a fitted curve of $(1.25n)^n$. However, special for Piterman is that it has the smallest complement for Michel 1 (together with Rank), the second-smallest for Michel 2, the third-smallest for Michel 3, and the largest for Michel 4. It is actually the large complement of Michel 4 that makes Piterman having the steepest fitted curve. However, it is still remarkable that the construction which is by far the most efficient on the GOAL test set produces so much worse results for the Michel automata than all the other constructions. Compared with the Rank construction, Piterman's complements of Michel 2 to 4 are 38.7%, 174.3%, and 573.6%, respectively, bigger. And this is exactly the Rank construction that performed so much worse on the GOAL test set that it made a comparison with the Piterman construction nearly impossible. Compared to the Fribourg construction, Piterman produces slightly smaller complements for Michel 1 and 2, but larger ones for Michel 3 and 4. Namely, they are 63.1% and 298.2% larger than the corresponding ones of the Fribourg construction.

Summarising we can say that the ranking of the constructions for the Michel test set is exactly the reverse of the ranking for the GOAL test set. The by far worst construction for the GOAL test set (Rank) is the best one for the Michel test set, and the by far best construction for the GOAL test set (Piterman) is the worst one for the Michel test set (at least for Michel 4). For the Fribourg construction this means that it “advances” from rank 3 of 4 for the GOAL test set to rank 2 of 4 for the Michel test set.

We also measured the execution times of the individual complementation tasks in CPU time seconds. The results are shown in Table 2.9.

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|----------|--------------|------------|
| Piterman+EQ+RO | 2.5 | 3.8 | 42.6 | 75,917.4 | $(1.08n)^n$ | 0.64% |
| Slice+P+RO+MADJ+EG | 2.3 | 3.6 | 11.4 | 159.5 | $(0.39n)^n$ | 0.38% |
| Rank+TR+RO | 2.2 | 3.0 | 6.4 | 30.0 | $(0.29n)^n$ | 0.18% |
| Fribourg+M1+M2+R2C | 2.5 | 3.5 | 10.8 | 2,332.6 | $(0.61n)^n$ | 0.62% |

Table 2.9: Execution times for the first four Michel automata.

Most interesting in this table is the column with the times for Michel 4. The time difference between the best and the worst construction is enormous. While the Rank construction took just 30 seconds to complement Michel 4, the Piterman construction took 75,917.4 seconds which is approximately 21 hours. This is more than 2500 times longer than the Rank construction. Of course the Piterman construction produced a bigger automata, which naturally requires more time, however, the automaton produced by the Piterman construction is just around 6.7 times bigger than the one of the Rank construction. This means that the Piterman construction must include very inefficient processes before finally arriving at the output automaton.

Furthermore, we can see in Table 2.9 that also the Fribourg construction took relatively long to complement Michel 4 compared to Rank, namely 2,332.6 seconds which are approximately 39 minutes. This is 77.8 times longer than the 30 seconds of Rank. At the same time, Fribourg's complement has just 68.2% more states than Rank's complement. Also compared to the Slice construction the Fribourg construction is slow for Michel 4. Slice's complement is 2.8 times bigger than Fribourg's complement, but with 159.5 seconds the complementation of slice was 14.6 times faster than the complementation of Fribourg.

So there seems to be an inefficiency in the Fribourg construction in terms of execution time for the complementation of Michel 4. However, this inefficiency is by far not as pronounced as for Piterman. While the complement of Piterman is just 4 times bigger, the execution time of Piterman is 32.5 times longer than the one of the Fribourg construction. One could also look at it from the other side and say that not the Fribourg construction is inefficient on Michel 4, but that Rank and Slice are extraordinarily efficient on this automaton.

Finally, these interesting differences in the execution time can only be observed for Michel 4. For Michel 3 there are also differences but they are by far not as pronounced as for Michel 4. If the computational resources would only allow it, it would be very interesting to see how the behaviour is for Michel 5 and beyond. One thing that stays the same for all the four Michel automata is that Rank is always the fastest and Piterman always the slowest construction.

2.3 Discussion

2.3.1 Summary of the Results

2.3.2 Validity

Appendix A

Plugin Installation and Usage

Since between the 2014-08-08 and 2014-11-17 releases of GOAL certain parts of the plugin interfaces have changed, and we adapted our plugin accordingly, the currently maintained version of the plugin works only with GOAL versions 2014-11-17 or newer. It is thus essential for any GOAL user to update to this version in order to use our plugin.

Appendix B

Median Complement Sizes of the GOAL Test Set

Bla bla bla

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|
| 1.0 | 130 | 117 | 109 | 77 | 69 | 61 | 56 | 40 | 40 | 29 | 1.0 | 171 | 174 | 166 | 124 | 118 | 117 | 100 | 67 | 84 | 35 |
| 1.2 | 387 | 456 | 352 | 281 | 155 | 136 | 101 | 105 | 75 | 45 | 1.2 | 622 | 833 | 803 | 877 | 529 | 398 | 320 | 372 | 215 | 53 |
| 1.4 | 822 | 683 | 394 | 376 | 230 | 204 | 151 | 120 | 105 | 63 | 1.4 | 2,086 | 1,618 | 1,367 | 1,676 | 1,065 | 967 | 664 | 682 | 494 | 78 |
| 1.6 | 890 | 594 | 458 | 321 | 237 | 178 | 134 | 114 | 113 | 61 | 1.6 | 2,465 | 2,073 | 2,182 | 1,959 | 1,518 | 1,259 | 767 | 545 | 623 | 78 |
| 1.8 | 624 | 507 | 324 | 275 | 196 | 136 | 110 | 92 | 89 | 41 | 1.8 | 2,310 | 1,963 | 1,950 | 1,988 | 1,485 | 1,095 | 746 | 418 | 346 | 57 |
| 2.0 | 362 | 286 | 211 | 176 | 117 | 103 | 79 | 64 | 59 | 34 | 2.0 | 1,318 | 1,482 | 1,393 | 1,461 | 981 | 871 | 434 | 338 | 228 | 50 |
| 2.2 | 248 | 222 | 124 | 116 | 82 | 73 | 56 | 52 | 50 | 28 | 2.2 | 1,068 | 1,145 | 1,085 | 1,067 | 772 | 747 | 263 | 235 | 158 | 40 |
| 2.4 | 147 | 145 | 114 | 87 | 56 | 48 | 43 | 39 | 35 | 19 | 2.4 | 689 | 838 | 809 | 751 | 524 | 466 | 240 | 159 | 93 | 30 |
| 2.6 | 115 | 117 | 67 | 61 | 47 | 42 | 32 | 29 | 29 | 15 | 2.6 | 469 | 531 | 555 | 565 | 437 | 360 | 169 | 94 | 71 | 23 |
| 2.8 | 95 | 71 | 52 | 45 | 38 | 29 | 27 | 25 | 23 | 13 | 2.8 | 369 | 421 | 536 | 405 | 329 | 224 | 130 | 81 | 58 | 21 |
| 3.0 | 59 | 60 | 47 | 35 | 32 | 27 | 22 | 21 | 20 | 10 | 3.0 | 244 | 327 | 360 | 322 | 219 | 176 | 85 | 64 | 49 | 16 |

(a) Piterman+EQ+RO

(b) Slice+P+RO+MADJ+EG

Figure B.1: Median complement sizes of the 10,998 effective samples of the external tests without the Rank construction. The rows (1.0 to 3.0) are the transition densities, and the columns (0.1 to 1.0) are the acceptance densities.

Appendix B. Median Complement Sizes of the GOAL Test Set

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 1.0 | 269 | 308 | 254 | 236 | 238 | 297 | 266 | 156 | 207 | 68 |
| 1.2 | 960 | 1,407 | 1,479 | 2,150 | 1,152 | 1,090 | 942 | 1,206 | 718 | 104 |
| 1.4 | 3,426 | 2,915 | 2,752 | 3,393 | 2,693 | 3,265 | 2,263 | 2,425 | 1,844 | 154 |
| 1.6 | 3,799 | 3,698 | 4,901 | 3,926 | 3,960 | 3,655 | 2,580 | 1,905 | 2,124 | 155 |
| 1.8 | 3,375 | 3,169 | 3,420 | 3,967 | 3,943 | 3,132 | 2,246 | 1,144 | 971 | 114 |
| 2.0 | 1,906 | 2,261 | 2,383 | 2,884 | 2,354 | 2,096 | 1,169 | 932 | 568 | 98 |
| 2.2 | 1,467 | 1,633 | 1,795 | 1,942 | 1,611 | 1,640 | 569 | 499 | 330 | 78 |
| 2.4 | 924 | 1,232 | 1,319 | 1,317 | 1,056 | 886 | 514 | 314 | 182 | 59 |
| 2.6 | 625 | 763 | 880 | 945 | 828 | 684 | 316 | 175 | 132 | 44 |
| 2.8 | 483 | 584 | 836 | 690 | 575 | 395 | 240 | 151 | 103 | 41 |
| 3.0 | 319 | 450 | 557 | 523 | 367 | 313 | 155 | 116 | 84 | 32 |

(a) Fribourg

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 1.0 | 390 | 438 | 434 | 324 | 328 | 459 | 337 | 204 | 227 | 40 |
| 1.2 | 1,576 | 2,394 | 2,505 | 2,996 | 1,613 | 1,551 | 1,166 | 1,542 | 1,002 | 58 |
| 1.4 | 5,007 | 4,336 | 4,652 | 4,877 | 3,458 | 3,956 | 3,169 | 3,380 | 1,868 | 86 |
| 1.6 | 5,067 | 5,032 | 6,444 | 4,868 | 4,575 | 3,864 | 3,211 | 1,731 | 1,892 | 85 |
| 1.8 | 4,016 | 3,701 | 3,647 | 4,523 | 3,548 | 3,009 | 1,808 | 451 | 336 | 62 |
| 2.0 | 1,663 | 2,276 | 2,676 | 3,035 | 1,925 | 1,932 | 464 | 307 | 150 | 54 |
| 2.2 | 989 | 1,514 | 1,621 | 1,826 | 1,121 | 846 | 155 | 127 | 93 | 45 |
| 2.4 | 560 | 821 | 919 | 771 | 529 | 267 | 133 | 87 | 55 | 32 |
| 2.6 | 388 | 519 | 524 | 441 | 259 | 219 | 84 | 50 | 41 | 26 |
| 2.8 | 311 | 317 | 396 | 242 | 165 | 95 | 64 | 44 | 33 | 22 |
| 3.0 | 173 | 224 | 211 | 169 | 102 | 72 | 41 | 34 | 27 | 18 |

(b) Fribourg+R2C

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-----|-----|-----|
| 1.0 | 225 | 223 | 195 | 181 | 187 | 199 | 189 | 124 | 161 | 68 |
| 1.2 | 731 | 971 | 946 | 1,071 | 629 | 562 | 488 | 568 | 388 | 104 |
| 1.4 | 2,228 | 1,701 | 1,543 | 1,732 | 1,241 | 1,287 | 945 | 944 | 727 | 154 |
| 1.6 | 2,489 | 2,263 | 2,331 | 2,133 | 1,777 | 1,443 | 964 | 757 | 889 | 155 |
| 1.8 | 2,381 | 2,027 | 2,009 | 2,075 | 1,618 | 1,243 | 1,005 | 592 | 515 | 114 |
| 2.0 | 1,390 | 1,569 | 1,416 | 1,573 | 1,093 | 1,008 | 594 | 464 | 330 | 98 |
| 2.2 | 1,118 | 1,197 | 1,150 | 1,151 | 879 | 809 | 317 | 330 | 241 | 78 |
| 2.4 | 712 | 885 | 836 | 809 | 580 | 535 | 316 | 231 | 145 | 59 |
| 2.6 | 498 | 569 | 601 | 627 | 497 | 412 | 217 | 137 | 113 | 44 |
| 2.8 | 391 | 455 | 578 | 456 | 374 | 263 | 173 | 119 | 90 | 41 |
| 3.0 | 258 | 350 | 392 | 354 | 253 | 208 | 119 | 97 | 74 | 32 |

(c) Fribourg+R2C+C

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-----|-----|
| 1.0 | 215 | 213 | 189 | 174 | 175 | 192 | 186 | 121 | 156 | 68 |
| 1.2 | 712 | 914 | 913 | 1,075 | 619 | 563 | 526 | 620 | 416 | 104 |
| 1.4 | 2,075 | 1,620 | 1,503 | 1,650 | 1,254 | 1,339 | 1,003 | 1,006 | 848 | 154 |
| 1.6 | 2,344 | 2,062 | 2,340 | 2,016 | 1,755 | 1,520 | 1,053 | 858 | 986 | 155 |
| 1.8 | 2,205 | 1,873 | 1,920 | 2,040 | 1,689 | 1,315 | 1,080 | 664 | 598 | 114 |
| 2.0 | 1,290 | 1,485 | 1,405 | 1,522 | 1,134 | 1,044 | 652 | 531 | 392 | 98 |
| 2.2 | 1,023 | 1,119 | 1,092 | 1,127 | 868 | 875 | 376 | 359 | 262 | 78 |
| 2.4 | 674 | 849 | 790 | 807 | 617 | 544 | 355 | 251 | 156 | 59 |
| 2.6 | 478 | 549 | 594 | 597 | 510 | 431 | 231 | 147 | 116 | 44 |
| 2.8 | 370 | 439 | 559 | 455 | 382 | 283 | 182 | 124 | 93 | 41 |
| 3.0 | 249 | 341 | 388 | 348 | 260 | 225 | 123 | 101 | 77 | 32 |

(d) Fribourg+M1

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-----|-----|-----|
| 1.0 | 225 | 223 | 195 | 181 | 187 | 199 | 189 | 124 | 161 | 68 |
| 1.2 | 731 | 971 | 946 | 1,071 | 629 | 562 | 488 | 568 | 388 | 104 |
| 1.4 | 2,228 | 1,701 | 1,543 | 1,732 | 1,241 | 1,287 | 945 | 944 | 727 | 154 |
| 1.6 | 2,489 | 2,263 | 2,331 | 2,133 | 1,777 | 1,443 | 964 | 757 | 889 | 155 |
| 1.8 | 2,381 | 2,027 | 2,009 | 2,075 | 1,618 | 1,215 | 1,005 | 592 | 515 | 114 |
| 2.0 | 1,390 | 1,513 | 1,416 | 1,542 | 1,093 | 1,003 | 594 | 441 | 330 | 97 |
| 2.2 | 1,019 | 1,156 | 1,064 | 1,104 | 859 | 785 | 304 | 303 | 221 | 78 |
| 2.4 | 672 | 867 | 789 | 772 | 544 | 478 | 269 | 191 | 139 | 55 |
| 2.6 | 466 | 542 | 572 | 568 | 452 | 348 | 183 | 129 | 99 | 43 |
| 2.8 | 368 | 407 | 480 | 337 | 260 | 197 | 129 | 96 | 75 | 36 |
| 3.0 | 201 | 261 | 266 | 272 | 199 | 136 | 83 | 74 | 50 | 27 |

(e) Fribourg+M1+M2

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-----|-----|
| 1.0 | 329 | 303 | 279 | 240 | 229 | 288 | 230 | 157 | 160 | 40 |
| 1.2 | 988 | 1,392 | 1,356 | 1,352 | 751 | 741 | 608 | 704 | 516 | 58 |
| 1.4 | 2,939 | 2,581 | 2,066 | 2,190 | 1,351 | 1,622 | 1,132 | 1,261 | 932 | 86 |
| 1.6 | 3,150 | 2,900 | 2,842 | 2,218 | 1,885 | 1,563 | 1,177 | 821 | 896 | 85 |
| 1.8 | 2,782 | 2,485 | 2,047 | 2,180 | 1,625 | 1,269 | 855 | 395 | 309 | 62 |
| 2.0 | 1,338 | 1,638 | 1,544 | 1,566 | 979 | 957 | 349 | 261 | 147 | 54 |
| 2.2 | 838 | 1,125 | 993 | 1,027 | 667 | 521 | 153 | 125 | 93 | 45 |
| 2.4 | 494 | 700 | 624 | 524 | 296 | 214 | 126 | 87 | 55 | 32 |
| 2.6 | 327 | 434 | 383 | 334 | 212 | 163 | 82 | 50 | 41 | 26 |
| 2.8 | 283 | 273 | 305 | 202 | 144 | 95 | 60 | 44 | 33 | 22 |
| 3.0 | 164 | 200 | 173 | 142 | 92 | 72 | 41 | 34 | 27 | 18 |

(f) Fribourg+M1+R2C

| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|-----|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.0 | 126 | 118 | 97 | 60 | 51 | 52 | 62 | 36 | 48 | 30 |
| 1.2 | 432 | 517 | 345 | 262 | 160 | 126 | 92 | 120 | 109 | 40 |
| 1.4 | 1,044 | 331 | 133 | 89 | 45 | 22 | 19 | 31 | 27 | 20 |
| 1.6 | 358 | 24 | 11 | 5 | 4 | 6 | 5 | 3 | 3 | 4 |
| 1.8 | 19 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2.0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2.2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2.4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2.6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2.8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3.0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

(g) Fribourg+M1+R2C+C

(h) Fribourg+R

Figure B.2: Median complement sizes of the 10,939 effective samples of the internal tests on the GOAL test set. The rows (1.0 to 3.0) are the transition densities, and the columns (0.1 to 1.0) are the acceptance densities.

Appendix C

Execution Times

| Construction | Mean | Min. | P25 | Median | P75 | Max. | Total | ≈ hours |
|-------------------|------|------|-----|--------|-----|-------|----------|---------|
| Fribourg | 8.5 | 2.5 | 3.3 | 4.9 | 7.3 | 586.0 | 93,351.2 | 259 |
| Fribourg+R2C | 6.6 | 2.2 | 2.9 | 4.2 | 6.4 | 219.7 | 72,545.7 | 202 |
| Fribourg+R2C+C | 8.5 | 2.2 | 2.6 | 3.5 | 6.4 | 582.9 | 93,396.2 | 259 |
| Fribourg+M1 | 4.9 | 2.5 | 3.2 | 4.1 | 5.9 | 55.1 | 54,061.3 | 150 |
| Fribourg+M1+M2 | 4.6 | 2.2 | 2.9 | 3.8 | 5.1 | 38.4 | 49,848.0 | 138 |
| Fribourg+M1+R2C | 4.4 | 2.2 | 2.8 | 3.6 | 5.3 | 42.5 | 48,572.0 | 135 |
| Fribourg+M1+R2C+C | 5.6 | 2.5 | 3.2 | 4.0 | 6.5 | 147.4 | 60,918.9 | 169 |
| Fribourg+R | 7.5 | 2.2 | 3.0 | 3.9 | 6.3 | 470.5 | 82,387.3 | 229 |

Table C.1: Execution times in CPU time seconds for the 10,939 effective samples of the GOAL test set.

| Construction | Mean | Min. | P25 | Median | P75 | Max. | Total | ≈ hours |
|--------------------|------|------|-----|--------|-----|-------|-----------|---------|
| Piterman+EQ+RO | 3.0 | 2.2 | 2.6 | 2.8 | 3.0 | 42.9 | 21,410.6 | 59 |
| Slice+P+RO+MADJ+EG | 3.7 | 2.2 | 2.7 | 3.2 | 4.1 | 36.7 | 26,398.9 | 73 |
| Rank+TR+RO | 16.0 | 2.3 | 2.8 | 3.7 | 9.3 | 443.3 | 115,563.9 | 321 |
| Fribourg+M1+R2C | 4.0 | 2.2 | 2.7 | 3.1 | 4.4 | 410.4 | 28,970.8 | 80 |

Table C.2: Execution times in CPU time seconds for the 7,204 effective samples of the GOAL test set.

| Construction | Mean | Min. | P25 | Median | P75 | Max. | Total | ≈ hours |
|--------------------|------|------|-----|--------|-----|-------|----------|---------|
| Piterman+EQ+RO | 3.6 | 2.2 | 2.7 | 2.9 | 3.4 | 365.7 | 39,663.4 | 110 |
| Slice+P+RO+MADJ+EG | 4.3 | 2.2 | 2.9 | 3.7 | 5.0 | 42.4 | 47,418.2 | 132 |
| Fribourg+M1+R2C | 4.7 | 2.2 | 2.8 | 3.6 | 5.3 | 410.4 | 52,149.0 | 145 |

Table C.3: Execution times in CPU time seconds for the 10,998 effective samples of the GOAL test set without the Rank construction.

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|-----------|--------------|------------|
| Fribourg | 2.3 | 4.0 | 88.8 | 100,976.0 | $(1.14n)^n$ | 0.64% |
| Fribourg+R2C | 2.3 | 3.4 | 27.4 | 27,938.3 | $(0.92n)^n$ | 0.64% |
| Fribourg+M1 | 2.2 | 3.6 | 17.9 | 6,508.4 | $(0.72n)^n$ | 0.63% |
| Fribourg+M1+M2 | 2.3 | 3.5 | 13.8 | 2,707.4 | $(0.62n)^n$ | 0.62% |
| Fribourg+M1+M2+R2C | 2.5 | 3.5 | 10.8 | 2,332.6 | $(0.61n)^n$ | 0.62% |
| Fribourg+R | 2.4 | 3.7 | 86.0 | 101,809.6 | $(1.14n)^n$ | 0.64% |

Table C.4: Execution times in CPU time seconds for the four Michel automata.

| Construction | Michel 1 | Michel 2 | Michel 3 | Michel 4 | Fitted curve | Std. error |
|--------------------|----------|----------|----------|----------|--------------|------------|
| Piterman+EQ+RO | 2.5 | 3.8 | 42.6 | 75,917.4 | $(1.08n)^n$ | 0.64% |
| Slice+P+RO+MADJ+EG | 2.3 | 3.6 | 11.4 | 159.5 | $(0.39n)^n$ | 0.38% |
| Rank+TR+RO | 2.2 | 3.0 | 6.4 | 30.0 | $(0.29n)^n$ | 0.18% |
| Fribourg+M1+M2+R2C | 2.5 | 3.5 | 10.8 | 2,332.6 | $(0.61n)^n$ | 0.62% |

Table C.5: Execution times in CPU time seconds for the four Michel automata.

Bibliography

- [1] C. Althoff, W. Thomas, N. Wallmeier. Observations on Determinization of Büchi Automata. In J. Farré, I. Litovsky, S. Schmitz, eds., *Implementation and Application of Automata*. vol. 3845 of *Lecture Notes in Computer Science*. pp. 262–272. Springer Berlin Heidelberg. 2006.
- [2] C. Göttel. Implementation of an Algorithm for Büchi Complementation. BSc Thesis, University of Fribourg, Switzerland. November 2013.
- [3] D. Kähler, T. Wilke. Complementation, Disambiguation, and Determinization of Büchi Automata Unified. In L. Aceto, I. Damgård, L. Goldberg, et al, eds., *Automata, Languages and Programming*. vol. 5125 of *Lecture Notes in Computer Science*. pp. 724–735. Springer Berlin Heidelberg. 2008.
- [4] O. Kupferman, M. Y. Vardi. Weak Alternating Automata Are Not that Weak. *ACM Trans. Comput. Logic*. 2(3):pp. 408–429. Jul. 2001.
- [5] M. Michel. Complementation is more difficult with automata on infinite words. *CNET, Paris*. 15. 1988.
- [6] D. E. Muller, P. E. Schupp. Simulating Alternating Tree Automata by Nondeterministic Automata: New Results and New Proofs of the Theorems of Rabin, McNaughton and Safra. *Theoretical Computer Science*. 141(1–2):pp. 69 – 107. 1995.
- [7] N. Piterman. From Nondeterministic Buchi and Streett Automata to Deterministic Parity Automata. *Logical Methods in Computer Science*. 3(5):pp. 1–21. 2007.
- [8] S. Safra. On the Complexity of Omega-Automata. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*. pp. 319–327. Oct 1988.
- [9] S. Schewe. Büchi Complementation Made Tight. In *26th International Symposium on Theoretical Aspects of Computer Science-STACS 2009*. pp. 661–672. 2009.
- [10] A. P. Sistla, M. Y. Vardi, P. Wolper. The Complementation Problem for Büchi Automata with Applications to Temporal Logic. *Theoretical Computer Science*. 49(2–3):pp. 217 – 237. 1987.
- [11] W. Thomas. Languages, Automata, and Logic. In G. Rozenberg, A. Salomaa, eds., *Handbook of Formal Languages*. pp. 389–455. Springer Berlin Heidelberg. 1997.
- [12] W. Thomas. Complementation of Büchi Automata Revisited. In J. Karhumäki, H. Maurer, G. Păun, et al, eds., *Jewels are Forever*. pp. 109–120. Springer Berlin Heidelberg. 1999.
- [13] M.-H. Tsai, S. Fogarty, M. Vardi, et al. State of Büchi Complementation. In M. Domaratzki, K. Salomaa, eds., *Implementation and Application of Automata*. vol. 6482 of *Lecture Notes in Computer Science*. pp. 261–271. Springer Berlin Heidelberg. 2011.
- [14] M.-H. Tsai, Y.-K. Tsay, Y.-S. Hwang. GOAL for Games, Omega-Automata, and Logics. In N. Sharygina, H. Veith, eds., *Computer Aided Verification*. vol. 8044 of *Lecture Notes in Computer Science*. pp. 883–889. Springer Berlin Heidelberg. 2013.
- [15] Y.-K. Tsay, Y.-F. Chen, M.-H. Tsai, et al. Goal: A Graphical Tool for Manipulating Büchi Automata and Temporal Formulae. In O. Grumberg, M. Huth, eds., *Tools and Algorithms for the Construction and Analysis of Systems*. vol. 4424 of *Lecture Notes in Computer Science*. pp. 466–471. Springer Berlin Heidelberg. 2007.

- [16] Y.-K. Tsay, Y.-F. Chen, M.-H. Tsai, et al. Goal Extended: Towards a Research Tool for Omega Automata and Temporal Logic. In C. Ramakrishnan, J. Rehof, eds., *Tools and Algorithms for the Construction and Analysis of Systems*. vol. 4963 of *Lecture Notes in Computer Science*. pp. 346–350. Springer Berlin Heidelberg. 2008.
- [17] Y.-K. Tsay, Y.-F. Chen, M.-H. Tsai, et al. Tool support for learning Büchi automata and linear temporal logic. *Formal Aspects of Computing*. 21(3):pp. 259–275. 2009.
- [18] M. Y. Vardi, T. Wilke. Automata: From Logics to Algorithms. In J. Flum, E. Grädel, T. Wilke, eds., *Logic and Automata: History and Perspectives*. vol. 2 of *Texts in Logic and Games*. pp. 629–736. Amsterdam University Press. 2007.