A Complete Proof System for 1-Free Regular Expressions Modulo Bisimilarity

Clemens Grabmayer
Department of Computer Science
Gran Sasso Science Institute
L'Aquila, Italy
clemens.grabmayer@gssi.it

Wan Fokkink
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
w.j.fokkink@vu.nl

Abstract

Robin Milner (1984) gave a sound proof system for bisimilarity of regular expressions interpreted as processes: Basic Process Algebra with unary Kleene star iteration, deadlock 0, successful termination 1, and a fixed-point rule. He asked whether this system is complete. Despite intensive research over the last 35 years, the problem is still open.

This paper gives a partial positive answer to Milner's problem. We prove that the adaptation of Milner's system over the subclass of regular expressions that arises by dropping the constant 1, and by changing to binary Kleene star iteration is complete. The crucial tool we use is a graph structure property that guarantees expressibility of a process graph by a regular expression, and that is preserved when going over from a process graph to its bisimulation collapse.

CCS Concepts: • **Theory of computation → Process calculi**; **Equational logic and rewriting**.

Keywords: regular expressions, process algebra, bisimilarity, process graphs, complete proof system

ACM Reference Format:
Clemens Grabmayer and Wan Fokkink. 2020. A Complete Proof System for 1-Free Regular Expressions Modulo Bisimilarity. In *Proceedings of the 35th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS '20), July 8–11, 2020, Saarbrücken, Germany*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3373718.3394744

1 Introduction

Regular expressions, introduced by Kleene [17], are widely studied in formal language theory, notably for string searching [29]. They are constructed from constants 0 (no strings), 1 (the empty string), and a (a single letter) from some alphabet; binary operators $+$ and $\cdot$ (union and concatenation); and the unary Kleene star * (zero or more iterations).

Their interpretations are Kleene algebras with as prime example the algebra of regular events, the language semantics of regular expressions, which is closely linked with finite-state automata. Aanderaa [1] and Salomaa [24] gave complete axiomatizations for the language semantics of regular expressions, with a non-algebraic fixed-point rule that has a non-empty-word property as side condition. Krob [20] found an infinitary algebraic axiomatization with equational implications, and Kozen [18] presented a finitary algebraic system.

Regular expressions also received significant attention in the process algebra community [5], where they are interpreted modulo the bisimulation process semantics [22]. Robin Milner [21] was the first to study regular expressions in this setting, where he called them star expressions. Here the interpretation of 0 is deadlock, 1 is (successful) termination, a is an atomic action, and $+$ and $\cdot$ are alternative and sequential composition of two processes, respectively. Milner adapted Salomaa's axiomatization to obtain a sound proof system for this setting, and posed the (still open) question whether this axiomatization is complete, meaning that if the process graphs of two star expressions are bisimilar, then they can be proven equal.

Milner's axiomatization contains a fixed-point rule, which is inevitable because due to the presence of 0 the underlying equational theory is not finitely based [25, 26]. Bergstra, Bethke, and Ponse [4] studied star expressions without 0 and 1, replaced the unary by the binary Kleene star $^{\circledast}$, which represents an iteration of the first argument, possibly eventually followed by the execution of the second argument. They obtained an axiomatization by basically omitting the axioms for 0 and 1 as well as the fixed-point rule from Milner's axiomatization, and adding Troeger's axiom [30]. This purely equational axiomatization was proven complete in [9, 11]. A sound and complete axiomatization for star expressions without unary Kleene star, but with 0 and 1 and a unary perpetual loop operator ^{*0} (equivalently, unary star is restricted to terms $e^* \cdot 0$), was given in [8, 10].

In contrast to the formal languages setting, not all finite-state process graphs can be expressed by a star expression modulo bisimilarity. Milner posed a second question in [21],

LICS '20, July 8–11, 2020, Saarbrücken, Germany

ACM ISBN 978-1-4503-7104-9/20/07…$15.00
https://doi.org/10.1145/3373718.3394744

namely, to characterize which finite-state process graphs can be expressed. This was shown to be decidable in [3] by defining and using 'well-behaved' specifications.

In this paper we prove completeness of Milner's axiomatization (tailored to the adapted setting) for star expressions with 0, but without 1 and with the binary Kleene star (note that with the unary star, 1 can be constructed as 0^*). This is substantially more difficult than in the absence of 0. Notably, the normalization approach of [11] cannot be used. In particular, Lemma 6 in [11], that $e \cdot g \leftrightarrows f \cdot g$ implies $e \leftrightarrows f$, where $\leftrightarrows$ means bisimilarity, fails with 0 for g, $a^{\circledast}a$ for e, and $(a \cdot (a + a \cdot 0))^{\circledast}a$ for f (counterexample adapted from [7]). Our result crucially extends the result proved in [10] for 0 and the perpetual loop operator ^{*0} by permitting iteration expressions with non-deadlocking exits (the extension [8] of [10] including 1 is not covered directly).

While earlier completeness proofs focus on the manipulation of terms, we follow in Milner's footsteps and focus on their process graphs. A key idea is to determine loops in graphs associated with star expressions. By a loop we mean a subgraph generated by a set of entry transitions from a vertex v in which (1) there is an infinite path from v, (2) each infinite path eventually returns to v, and (3) termination is not permitted. A graph is said to satisfy LLEE (Layered Loop Existence and Elimination) if repeatedly eliminating the entry transitions of a loop, and performing garbage collection, leads to a graph without infinite paths. LLEE offers a generalization (and a more elegant definition) of the notion of a well-behaved specification.

Our completeness proof roughly works as follows (for more details see Sect. 4). Let e_1 and e_2 be star expressions with bisimilar process graph interpretations G_1 and G_2. We show that G_1 and G_2 satisfy LLEE. We moreover prove that LLEE is preserved under bisimulation collapse. And we construct for each graph that satisfies LLEE a star expression that corresponds to this graph, modulo bisimilarity. In particular such a star expression f can be constructed for the bisimulation collapse C of G_1 and G_2. We show that both e_1 and e_2 can be proven equal to f, by a pull-back of f over the functional bisimulations from G_1 and G_2 to the bisimulation collapse C. This yields the desired completeness result.

In our proof, the minimization of terms (and thereby of the associated process graphs) in the left-hand side of a binary Kleene star modulo bisimilarity is partly inspired by [8, 10]. Interestingly, we will be able to use as running example the process graph interpretation of the star expression that at the end of [10] is mentioned as problematic for a completeness proof for our current setting. Our crucial use of witnesses for the graph property LLEE borrows from the representation of cyclic λ-terms [15] as structure-constrained term graphs, as used for defining and implementing maximal sharing in the λ-calculus with `letrec` [16] (see also [13]).

The completeness result for star expressions with 0 but without 1 and with the binary Kleene star settles a natural question. We are also hopeful that the property LLEE provides a strong conceptual tool for approaching Milner's long-standing open question regarding the class of all star expressions. The presence of 1-transitions in graphs presents new challenges, such as that LLEE is not always preserved under bisimulation collapse. In order to be able to still work with this concept, we will need workarounds.

Please see the extended version [14] for details of proofs that have been omitted or are only sketched.

2 Preliminaries

In this section we define star expressions, their process semantics as 'charts', the proof system BBP for bisimilarity of their chart interpretations, and provable solutions of charts.

Definition 2.1. Given a set A of *actions*, the set $StExp(A)$ of *star expressions over* A is generated by the grammar:

$$e ::= 0 \mid a \mid (e_1 + e_2) \mid (e_1 \cdot e_2) \mid (e_1{}^{\circledast}e_2) \quad (\text{with } a \in A).$$

0 represents deadlock (i.e., does not perform any action), a an atomic action, $+$ alternative and $\cdot$ sequential composition, and $^{\circledast}$ the binary Kleene star. Note that 1 (for empty steps) is missing from the syntax. $\sum_{i=1}^{k} e_i$ is defined recursively as 0 if $k = 0$, e_1 if $k = 1$, and $(\sum_{i=1}^{k-1} e_i) + e_k$ if $k > 1$.

The *star height* $|e|_{\circledast}$ of a star expression $e \in StExp(A)$ denotes the maximum number of nestings of Kleene stars in e: it is defined by $|0|_{\circledast} := |a|_{\circledast} := 0$, $|f + g|_{\circledast} := |f \cdot g|_{\circledast} := \max\{|f|_{\circledast}, |g|_{\circledast}\}$, and $|f^{\circledast}g|_{\circledast} := \max\{|f|_{\circledast} + 1, |g|_{\circledast}\}$.

Definition 2.2. By a (finite sink-termination) *chart* C we understand a 5-tuple $\langle V, \surd, v_s, A, T\rangle$ where V is a finite set of *vertices*, $\surd$ is, in case $\surd \in V$, a special vertex with no outgoing transitions (a sink) that indicates termination (in case $\surd \notin V$, the chart does not admit termination), $v_s \in V \backslash \{\surd\}$ is the *start vertex*, A is a set *actions*, and $T \subseteq V \times A \times V$ the set of *transitions*. Since A can be reconstructed from T, we will frequently keep A implicit, denote a chart as a 4-tuple $\langle V, \surd, v_s, T\rangle$. A chart is *start-vertex connected* if every vertex is reachable by a path from the start vertex. This property can be achieved by removing unreachable vertices ('garbage collection'). We will assume charts to be start-vertex connected.

In a chart C, let $v \in V$ and $U \subseteq T$ be a set of transitions from v. By the $\langle v, U\rangle$*-generated subchart of* C we mean the chart $C_0 = \langle V_0, \surd, v, A, T_0\rangle$ with start vertex v where V_0 is the set of vertices and T_0 the set of transitions that are on paths in C from v that first take a transition in U, and then, until v is reached again, continue with other transitions of C.
We use the standard notation $v \xrightarrow{a} v'$ in lieu of $\langle w, a, w'\rangle \in T$.

Definition 2.3. Let $C_i = \langle V_i, \surd, v_{s,i}, T_i\rangle$ for $i \in \{1, 2\}$ be two charts. A *bisimulation* between C_1 and C_2 is a relation $B \subseteq V_1 \times V_2$ that satisfies the following conditions:

(*start*) $v_{s,1}\, B\, v_{s,2}$ (it relates the start vertices), and for all $v_1, v_2 \in V$ with $v_1 \, B \, v_2$:

(*forth*) for every transition $v_1 \xrightarrow{a} v_1'$ in C_1 there is a transition $v_2 \xrightarrow{a} v_2'$ in C_2 with $v_1' \, B \, v_2'$,

(*back*) for every transition $v_2 \xrightarrow{a} v_2'$ in C_2 there is a transition $v_1 \xrightarrow{a} v_1'$ in C_1 with $v_1' \, B \, v_2'$,

(*termination*) $v_1 = \surd$ if and only if $v_2 = \surd$.

If there is a bisimulation between C_1 and C_2, then we write $C_1 \leftrightarrows C_2$ and say that C_1 and C_2 are *bisimilar*. If a bisimulation is the graph of a function, we say that it is a *functional* bisimulation. We write $C_1 \rightrightarrows C_2$ if there is a functional bisimulation between C_1 and C_2.

Definition 2.4. For every star expression $e \in \mathit{StExp}(A)$ the *chart interpretation* $C(e) = \langle V(e), \surd, e, A, T(e) \rangle$ *of* e is the chart with start vertex e that is specified by iteration via the following transition rules, which form a transition system specification (TSS), with $e, e_1, e_2, e_1' \in \mathit{StExp}(A)$, $a \in A$:

$$\frac{}{a \xrightarrow{a} \surd} \qquad \frac{e_i \xrightarrow{a} \xi}{e_1 + e_2 \xrightarrow{a} \xi} \; (i = 1, 2)$$

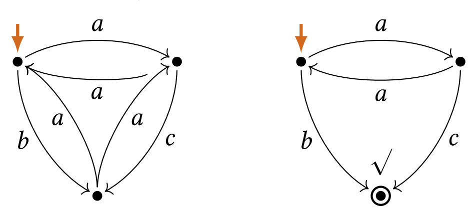
$$\frac{e_1 \xrightarrow{a} e_1'}{e_1 \cdot e_2 \xrightarrow{a} e_1' \cdot e_2} \qquad \frac{e_1 \xrightarrow{a} \surd}{e_1 \cdot e_2 \xrightarrow{a} e_2}$$

$$\frac{e_1 \xrightarrow{a} e_1'}{e_1^{\circledast} e_2 \xrightarrow{a} e_1' \cdot (e_1^{\circledast} e_2)} \qquad \frac{e_1 \xrightarrow{a} \surd}{e_1^{\circledast} e_2 \xrightarrow{a} e_1^{\circledast} e_2} \qquad \frac{e_2 \xrightarrow{a} \xi}{e_1^{\circledast} e_2 \xrightarrow{a} \xi}$$

with $\xi \in \mathit{StExp}(A)_{\surd} := \mathit{StExp}(A) \cup \{\surd\}$, where $\surd$ indicates termination. If $e \xrightarrow{a} \xi$ can be proved, ξ is called an *a-derivative*, or just *derivative*, of e. The set $V(e) \subseteq \mathit{StExp}(A)_{\surd}$ consists of the *iterated derivatives* of e. To see that $C(e)$ is finite, Antimirov's result [2], that a regular expression has only finitely many iterated derivatives, can be adapted.

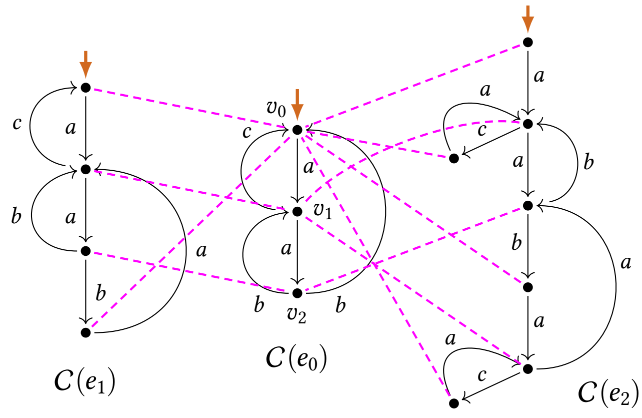
We say that a star expression $e \in \mathit{StExp}(A)$ is *normed* if there is a path of transitions from e to $\surd$ in $C(e)$.

Example 2.5. The left chart below does not admit termination. The right chart is a double-exit graph with the sink termination vertex $\surd$ at the bottom.



These charts are not bisimilar to chart interpretations of star expressions. For the left chart this was shown by Milner [21], and for the right chart by Bosscher [6].

Example 2.6. By the rules in Def. 2.4, $e_0 := a \cdot e_0'$ with $e_0' := (c \cdot a + a \cdot (b + b \cdot a))^{\circledast} 0$ has the chart $C(e_0)$ as below, with $v_0 := e_0$, $v_1 := e_0'$ and $v_2 := (b + b \cdot a) \cdot e_0'$. This chart is the bisimulation collapse of the charts $C(e_1)$ and $C(e_2)$ of star expressions $e_1 := (a \cdot ((a \cdot (b + b \cdot a))^{\circledast} c))^{\circledast} 0$, and $e_2 := a \cdot ((c \cdot a + a \cdot (b \cdot a \cdot ((c \cdot a)^{\circledast} a))^{\circledast} b)^{\circledast} 0)$. Bisimulations between $C(e_1)$ and $C(e_0)$, and between $C(e_0)$ and $C(e_1)$ are indicated by the broken lines. The chart $C(e_0)$ was considered problematic in [10].



Definition 2.7. The proof system BBP or the class of star expressions has the axioms (B1)–(B6), (BKS1), (BKS2), the inference rules of equational logic, and the rule RSP$^{\circledast}$:

$$\begin{array}{lrcl}
\text{(B1)} & x + y & = & y + x \\
\text{(B2)} & (x + y) + z & = & x + (y + z) \\
\text{(B3)} & x + x & = & x \\
\text{(B4)} & (x + y) \cdot z & = & x \cdot z + y \cdot z \\
\text{(B5)} & (x \cdot y) \cdot z & = & x \cdot (y \cdot z) \\
\text{(B6)} & x + 0 & = & x \\
\text{(B7)} & 0 \cdot x & = & 0 \\
\text{(BKS1)} & x \cdot (x^{\circledast} y) + y & = & x^{\circledast} y \\
\text{(BKS2)} & (x^{\circledast} y) \cdot z & = & x^{\circledast} (y \cdot z) \\
(\text{RSP}^{\circledast}) & & & \dfrac{x = (y \cdot x) + z}{x = y^{\circledast} z}
\end{array}$$

By $e_1 =_{\text{BBP}} e_2$ we denote that $e_1 = e_2$ is derivable in BBP.

BBP is a finite 'implicational' proof system [28], because unlike in Salomaa's and Milner's systems for regular expressions with 1 the fixed-point rule does not require any side-condition to ensure 'guardedness'.

Definition 2.8. For a chart $C = \langle V, \surd, v_s, A, T \rangle$, a *provable solution of* C is a function $s : V \backslash \{\surd\} \to \mathit{StExp}(A)$ such that:

$$s(v) =_{\text{BBP}} \Big(\sum_{i=1}^{m} a_i \Big) + \Big(\sum_{j=1}^{n} b_j \cdot s(w_j) \Big) \quad (\text{for all } v \in V \backslash \{\surd\})$$

holds, given that the union of $\{ v \xrightarrow{a_i} \surd \mid i = 1, \ldots, m \}$ and $\{ v \xrightarrow{b_j} w_j \mid j = 1, \ldots, n, \; w_j \neq \surd \}$ is the set of transitions from v in C. We call $s(v_s)$ the *principal value* of s.

Proposition 2.9 (uses BBP-axioms (B1)–(B7), (BKS1))**.** *For every $e \in \mathit{StExp}(A)$, the identity function $\mathrm{id}_{V(e)} : V(e) \to V(e) \subseteq \mathit{StExp}(A)$, $e' \mapsto e'$, is a provable solution of the chart interpretation $C(e)$ of e.*

Proof (Idea). Each e in $StExp(A)$ is the BBP-provable sum of expressions a and $a \cdot e'$ over all $a \in A$ for a-derivatives $\surd$ and e', respectively, of e. This 'fundamental theorem[1] of a differential calculus for star expressions' implies, quite directly, that $\mathrm{id}_{V(e)}$ is a provable solution of $C(e)$. ◻

3 Layered loop existence and elimination

As preparation for the definition of the central concept of 'LLEE-witness', we start with an informal explanation of the structural chart property 'LEE'. It is a necessary condition for a chart to be the chart interpretation of a star expression. LEE is defined by a dynamic elimination procedure that analyses the structure of the graph by peeling off 'loop subcharts'. Such subcharts capture, within the chart interpretation of a star expression e, the behaviour of the iteration of f_1 within innermost subterms $f_1{}^{\circledast}f_2$ in e. (A weaker form of 'loop' by Milner [21], which describes the behavior of general iteration subterms, is not sufficient for our aims.)

Definition 3.1. A chart $\mathcal{L} = \langle V, \surd, v_s, T \rangle$ is a *loop chart* if:

(L1) There is an infinite path from the start vertex v_s.

(L2) Every infinite path from v_s returns to v_s after a positive number of transitions (and so visits v_s infinitely often).

(L3) V does not contain the vertex $\surd$.

In such a loop chart we call the transitions from v_s *loop-entry transitions*, and all other transitions *loop-body transitions*.

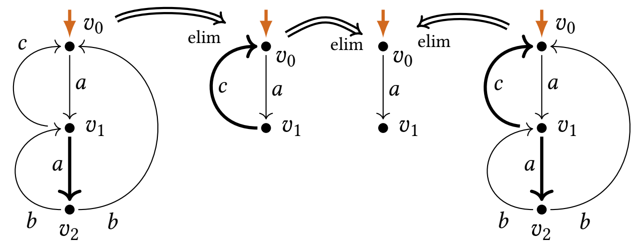
Let C be a chart. A loop chart $\mathcal{L}$ is called a *loop subchart of* C if $\mathcal{L}$ is the $\langle v, U \rangle$-generated subchart of C for some vertex v of C, and a set U of transitions of C that depart from v (so the transitions in U are the loop-entry transitions of $\mathcal{L}$).

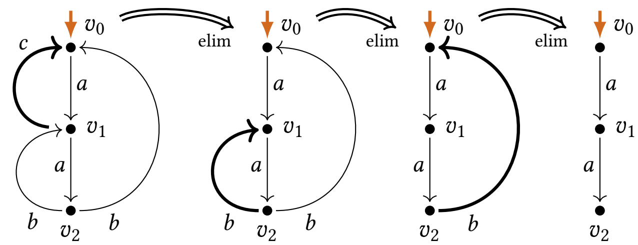
Note that the two charts in Ex. 2.5 are not loop charts: the left one violates (L2), and the right one violates (L3). Moreover, none of these charts contains a loop subchart. While the chart $C(e_0)$ in Ex. 2.6 is not a loop chart either, as it violates (L2), we will see that it has loop subcharts.

Let $\mathcal{L}$ be a loop subchart of a chart C. Then the result of *eliminating* $\mathcal{L}$ *from* C arises by removing all loop-entry transitions of $\mathcal{L}$ from C, and then removing all vertices and transitions that get unreachable. We say that a chart C has the *loop existence and elimination property (LEE)* if the process, started on C, of repeated eliminations of loop subcharts results in a chart that does not have an infinite path.

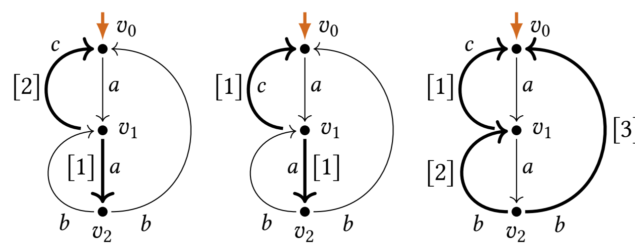
For the charts in Ex. 2.5 the procedure stops immediately, as they do not contain loop subcharts. Since both of them have infinite paths, it follows that they do not satisfy LEE.

We consider three runs of the elimination procedure for the chart $C(e_0)$ in Ex. 2.6. The loop-entry transitions of loop subcharts that are removed in each step are marked in bold. Each run witnesses that $C(e_0)$ satisfies LEE.





Note that loop elimination does not yield a unique result.[2] Runs can be recorded by attaching, in the original chart, to transitions that get removed in the elimination procedure as marking label the sequence number of the appertaining elimination step. For the three runs of loop elimination above we get the following marking labeled versions of C, respectively:



Since all three runs were successful (as they yield charts without infinite paths), these recordings (marking-labeled charts) can be viewed as 'LEE-witnesses'. We now will define a concept of a 'layered LEE-witness' (LLEE-witness), i.e., a LEE-witness with the added constraint that in the formulated run of the loop elimination procedure it never happens that a loop-entry transition is removed from within the body of a previously removed loop subchart. This refined concept has simpler properties, and it will fit our purpose.

Before introducing 'LLEE-witnesses', we first define chart labelings that mark transitions in a chart as '(loop-)entry' and as '(loop-)body' transitions, but without safeguarding that these markings refer to actual loops.

Definition 3.2. Let $C = \langle V, v_s, \surd, A, T \rangle$ be a chart. An *entry/body-labeling* $\hat{C} = \langle V, v_s, \surd, A \times \mathbb{N}, \widehat{T} \rangle$ of C is a chart that arises from C by adding, for each transition $\tau = \langle v_1, a, v_2 \rangle \in$

[1] Rutten [23] used this name for an analogous result on infinite streams. The first author [12], and Kozen and Silva [19, 27] used it for the provable synthesis of regular expressions from their Brzozowski derivatives. The result here can be viewed as stating the provable synthesis of regular expressions from their partial derivatives (due to Antimirov [2]).

[2] Confluence, and unique normalization, can be shown if a pruning operation is added that permits to drop transitions to deadlocking vertices.

T, to the action label a of τ a *marking label* $\alpha \in \mathbb{N}$, yielding $\widehat{\tau} = \langle v_1, \langle a, \alpha \rangle, v_2 \rangle \in \widehat{T}$. In such an entry/body-labeling we call transitions with marking label 0 *body transitions*, and transitions with marking labels in $\mathbb{N}^+$ *entry transitions*.

Let $\hat{C}$ be an entry/body-labeling of C, and let v and w be vertices of C and $\hat{C}$. We denote by $v \rightarrow_{\text{bo}} w$ that there is a body transition $v \xrightarrow{\langle a,0\rangle} w$ in $\hat{C}$ for some $a \in A$, and by $v \rightarrow_{[\alpha]} w$, for $\alpha \in \mathbb{N}^+$ that there is an entry transition $v \xrightarrow{\langle a,\alpha\rangle} w$ in $\hat{C}$ for some $a \in A$. We will use $\alpha, \beta, \gamma, \ldots$ for marking labels in $\mathbb{N}^+$ of entry transitions. By the set $E(\hat{C})$ of *entry transition identifiers* we denote the set of pairs $\langle v, \alpha \rangle \in V \times \mathbb{N}^+$ such that an entry transition $\rightarrow_{[\alpha]}$ departs from v in $\hat{C}$. For $\langle v, \alpha \rangle \in E(\hat{C})$, we define by $C_{\hat{C}}(v, \alpha)$ the subchart of C with start vertex v_s that consists of the vertices and transitions which occur on paths in C as follows: they start with a $\rightarrow_{[\alpha]}$ entry transition from v, continue with body transitions only, and halt immediately if v is revisited.

Definition 3.3. A *LLEE-witness* $\hat{C}$ of a chart C is an entry/body-labeling of C that satisfies the following properties:

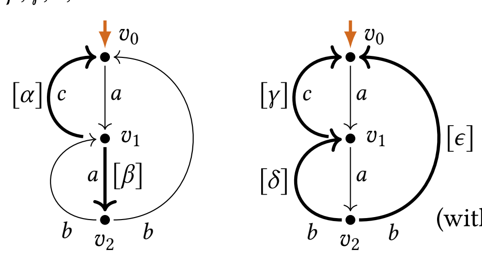
(W1) There is no infinite path of $\rightarrow_{\text{bo}}$ transitions from v_s.

(W2) For all $\langle v, \alpha \rangle \in E(\hat{C})$, (a) $C_{\hat{C}}(v, \alpha)$ is a loop chart, and (b) (*layeredness*) from no vertex $w \neq v$ of $C_{\hat{C}}(v, \alpha)$ there departs in $\hat{C}$ an entry transition $\rightarrow_{[\beta]}$ with $\beta \geqslant \alpha$.

The stipulation in (W2)(a) justifies to call entry transitions in a LLEE-witness a *loop-entry transition*. For a loop-entry transition $\rightarrow_{[\beta]}$ with $\beta \in \mathbb{N}^+$, we call β its *loop level*.

A chart is a *LLEE-chart* if it has a LLEE-witness.

Example 3.4. The three labelings of the chart $C(e_0)$ in Ex. 2.6 that arose as recordings of runs of the loop elimination procedure can be viewed as entry/body-labelings of that chart. There, and below, we dropped the body labels of transitions, and instead only indicated the entry labels in boldface together with their levels. By checking conditions (W1) and (W2),(a)-(b), it is easy to verify that these entry/body-labelings are LLEE-witnesses. In fact it is not difficult to establish that every LLEE-witness of $C(e_0)$ in Ex. 2.6 is of either of the following two forms, with marking labels $\alpha, \beta, \gamma, \delta, \epsilon \in \mathbb{N}^+$:



(with $\gamma < \delta, \epsilon$)

We now argue that LLEE-witnesses guarantee the property LEE. Let $\hat{C}$ be a LLEE-witness of a chart C. Repeatedly pick an entry transition identifier $\langle v, \alpha \rangle$ with $\alpha \in \mathbb{N}^+$ minimal, remove the loop subchart that is generated by loop-entry transitions of level α from v (it is indeed a loop by (W2)(a), and minimality of α and (W2)(b) ensure the absence of departing loop-entry transitions of lower level), and perform garbage collection. Eventually the part of C that is reachable by body transitions from the start vertex is obtained. This subchart does not have an infinite path due to (W1). Therefore C indeed satisfies LEE, as witnessed by $\hat{C}$.

The property LEE and the concept of LLEE-witness are closely linked with the process semantics of star expressions. In fact, we now define a labeling of the TSS in Def. 2.4 that permits to define, for every star expression e, an entry/body-labeling of the chart interpretation $C(e)$ of e, which can then be recognized as a LLEE-witness of $C(e)$.

We refine the TSS rules in Def. 2.4 as follows: A body label is added to transitions that cannot return to the star expression in their left-hand side. The rule for transitions into the iteration part e_1 of an iteration $e_1{}^{\circledast}e_2$ is split into the cases where e_1 is normed or not. Only in the normed case can $e_1{}^{\circledast}e_2$ return to itself, and then a loop-entry transition with the star height $|e_1|_{\circledast}$ of e_1 plus 1 as its level is created.

Definition 3.5. For every $e \in StExp(A)$, we define *the entry/body-labeling* $\widehat{C(e)}$ *of* the chart interpretation $C(e)$ of e in analogy with $C(e)$ by using the following transition rules that refine the rules in Def. 2.4 by adding marking labels:

$$\frac{}{a \xrightarrow{a}_{\text{bo}} \surd} \qquad \frac{e_i \xrightarrow{a}_l \xi}{e_1 + e_2 \xrightarrow{a}_{\text{bo}} \xi}\; i \in \{1,2\}$$

$$\frac{e_1 \xrightarrow{a}_l e_1'}{e_1 \cdot e_2 \xrightarrow{a}_l e_1' \cdot e_2} \qquad \frac{e_1 \xrightarrow{a}_{\text{bo}} \surd}{e_1 \cdot e_2 \xrightarrow{a}_{\text{bo}} e_2}$$

$$\frac{e_1 \xrightarrow{a}_l e_1'}{e_1{}^{\circledast}e_2 \xrightarrow{a}_{[|e_1|_{\circledast}+1]} e_1' \cdot (e_1{}^{\circledast}e_2)}\; \text{if } e_1 \text{ is normed}$$

$$\frac{e_1 \xrightarrow{a}_l e_1'}{e_1{}^{\circledast}e_2 \xrightarrow{a}_{\text{bo}} e_1' \cdot (e_1{}^{\circledast}e_2)}\; \text{if } e_1 \text{ is not normed}$$

$$\frac{e_1 \xrightarrow{a}_{\text{bo}} \surd}{e_1{}^{\circledast}e_2 \xrightarrow{a}_{[|e_1|_{\circledast}+1]} e_1{}^{\circledast}e_2} \qquad \frac{e_2 \xrightarrow{a}_l \xi}{e_1{}^{\circledast}e_2 \xrightarrow{a}_{\text{bo}} \xi}$$

for $l \in \{\text{bo}\} \cup \{[\alpha] \mid \alpha \in \mathbb{N}^+\}$, where we employed notation defined in Def. 2.4 for writing marking labels as subscripts.

Example 3.6. In Fig. 1 we depict the entry/body-labelings, as defined in Def. 3.2, for star expressions e_1, e_0, and e_2 in Ex. 2.6. It is easy to verify that these labelings are LLEE-witnesses of the charts $C(e_0)$, $C(e_1)$, and $C(e_2)$ in Ex. 2.6, resp..

Proposition 3.7. *For every $e \in StExp(A)$, the entry/body-labeling $\widehat{C(e)}$ of $C(e)$ is a* LLEE*-witness of $C(e)$.*

For a binary relation R, let R^+ and R^* be its transitive and transitive-reflexive closures. $u \rightarrow_l v$ denotes that there is a transition $u \xrightarrow{a}_l v$ for an $a \in A$, and in proofs (but not pictures) $u \rightarrow v$ denotes that $u \rightarrow_l v$ for some label l. By $u \xrightarrow[\not\in(w)]{}_l v$ we denote that $u \rightarrow_l v$ and $v \neq w$ (this transition

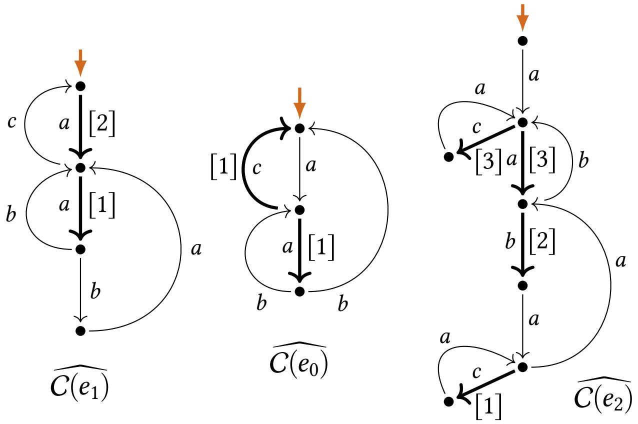


Figure 1. LLEE-witness entry/body-labelings as defined by Def. 3.5 for the chart interpretations of e_0, e_1, and e_2 in Ex. 2.6.

avoids target w). Likewise, $u \xrightarrow[\not\leftarrow(w)]{} v$ denotes that $u \xrightarrow[\not\leftarrow(w)]{}_{l} v$ for some label l. By $\mathrm{scc}(u)$ we denote the strongly connected component (scc) to which u belongs.

Definition 3.8. Let $\hat{C}$ be a LLEE-witness of chart C. If there is a path $v \xrightarrow[\not\leftarrow(v)]{}_{[\alpha]} \cdot \xrightarrow[\not\leftarrow(v)]{}^{*}_{\mathrm{bo}} w$, then we write $v \,{}^{\alpha}\!\!\rightsquigarrow w$. (Note that $v \,{}^{\alpha}\!\!\rightsquigarrow w$ holds if and only if w is a vertex $\neq v$ of the loop chart $C_{\hat{C}}(v, \alpha)$ that is generated by the $\rightarrow_{[\alpha]}$ entry transitions at v in C.) We write $v \rightsquigarrow w$ and say that v *descends in a loop to* w if $v \,{}^{\alpha}\!\!\rightsquigarrow w$ for some $\alpha \in \mathbb{N}^{+}$.

We write $w \circlearrowleft v$ (or $v \circlearrowright w$), and say that w *loops back to* v, if $v \rightsquigarrow w \rightarrow^{+}_{\mathrm{bo}} v$. The loops-back-to relation $\circlearrowleft$ totally orders its successors (see Lem. 3.9, (vi)). Therefore we define the 'direct successor relation' ${}_{d}\!\circlearrowleft$ of $\circlearrowleft$ as follows: We write $w \,{}_{d}\!\circlearrowleft v$ (or $v \,{}_{d}\!\circlearrowright w$), and say that w *directly loops back to* v, if $w \circlearrowleft v$ and for all u with $w \circlearrowleft u$ either $u = v$ or $v \circlearrowleft u$.

Lemma 3.9. *The relations* $\rightarrow_{\mathrm{bo}}$, $\rightsquigarrow$, $\circlearrowleft$, ${}_{d}\!\circlearrowleft$ *as defined by a* LLEE-*witness* $\hat{C}$ *on a chart* C *satisfy the following properties:*

(i) There are no infinite $\rightarrow_{\mathrm{bo}}$ *paths (so no* $\rightarrow_{\mathrm{bo}}$ *cycles).*
(ii) If $\mathrm{scc}(u) = \mathrm{scc}(v)$, *then* $u \rightsquigarrow^{*} v$ *implies* $v \circlearrowleft^{*} u$.
(iii) If $v \rightsquigarrow w$ *and* $\neg(w \circlearrowleft)$, *then* w *is not normed.*
(iv) $\mathrm{scc}(u) = \mathrm{scc}(v)$ *if and only if* $u \circlearrowleft^{*} w$ *and* $v \circlearrowleft^{*} w$ *for some vertex* w.
(v) $\circlearrowleft^{*}$ *is a partial order with the least-upper-bound property: if a nonempty set of vertices has an upper bound with respect to* $\circlearrowleft^{*}$, *then it has a least upper bound.*
(vi) $\circlearrowleft$ *is a total order on* $\circlearrowleft$*-successor vertices: if* $w \circlearrowleft v_1$ *and* $w \circlearrowleft v_2$, *then* $v_1 \circlearrowleft v_2$ *or* $v_1 = v_2$ *or* $v_2 \circlearrowleft v_1$.
(vii) If $v_1 \,{}_{d}\!\circlearrowleft u$ *and* $v_2 \,{}_{d}\!\circlearrowleft u$ *for distinct* v_1, v_2, *then there is no vertex* w *such that both* $w \circlearrowleft^{*} v_1$ *and* $w \circlearrowleft^{*} v_2$.

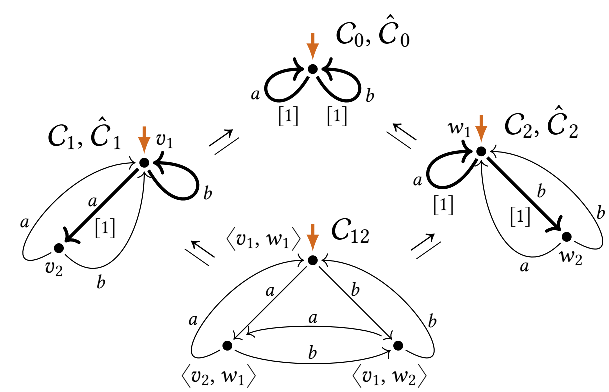
4 The completeness proof, anticipated

After having introduced LLEE-charts as our crucial auxiliary concept, we now sketch the completeness proof. In doing so we need to anticipate four results that will be developed in the next two sections: **(C)** The bisimulation collapse of a LLEE-chart is again a LLEE-chart. **(E)** From every LLEE-chart a provable solution can be extracted. **(S)** All provable solutions of LLEE-charts are provably equal. **(P)** All provable solutions can be pulled back from the target to the source chart of a functional bisimulation.

Then completeness of BBP can be argued as follows. Suppose that we are given two star expressions e_1 and e_2 with bisimilar chart interpretations $C(e_1)$ and $C(e_2)$. First we find by Prop. 3.7 that $C(e_1)$ and $C(e_2)$ are LLEE-charts. Then we obtain by Prop. 2.9 that e_1 and e_2 are principal values of provable solutions of $C(e_1)$ and $C(e_2)$, respectively. These charts have the same bisimulation collapse C. By (**C**, Thm. 6.9), C is again a LLEE-chart. Use (**E**, Prop. 5.5) to build a provable solution s of C; let its principal value be e. Apply (**P**, Prop. 5.1) to transfer s backwards over the functional bisimulations to obtain provable solutions s_1 and s_2 of $C(e_1)$ and $C(e_2)$, respectively. By construction, s_1 and s_2 have the same principal value e as s. Finally, by using (**S**, Prop. 5.8), e_1 and e_2 are both provably equal to e. Hence, $e_1 =_{\mathrm{BBP}} e =_{\mathrm{BBP}} e_2$.

In his completeness proof for regular expressions in formal language theory, Salomaa [24] argued 'upwards' from two equivalent regular expressions to a larger regular expression that can be homomorphically collapsed onto both of them. In contrast, our proof approach forces us 'downwards' to the bisimulation collapse, because in the opposite direction the property of being a LLEE-chart may be lost.

Example 4.1. The picture below highlights why we cannot adopt Salomaa's proof strategy of linking two language-equivalent regular expressions via the product of the DFAs they represent. The bisimilar LLEE-charts C_1 and C_2 are interpretations of $(a \cdot (a + b) + b)^{\circledast}0$ and $(b \cdot (a + b) + a)^{\circledast}0$, respectively (the indicated labelings $\hat{C}_1$ and $\hat{C}_2$ are LLEE-witnesses). But their product C_{12} is a not a LLEE-chart; it is of the form of one the not expressible charts from Ex. 2.5 (but has different transition labels). Yet their common bisimulation collapse C_0, the chart interpretation of $(a + b)^{\circledast}0$, is a LLEE-chart with LLEE-witness $\hat{C}_0$.



In view of $C_1 \leftleftarrows C_{12} \rightrightarrows C_2$ this also shows that LLEE-charts are not closed under converse functional bisimilarity $\leftleftarrows$.

5 Extraction of star expressions from, and transferral between, LLEE-charts

In this section we develop the results **(E)**, **(S)**, and **(P)** as mentioned in Sect. 4. We start with the statement **(P)**.

Proposition 5.1 (requires BBP-axioms (B1), (B2), (B3))**.** *Let $\phi : V_1 \to V_2$ be a functional bisimulation between charts C_1 and C_2. If $s_2 : V_2 \setminus \{\surd\} \to StExp(A)$ is a provable solution of C_2, then $s_2 \circ \phi : V_1 \setminus \{\surd\} \to StExp(A)$ is a provable solution of C_1 with the same principal value as s_2.*

Proof (Idea). The bisimulation clauses make it possible to demonstrate the condition for $s_2 \circ \phi$ to be a provable solution of C_1 at w by using the condition for the provable solution s_2 of C_2 at $\phi(w)$, together with the axioms (B1), (B2), (B3). □

We now turn to proving results **(E)** and **(S)** from Sect. 4.

We show that from every chart C with LLEE-witness $\hat{C}$ a provable solution $s_{\hat{C}}$ of C can be extracted. Intuitively, the extraction process follows a specifically chosen run of the loop-elimination procedure on C that is guided by the LLEE-witness $\hat{C}$. For the run we demand that loop subcharts that are generated by the loop-entry transitions from the same vertex v are removed in a row. For this, we can pick vertices v in the remaining LLEE-witness with entry step level $|v|_{\mathrm{en}}$ (see below) minimal. Now extraction synthesizes a star expression e_1 whose behavior captures the behaviour that is accessible via loop-entry transitions from v in the eliminated loop subcharts at v, and in previously eliminated inner loop subcharts. Together with a (later synthesized) expression e_2 that represents the behaviour that is accessible via body transitions from v, the expression e_1 is part of an iteration expression $e_1{}^{\circledast} e_2$ that forms the solution value at v, which represents the behaviour that is accessible from v.

This idea motivates an inside-out extraction process that works with partial solutions, and eventually builds up a provable solution of C. In particular, we inductively define 'relative extracted solutions' $t_{\hat{C}}(w, v)$ for vertices v and w where w is in a loop subchart $C_{\hat{C}}(v, \alpha)$ at v, for some $\alpha \in \mathbb{N}^+$, that is, $v \mathrel{{}^{\alpha}{\curvearrowright}} w$. Hereby $t_{\hat{C}}(w, v)$ captures the part of the behavior in C from w until v is reached. Then we define the from $\hat{C}$ 'extracted solution' $s_{\hat{C}}(v)$ at v by using the relative solutions $t_{\hat{C}}(w_j, v)$ for all targets w_j of loop-entry transitions from v to define the iteration part e_1 of the extracted solution $s_{\hat{C}}(v) = e_1{}^{\circledast} e_2$ at v. We start with a preparation.

Let $\hat{C}$ be a LLEE-witness, and let v be a vertex of $\hat{C}$. By the *entry step level* $|v|_{\mathrm{en}}$ *of* v we mean the maximum loop level of a loop-entry transition in $\hat{C}$ that departs from v, or 0 if no loop-entry transition departs from v. By the *body step norm* $\|v\|_{\mathrm{bo}}$ *of* v we mean the maximal length of a body transition path in C from v (well-defined by Lem. 3.9, (i)).

Lemma 5.2. *For all vertices v, w in a chart C with* LLEE*-witness $\hat{C}$ it holds (for the concepts as defined with respect to $\hat{C}$):*

(i) $v \to_{\mathrm{bo}} w \Rightarrow \|v\|_{\mathrm{bo}} > \|w\|_{\mathrm{bo}}$,

(ii) $v \curvearrowright w \Rightarrow |v|_{\mathrm{en}} > |w|_{\mathrm{en}}$.

Definition 5.3. Let $\hat{C}$ be a LLEE-witness of a chart C. Then the *relative extraction function of* $\hat{C}$ is defined inductively as:

$$t_{\hat{C}} : \{\langle w, v\rangle \mid v, w \in V\setminus\{\surd\}, v \curvearrowright w\} \to StExp(A),$$

$$t_{\hat{C}}(w, v) := \Big(\Big(\Big(\sum_{i=1}^{m} a_i\Big) + \Big(\sum_{j=1}^{n} b_i \cdot t_{\hat{C}}(w_j, w)\Big)\Big)^{\circledast} \Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot t_{\hat{C}}(u_j, v)\Big)\Big)\Big),$$

provided that w has loop-entry transitions $\{w \xrightarrow{a_i}_{[\alpha_i]} w \mid i = 1, \ldots, m\} \cup \{w \xrightarrow{b_j}_{[\beta_j]} w_j \mid j = 1, \ldots, n \wedge w_j \neq w\}$ and body transitions $\{w \xrightarrow{c_i}_{\mathrm{bo}} v \mid i = 1, \ldots, p\} \cup \{w \xrightarrow{d_j}_{\mathrm{bo}} u_j \mid j = 1, \ldots, q \wedge u_j \neq v\}$. Hereby the induction proceeds on $\langle |v|_{\mathrm{en}}, \|w\|_{\mathrm{bo}}\rangle$ with the lexicographic order $<_{lex}$ on $\mathbb{N} \times \mathbb{N}$: For $t_{\hat{C}}(w_j, w)$ we have $\langle |w|_{\mathrm{en}}, \|w_j\|_{\mathrm{bo}}\rangle <_{lex} \langle |v|_{\mathrm{en}}, \|w\|_{\mathrm{bo}}\rangle$ due to $|w_j|_{\mathrm{en}} < |v|_{\mathrm{en}}$, which follows from $v \curvearrowright w$ by Lem. 5.2, (ii). For $t_{\hat{C}}(u_j, v)$ we have $\langle |v|_{\mathrm{en}}, \|u_j\|_{\mathrm{bo}}\rangle <_{lex} \langle |v|_{\mathrm{en}}, \|w\|_{\mathrm{bo}}\rangle$ due to $\|u_j\|_{\mathrm{bo}} < \|w\|_{\mathrm{bo}}$, which follows from $w \to_{\mathrm{bo}} u_j$ by Lem 5.2, (i).

The *extraction function of* $\hat{C}$ is defined by:

$$s_{\hat{C}} : V\setminus\{\surd\} \to StExp(A),$$

$$s_{\hat{C}}(w) := \Big(\Big(\Big(\sum_{i=1}^{m} a_i\Big) + \Big(\sum_{j=1}^{n} b_j \cdot t_{\hat{C}}(w_j, w)\Big)\Big)^{\circledast} \Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot s_{\hat{C}}(u_j)\Big)\Big)\Big),$$

with induction on $\|w\|_{\mathrm{bo}}$, provided that w has loop-entry transitions $\{w \xrightarrow{a_i}_{[\alpha_i]} w \mid i = 1, \ldots, m\} \cup \{w \xrightarrow{b_j}_{[\beta_j]} w_j \mid j = 1, \ldots, n \wedge w_j \neq w\}$ and body transitions $\{w \xrightarrow{c_i}_{\mathrm{bo}} \surd \mid i = 1, \ldots, p\} \cup \{w \xrightarrow{d_j}_{\mathrm{bo}} u_j \mid j = 1, \ldots, q \wedge u_j \neq \surd\}$. For $s_{\hat{C}}(u_j)$ the induction hypothesis holds due to $\|u_j\|_{\mathrm{bo}} < \|w\|_{\mathrm{bo}}$, which follows from $w \to_{\mathrm{bo}} u_j$ by Lem. 5.2, (i).

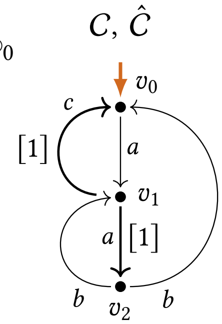
Lemma 5.4 (uses the BBP-axioms (B1)–(B6), (BKS2), but not the rule RSP$^{\circledast}$)**.** *In a chart C with* LLEE*-witness $\hat{C}$, if $v \curvearrowright w$, then $s_{\hat{C}}(w) =_{\mathrm{BBP}} t_{\hat{C}}(w, v) \cdot s_{\hat{C}}(v)$.*

Proposition 5.5 (uses the BBP-axioms (B1)–(B6), (BKS1), (BKS2), but not the rule RSP$^{\circledast}$)**.** *For every* LLEE*-witness $\hat{C}$ of a chart C, the extraction function $s_{\hat{C}}$ is a provable solution of C.*

The proof of Lem. 5.4 proceeds by induction on $\|w\|_{\mathrm{bo}}$; no induction is needed for the proof of Prop. 5.5 (see in [14]).

Example 5.6. Left in Fig. 2 we illustrate the extraction of a provable solution for the LLEE-witness $\hat{C} = \widehat{C(e_0)}$ in Ex. 3.6 of the chart $C = C(e_0)$ in Ex. 2.6. In order to obtain the principal value $s_{\hat{C}}(v_0)$ of the extracted solution $s_{\hat{C}}$, its definition is expanded. It recurs on $s_{\hat{C}}(v_1)$, and then on $t_{\hat{C}}(v_0, v_1)$ and $t_{\hat{C}}(v_2, v_1)$. After computing those star expressions by using the definition of $t_{\hat{C}}$, the principal value can be obtained by substitution. The star expressions $s_{\hat{C}}(v_1)$ and $s_{\hat{C}}(v_2)$ are

$$
\begin{aligned}
s_{\hat{C}}(v_0) &:= 0^{\circledast}(a \cdot s_{\hat{C}}(v_1)) \\
&=_{\mathrm{BBP}} a \cdot s_{\hat{C}}(v_1) \\
&=_{\mathrm{BBP}} a \cdot (c \cdot a + a \cdot (b + b \cdot a))^{\circledast} 0 \\
s_{\hat{C}}(v_1) &:= (c \cdot t_{\hat{C}}(v_0, v_1) + a \cdot t_{\hat{C}}(v_2, v_1))^{\circledast} 0 \\
&=_{\mathrm{BBP}} (c \cdot a + a \cdot (b + b \cdot a))^{\circledast} 0 \\
t_{\hat{C}}(v_0, v_1) &:= 0^{\circledast} a \\
&=_{\mathrm{BBP}} a \\
t_{\hat{C}}(v_2, v_1) &:= 0^{\circledast}(b + b \cdot t_{\hat{C}}(v_0, v_1)) \\
&=_{\mathrm{BBP}} b + b \cdot a \\
s_{\hat{C}}(v_2) &:= 0^{\circledast}(b \cdot s_{\hat{C}}(v_1) + b \cdot s_{\hat{C}}(v_0)) \\
&=_{\mathrm{BBP}} b \cdot s_{\hat{C}}(v_1) + b \cdot (a \cdot s_{\hat{C}}(v_1)) \\
&=_{\mathrm{BBP}} (b + b \cdot a) \cdot s_{\hat{C}}(v_1) \\
&=_{\mathrm{BBP}} (b + b \cdot a) \cdot ((c \cdot a + a \cdot (b + b \cdot a))^{\circledast} 0)
\end{aligned}
$$



$$
\begin{aligned}
s(v_0) &=^{(\mathrm{sol})}_{\mathrm{BBP}} a \cdot s(v_1) \quad (^{(\mathrm{sol})} \text{ means use of 'is provable solution'}) \\
s(v_1) &=^{(\mathrm{sol})}_{\mathrm{BBP}} c \cdot s(v_0) + a \cdot s(v_2) \\
&=^{(\mathrm{sol})}_{\mathrm{BBP}} c \cdot (a \cdot s(v_1)) + a \cdot (b \cdot s(v_1) + b \cdot s(v_0)) \\
&=^{(\mathrm{sol})}_{\mathrm{BBP}} c \cdot (a \cdot s(v_1)) + a \cdot (b \cdot s(v_1) + b \cdot (a \cdot s(v_1))) \\
&=_{\mathrm{BBP}} (c \cdot a + a \cdot (b + b \cdot a)) \cdot s(v_1) + 0 \\
&\Downarrow \text{ applying RSP}^{\circledast} \\
s(v_1) &=_{\mathrm{BBP}} (c \cdot a + a \cdot (b + b \cdot a))^{\circledast} 0 \\
&=_{\mathrm{BBP}} s_{\hat{C}}(v_1) \quad \text{(see in the derivation on the left)} \\
&\Downarrow \\
s(v_0) &=^{(\mathrm{sol})}_{\mathrm{BBP}} a \cdot s(v_1) =_{\mathrm{BBP}} a \cdot s_{\hat{C}}(v_1) =^{(\mathrm{sol})}_{\mathrm{BBP}} s_{\hat{C}}(v_0) \\
&\Downarrow \\
s(v_2) &=^{(\mathrm{sol})}_{\mathrm{BBP}} b \cdot s(v_1) + b \cdot s(v_0) \\
&=_{\mathrm{BBP}} b \cdot s_{\hat{C}}(v_1) + b \cdot s_{\hat{C}}(v_0) =^{(\mathrm{sol})}_{\mathrm{BBP}} s_{\hat{C}}(v_2)
\end{aligned}
$$

Figure 2. Left: the process of extracting the provable solution $s_{\hat{C}}$ of a chart C from an LLEE-witness $\hat{C}$ of C as in the middle. Right: steps for showing that an arbitrary provable solution s of C is BBP-provably equal to the extracted solution $s_{\hat{C}}$.

obtained similarly. For readability we have simplified the arising terms on the way by using the equality $0^{\circledast}x =_{\mathrm{BBP}} x$ (which follows by (B1), (B6), (B7), and (BKS1)).

Lemma 5.7 (uses the BBP-axioms (B1)–(B6), and the rule RSP$^{\circledast}$)**.** *If $v \rightharpoonup w$, then $s(w) =_{\mathrm{BBP}} t_{\hat{C}}(w, v) \cdot s(v)$ for every provable solution s of a chart C with* LLEE*-witness $\hat{C}$.*

Proposition 5.8 (uses the BBP-axioms (B1)–(B6), and the rule RSP$^{\circledast}$)**.** *Let s_1 and s_2 be provable solutions of a* LLEE*-chart. Then $s_1(w) =_{\mathrm{BBP}} s_2(w)$ for all vertices $w \neq \surd$.*

For the *proof* of this proposition, see Fig. 3. The proof of Lem. 5.7 (see in [14]) proceeds by the same induction measure as we used for the relative extraction function.

Example 5.9. In the right half of Fig. 2 we prove that an arbitrary provable solution s of LLEE-chart $C = C(e_0)$ in Ex. 2.6 with LLEE-witness $\hat{C} = \widehat{C(e_0)}$ in Ex. 3.6 is provably equal to the extracted solution $s_{\hat{C}}$ of C. Crucially, the defining conditions for s as a provable solution of C are expanded along the loop at v_1. The loop behavior obtained is the same as that which is used in the definition of $s_{\hat{C}}(v_1)$. By applying the fixed-point rule RSP$^{\circledast}$ we can then deduce BBP-provable equality of $s(v_1)$ and $s_{\hat{C}}(v_1)$. By using the solution conditions for s again, provable equality is then transferred to v_0 and v_1.

6 Preservation of LLEE under collapse

In this section we establish the remaining result **(C)** from Sect. 4 that is crucial for the completeness proof: that the bisimulation collapse of a LLEE-chart is again a LLEE-chart.

This result is achieved by a step-wise construction of a bisimulation collapse. Pairs of bisimilar vertices w_1 and w_2 are collapsed one at a time, whereby the incoming transitions of w_1 are redirected to w_2. The crux is to take care, and to prove, that the resulting chart has again a LLEE-witness.

Definition 6.1. Let C be a chart, with vertices w_1 and w_2.

The *connect-w_1-through-to-w_2 chart* $C^{(w_1)}_{w_2}$ of C is obtained by redirecting all incoming transitions at w_1 over to w_2, and, if w_1 is the start vertex of C, making w_2 the new start vertex; in this way w_1 gets unreachable, and it is removed with other unreachable vertices to obtain a start-vertex connected chart.

Let $\hat{C}$ be an entry/body-labeling of C. Then we define the entry/body-labeling $\hat{C}^{(w_1)}_{w_2}$ of $C^{(w_1)}_{w_2}$ as follows: every transition in $C^{(w_1)}_{w_2}$ that was already a transition τ in C inherits its marking label from τ in $\hat{C}$; and every transition in $C^{(w_1)}_{w_2}$ that arises as the redirection τ_{w_2} to w_2 of a transition τ to w_1 in C such that τ_{w_2} does not coincide with a transition already in C inherits its marking label from τ in $\hat{C}$. This definition of $\hat{C}^{(w_1)}_{w_2}$ prevents the formation of two transitions with the same source, action label, and target, but with different marking labels, which is not permitted for entry/body-labelings. The choice to define $\hat{C}^{(w_1)}_{w_2}$ by giving precedence to marking labels of already existing transitions in $\hat{C}$ over the marking labels of redirections of transitions will be expedient in showing LLEE-witness preservation under transformations.

Lemma 6.2. *If $w_1 \leftrightarrow w_2$ in C, then $C^{(w_1)}_{w_2} \leftrightarrow C$.*

While the connect-through operation of bisimilar vertices in a chart thus results in a bisimilar chart, its application to a LLEE-witness (an entry/body-labeling) does not need to yield a LLEE-witness again: the property LEE may be lost.

Example 6.3. Consider the LLEE-witness $\hat{C}$ in the middle below. The unspecified action labels are assumed to facilitate that w_1 and w_2 are bisimilar. Hence also $\widehat{w}_1$ and $\widehat{w}_2$ are

Proof (of Prop. 5.8). Let $\hat{C}$ be a LLEE-witness of a chart C. Let s be a provable solution of C. We have to show that $s(w) =_{\mathrm{BBP}} s_{\hat{C}}(w)$ for all $w \neq \surd$. For this, let $w \neq \surd$. The derivation below is based on the set representation of transitions from w in $\hat{C}$ as formulated in the definition of $s_{\hat{C}}(w)$. The first derivation step uses that s is a provable solution of C and axioms (B1), (B2), and (B3), the second step uses Lem. 5.7 in view of $w \rightharpoonup w_j$ for $j = 1, \ldots, n$, and the third step uses axioms (B4), (B5), and (B6).

$$\begin{aligned} s(w) &=_{\mathrm{BBP}} \Big(\Big(\sum_{i=1}^{m} a_i \cdot s(w)\Big) + \Big(\sum_{j=1}^{n} b_j \cdot s(w_j)\Big)\Big) + \Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot s(u_j)\Big)\Big) \\ &=_{\mathrm{BBP}} \Big(\Big(\sum_{i=1}^{m} a_i \cdot s(w)\Big) + \Big(\sum_{j=1}^{n} b_j \cdot \big(t_{\hat{C}}(w_j, w) \cdot s(w)\big)\Big)\Big) + \Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot s(u_j)\Big)\Big) \\ &=_{\mathrm{BBP}} \Big(\Big(\sum_{i=1}^{m} a_i\Big) + \Big(\sum_{j=1}^{n} \big(b_j \cdot t_{\hat{C}}(w_j, w)\big)\Big)\Big) \cdot s(w) + \Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot s(u_j)\Big)\Big) \end{aligned}$$

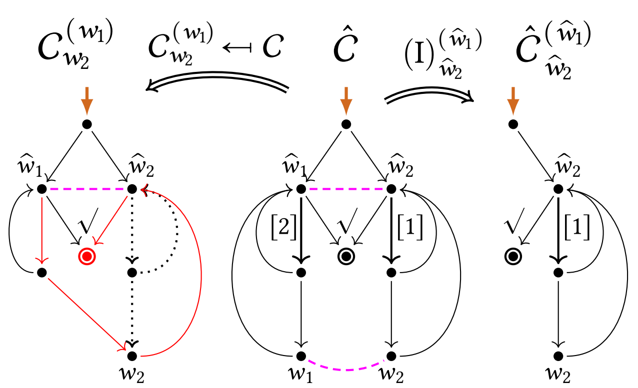
In view of this derived provable equality for $s(w)$, we can now apply the rule $\mathrm{RSP}^{\circledast}$ in order to obtain:

$$s(w) =_{\mathrm{BBP}} \Big(\Big(\sum_{i=1}^{m} a_i\Big) + \Big(\sum_{j=1}^{n} b_j \cdot t_{\hat{C}}(w_j, w)\Big)\Big)^{\circledast}\Big(\Big(\sum_{i=1}^{p} c_i\Big) + \Big(\sum_{j=1}^{q} d_j \cdot s_{\hat{C}}(u_j)\Big)\Big) \equiv s_{\hat{C}}(w)$$

In this last step we have used the definition of $s_{\hat{C}}(w)$. $\square$

Figure 3. Proof of Prop. 5.8.

bisimilar. Bisimilarity is indicated by the broken lines. The connect-w_1-through-to-w_2 chart on the left is not a LLEE-chart, because it does not satisfy LEE: after the loop subchart induced by the downwards transition from $\widehat{w}_2$ is eliminated, and garbage collection is done, the remaining chart without the dotted transitions still has an infinite path; yet it does not contain another loop subchart, because each infinite path can reach $\surd$ without returning to its source. An example of this is the red path from $\widehat{w}_1$ via w_2 and $\widehat{w}_2$ to $\surd$. In $\hat{C}$, the bisimilar pair w_1, w_2 progresses to the bisimilar pair $\widehat{w}_1, \widehat{w}_2$. The connect-$\widehat{w}_1$-through-to-$\widehat{w}_2$ chart on the right is a LLEE-chart, as witnessed by the entry/body-labeling $\hat{C}^{(\widehat{w}_1)}_{\widehat{w}_2}$.



This illustrates that bisimilar pairs of vertices must be selected carefully, to safeguard that the connect-through construction preserves LLEE. The proposition below expresses that a pair of distinct bisimilar vertices can always be selected in one of three mutually exclusive categories. Later, three LLEE-preserving transformations I, II, and III will be defined for each of these categories.

Proposition 6.4. *If a* LLEE*-chart C is not a bisimulation collapse, then it contains a pair of bisimilar vertices w_1, w_2 that satisfy, for a* LLEE*-witness of C, one of the conditions:*
(C1) $\neg(w_2 \to^* w_1) \wedge (\rightharpoonup w_1 \Rightarrow w_2$ *is not normed*$)$,
(C2) $w_2 \circlearrowleft^+ w_1$,
(C3) $\exists v \in V\big(w_1 \,{}_d\!\circlearrowleft v \wedge w_2 \circlearrowleft^+ v\big) \wedge \neg(w_2 \to^*_{\mathrm{bo}} w_1)$.

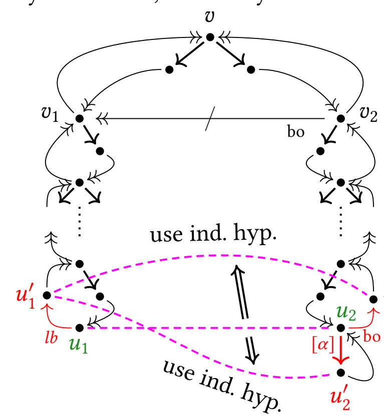
Condition (C1) requires that w_1 and w_2 are in different scc's, as there is no path from w_2 to w_1. The additional proviso in (C1) constrains the pair in such a way that if both are normed, then w_1 must be outside of all loops (otherwise the connect-w_1-through-to-w_2 operation does not preserve LLEE-charts, see Ex. 6.3); its asymmetric formulation helps to avoid the assumption of bisimilarity in Prop. 6.8 below. The two other conditions concern the situation that w_1 and w_2 are in the same scc. While in (C2) w_1 and w_2 are comparable (but different) by the loops-back-to relation $\circlearrowleft^*$, they are incomparable in (C3). In the situation that w_1, w_2 loop back to the same vertex v, but w_1 directly loops back to v, (C3) also demands that no body step path exists from w_2 to w_1 (otherwise the connect-w_1-through-to-w_2 construction does not preserve LLEE-charts, see an example in [14]).

In the proof of Prop. 6.4 we progress, from a given pair of distinct bisimilar vertices, repeatedly via transitions, at one side picking loop-back transitions, over pairs of distinct bisimilar vertices, until one of the conditions (C1) , (C2) , (C3) is met. We will use a subset of the body transitions in a LLEE-witness. By a *loop-back transition*, written as $u \to_{lb} v$, we mean a transition $u \to_{\mathrm{bo}} v$ that stays within an scc, that is, $\mathrm{scc}(u) = \mathrm{scc}(v)$. The *loops-back-to norm* $\|u\|^{\min}_{lb}$ of u is the maximal length of a $\to_{lb}$ path from u (which is well-defined by Lem. 3.9, (i) and chart finiteness). Note that $\|u\|_{lb}^{\min} = 0$ if and only if u does not loop back (denoted by $\neg(u \circlearrowright)$).

Proof of Prop. 6.4. We pick distinct bisimilar vertices u_1, u_2. First we consider the case $\mathrm{scc}(u_1) \neq \mathrm{scc}(u_2)$. Without loss of generality, suppose $\neg(u_2 \to^* u_1)$. We progress to a pair of vertices where (C1) holds, using induction on $\|u_1\|_{lb}^{\min}$. In the base case, $\|u_1\|_{lb}^{\min} = 0$, it suffices to show that it is not possible that both $\curvearrowright u_1$ holds and u_2 is normed, because then we can define $w_1 = u_1$ and $w_2 = u_2$, and are done. Therefore suppose, toward a contradiction, that $\curvearrowright u_1$ holds and u_2 is normed. Then u_1 is normed, too, since u_1 and u_2 are bisimilar. Also $\neg(u_1 \circlearrowright)$ follows from $\|u_1\|_{lb}^{\min} = 0$, which says that there are no loops-back-to steps from u_1. So we get that $\curvearrowright u_1$, $\neg(u_1 \circlearrowright)$, and u_1 is normed. This contradicts Lemma 3.9, (iii). In the induction step, $\|u_1\|_{lb}^{\min} > 0$ implies $u_1 \to_{lb} u_1'$ and $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$ for some u_1'. Since $u_1 \mathrel{\underline{\leftrightarrow}} u_2$, we have $u_2 \to u_2'$ and $u_1' \mathrel{\underline{\leftrightarrow}} u_2'$ for some u_2'. Since $u_1 \to_{lb} u_1'$, by definition, u_1 and u_1' are in the same scc. Hence $u_1' \to^* u_1$. This implies $\neg(u_2' \to^* u_1')$, for else $u_2 \to u_2' \to^* u_1' \to^* u_1$, which contradicts the assumption $\neg(u_2 \to^* u_1)$. Since $u_1' \mathrel{\underline{\leftrightarrow}} u_2'$ and $\neg(u_2' \to^* u_1')$ and $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$, by induction there exists a bisimilar pair w_1, w_2 for which (C1) holds.

Now let $\mathrm{scc}(u_1) = \mathrm{scc}(u_2)$. Then by Lem. 3.9, (iv), $u_1 \circlearrowright^* v$ and $u_2 \circlearrowright^* v$ for some v. By Lem. 3.9, (v) we pick v as the least upper bound of u_1, u_2 with regard to $\circlearrowright^*$. If $u_1 = v$, then $u_2 \circlearrowright^+ u_1$, so (C2) holds for $w_1 = u_1$ and $w_2 = u_2$. If $u_2 = v$, then likewise (C2) holds for $w_1 = u_2$ and $w_2 = u_1$. Now let $u_1, u_2 \neq v$. Since v is the least upper bound, $u_1 \circlearrowright^* v_1 \;{}_d\!\circlearrowright v \;{}_d\!\circlearrowleft v_2 \circlearrowleft^* u_2$ for distinct $v_1, v_2 \in V$. There cannot be a cycle of body transitions, so $\neg(v_2 \to_{\mathrm{bo}}^* v_1)$ or $\neg(v_1 \to_{\mathrm{bo}}^* v_2)$. By symmetry it suffices to consider $\neg(v_2 \to_{\mathrm{bo}}^* v_1)$. Summarizing, $u_1 \circlearrowright^* v_1 \;{}_d\!\circlearrowright v \;{}_d\!\circlearrowleft v_2 \circlearrowleft^* u_2$ and $\neg(v_2 \to_{\mathrm{bo}}^* v_1)$. For this situation we use induction on $\|u_1\|_{lb}^{\min}$. If $u_1 = v_1$, then $u_1 \;{}_d\!\circlearrowright v$; taking $w_1 = u_1$ and $w_2 = u_2$, (C3) holds. So we can assume $u_1 \circlearrowright^+ v_1 \;{}_d\!\circlearrowright v$. Pick a transition $u_1 \to_{lb} u_1'$ with $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$; by definition, $\mathrm{scc}(u_1') = \mathrm{scc}(u_1)$. Since $u_1 \mathrel{\underline{\leftrightarrow}} u_2$, there is a transition $u_2 \to u_2'$ with $u_1' \mathrel{\underline{\leftrightarrow}} u_2'$ for some u_2'. If $\mathrm{scc}(u_1') \neq \mathrm{scc}(u_2')$, then as before we can find bisimilar w_1, w_2 for which (C1) holds. Now let $\mathrm{scc}(u_1') = \mathrm{scc}(u_2')$, so u_1, u_2, u_1', u_2' are in the same scc. Since $u_1 \circlearrowright^+ v_1$ and $u_1 \to u_1'$, either $u_1' = v_1$ or $v_1 \curvearrowright^+ u_1'$. Moreover, $\mathrm{scc}(u_1') = \mathrm{scc}(u_1) = \mathrm{scc}(v_1)$, so by Lem. 3.9, (ii), $u_1' \circlearrowright^* v_1$. Since $u_2 \circlearrowright^* v_2$, we can distinguish two cases (for illustrations for each of the subcases, see the version [14] with appendix).

Case 1: $u_2 \circlearrowright^+ v_2$. Since $u_2 \to u_2'$, either $u_2' = v_2$ or $v_2 \curvearrowright^+ u_2'$. Moreover, $\mathrm{scc}(u_2') = \mathrm{scc}(u_2) = \mathrm{scc}(v_2)$, so by Lem. 3.9, (ii), $u_2' \circlearrowright^* v_2$. Hence, $u_1' \circlearrowright^* v_1 \;{}_d\!\circlearrowright v \;{}_d\!\circlearrowleft v_2 \circlearrowleft^* u_2' \wedge \neg(v_2 \to_{\mathrm{bo}}^* v_1)$, and $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$. We apply the induction hypothesis to obtain a bisimilar pair w_1, w_2 for which (C1), (C2), or (C3) holds. See above for an illustration of both of the cases in which $u_2 \to u_2'$ is a loop-entry transition, or a body transition.



Case 2: $u_2 = v_2$. We distinguish two cases.

Case 2.1: $u_2 \to_{[\alpha]} u_2'$. Then either $u_2' = u_2$ or $u_2 \curvearrowright^+ u_2'$. Moreover, $\mathrm{scc}(u_2') = \mathrm{scc}(u_2)$, so by Lem. 3.9, (ii), $u_2' \circlearrowright^* u_2$, and hence $u_2' \circlearrowright^* v_2$. Thus we have obtained $u_1' \circlearrowright^* v_1 \;{}_d\!\circlearrowright v \;{}_d\!\circlearrowleft v_2 \circlearrowleft^* u_2' \wedge \neg(v_2 \to_{\mathrm{bo}}^* v_1)$. Due to $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$, we can apply the induction hypothesis again.

Case 2.2: $u_2 \to_{\mathrm{bo}} u_2'$. Then $\neg(v_2 \to_{\mathrm{bo}}^* v_1)$ together with $v_2 = u_2 \to_{\mathrm{bo}} u_2'$ and $u_1' \to_{\mathrm{bo}}^* v_1$ (because $u_1' \circlearrowright^* v_1$) imply $u_1' \neq u_2'$. We distinguish two cases.

Case 2.2.1: $u_2' = v$. Then $u_1' \circlearrowright^* v_1 \;{}_d\!\circlearrowright v = u_2'$, i.e., $u_1' \circlearrowright^+ u_2'$, so we are done, because (C2) holds for $w_1 = u_2'$ and $w_2 = u_1'$.

Case 2.2.2: $u_2' \neq v$. By Lem. 3.9, (ii), $u_2' \circlearrowright^+ v$. Hence, $u_2' \circlearrowright^* v_2' \;{}_d\!\circlearrowright v$ for some v_2'. Since $v_2 = u_2 \to_{\mathrm{bo}} u_2' \circlearrowright^* v_2'$ and $\neg(v_2 \to_{\mathrm{bo}}^* v_1)$, it follows that $\neg(v_2' \to_{\mathrm{bo}}^* v_1)$. So $u_1' \circlearrowright^* v_1 \;{}_d\!\circlearrowright v \;{}_d\!\circlearrowleft v_2' \circlearrowleft^* u_2' \wedge \neg(v_2 \to_{\mathrm{bo}}^* v_1')$. Due to $\|u_1'\|_{lb}^{\min} < \|u_1\|_{lb}^{\min}$, we can apply the induction hypothesis again.

This exhaustive case analysis concludes the proof. ◻

Now we define, for LLEE-witnesses $\hat{C}$ of a LLEE-chart C, and for bisimilar vertices w_1, w_2 in C, in each of the three cases (C1), (C2), or (C3) of Prop. 6.4 a transformation of $\hat{C}$ into an entry/body-labeling of the connect-w_1-through-to-w_2 chart $C_{w_2}^{(w_1)}$ that can be shown to be a LLEE-witness again. We number the *transformations* for (C1), (C2), and (C3) as *I*, *II*, and *III*, respectively. Each transformation makes use of the *connect-through construction for entry/body-labelings* as defined in Def. 6.1. Additionally, in each transformation an adaptation of labels of transitions is performed, to avoid violations of LLEE-witness properties. In transformations I and III the adaptation is performed <u>before</u> connecting w_1 through to w_2, and is needed to guarantee that layeredness is preserved; in transformation II it is performed right <u>after</u> eliminating w_1, and avoids the creation of body step cycles. The *level adaptations* for the three transformations are:

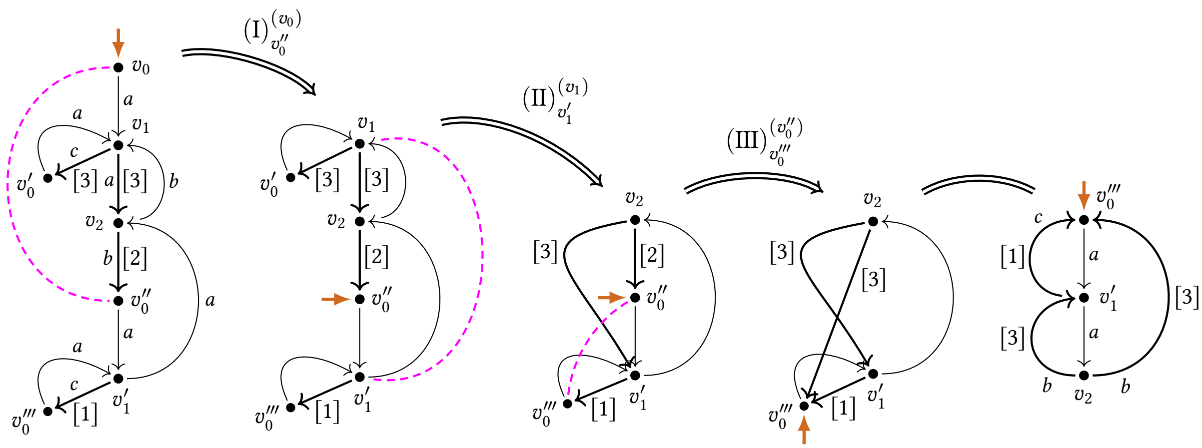


Figure 4. Three connect-through-steps according to the transformations I, II, and III from the LLEE-witness on the left, and a final isomorphic deformation, leading to the LLEE-witness on the right. For clarity, we neglected action labels in the middle.

L_{I} Let $m = \max\{\, \beta \,:\, \text{there is a path } w_2 \rightarrow^* \cdot \rightarrow_{[\beta]} \text{ in } \hat{C}\,\}$. In loop-entry transitions $u \rightarrow_{[\alpha]} v$ for which there is a path $v \rightarrow^* w_1$ in C, replace α by an α' with $\alpha' = \alpha + m$. This increases the labels of loop-entry transitions that descend to w_1 in $\hat{C}$ to a higher level than the loop labels reachable from w_2.

L_{II} Since $w_2 \circlearrowleft^+ w_1$, there exists a $\widehat{w}_2$ with $w_2 \circlearrowleft^* \widehat{w}_2 \,{}_{d}\!\circlearrowleft w_1$. Let γ be the maximum loop level among the loop-entries at w_1 in $\hat{C}$. (Note that since $w_2 \circlearrowleft^+ w_1$, there is at least one such transition.) Turn the body transitions from $\widehat{w}_2$ into loop-entry transitions with loop label γ.

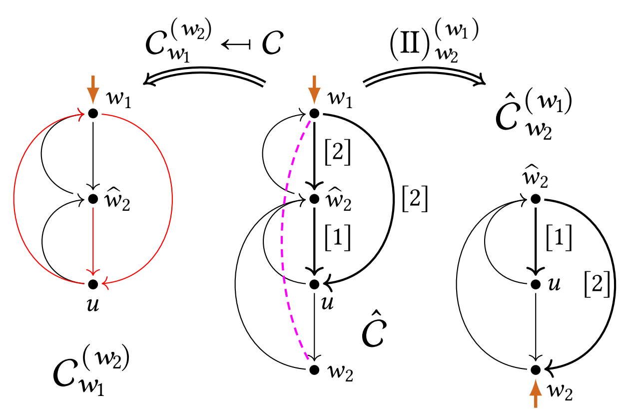
L_{III} Let γ be a loop label of maximum level among the loop-entry transitions at v in $\hat{C}$. (Note that since $w_1 \circlearrowleft v$, there is at least one such transition.) Turn the loop labels of the loop-entry transitions from v into γ.

Each of these transformations ends with a *clean-up step*: if the loop-entry transitions from a vertex with the same loop label no longer induce an infinite path (due to the removal of w_1), then they are changed into body transitions.

Example 6.5. The LLEE-witness on the left in Fig. 4 is reduced in three transformation steps to a LLEE-witness of the chart $C(e_0)$ in Ex. 2.6. Broken lines are between bisimilar vertices. In step one, a transformation I, the start state v_0 is connected through to the bisimilar vertex v_0'', whereby v_0'' becomes the start vertex; note that there is no path from v_0'' to v_0, and no vertex descends into a loop to v_0. In step two, a transformation II, v_1 is connected through to the bisimilar vertex v_1'; note that $v_1' \circlearrowleft^+ v_1$. In step three, a transformation III, the start vertex v_0'' is connected through to the bisimilar vertex v_0''', whereby v_0''' becomes the start vertex; note that $v_0'' \,{}_{d}\!\circlearrowleft v_2$ and $v_0''' \circlearrowleft^+ v_2$ and there is no body step path from v_0''' to v_0''. By the loop level adaptation L_{III}, all loop entries from v_2 get level 3. The final step is an isomorphic deformation. Only the left and right charts depict actions.

The following examples provide more illustrations of the transformations II and III. Similarly as Ex. 6.3 does so for transformation I and (C1), they also show that the conditions (C2) and (C3) mark rather sharp borders between whether, on a given LLEE-witness, a connect-through operation is possible while preserving LLEE, or not.

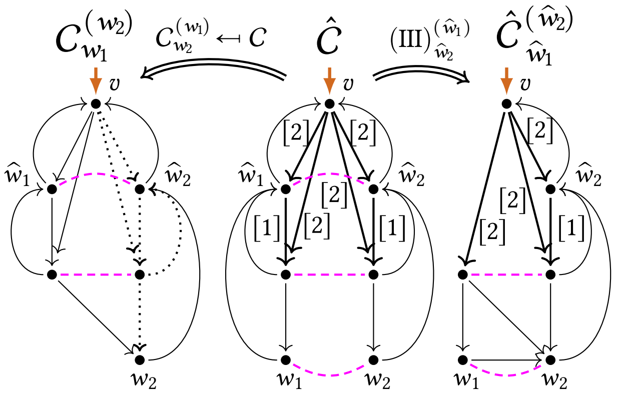
Example 6.6. For the LLEE-witness $\hat{C}$ below in the middle, the chart $C^{(w_2)}_{w_1}$ on the left has no LLEE-witness. It does not satisfy LEE: it has no loop subchart, since from each of its three vertices an infinite path starts that does not return to this vertex; from $\widehat{w}_2$ this path, drawn in red, cycles between u and w_1. Transformation II applied to the pair w_1, w_2 (instead of w_2, w_1) in $\hat{C}$ yields the entry/body-labeling $\hat{C}^{(w_1)}_{w_2}$ where $\widehat{w}_2 \rightarrow_{\mathrm{bo}} w_2$ is turned into $\widehat{w}_2 \rightarrow_{[2]} w_2$. As the pair w_1, w_2 satisfies (C2) , the proof of Prop. 6.8 ensures that this labeling, drawn on the right, is a LLEE-witness.



Example 6.7. In the LLEE-witness $\hat{C}$ below in the middle, $w_1, w_2 \circlearrowleft^+ v$ and there is no body step path from w_2 to w_1, but (C3) does not hold for the pair w_1, w_2 due to $\neg(w_1 \,{}_{d}\!\circlearrowleft v)$. The chart $C^{(w_1)}_{w_2}$ on the left has no LLEE-witness. It does not satisfy LEE: the downwards loop-entry transition from $\widehat{w}_2$ can be eliminated, and then two more arising loop-entry

transitions from v; the remaining chart of solid arrows has no further loop subchart, because from each of its vertices an infinite path starts that does not return to this vertex.

In $\hat{C}$, loop-entry transitions from v have the same loop label, so the preprocessing step of transformation III is void. The bisimilar pair w_1, w_2 progresses to the bisimilar pair $\widehat{w}_1, \widehat{w}_2$ in $\hat{C}$, for which (C3) holds because $\widehat{w}_1 \;{}_d\!\circlearrowleft v \circlearrowleft \widehat{w}_2$ and $\neg(\widehat{w}_2 \to^*_{\mathrm{bo}} \widehat{w}_1)$. Transformation III applied to this pair yields the labeling $\hat{C}^{(\widehat{w}_1)}_{\widehat{w}_2}$ on the right. In the proof of Prop. 6.8 it is argued that this is guaranteed to be a LLEE-witness. The remaining two bisimilar pairs can be eliminated by one or by two further applications of transformation III.



Proposition 6.8. *Let C be a LLEE-chart. If a pair $\langle w_1, w_2\rangle$ of vertices satisfies (C1), (C2), or (C3) with respect to a* LLEE*-witness of C, then $C^{(w_1)}_{w_2}$ is a* LLEE*-chart.*

Proof. Let $\hat{C}$ be a LLEE-witness. For vertices w_1, w_2 such that (C1) , (C2) , or (C3) holds, transformation I, II, or III, respectively, produces an entry/body-labeling $\hat{C}^{(w_1)}_{w_2}$. We prove for transformation I that this is a LLEE-witness, and refer to the report version [14] with regard to transformations II, and III.

We first argue it suffices to show that each of the transformations produces, before the final clean-up step, a labeling that satisfies the LLEE-witness conditions, except possible violations of loop property (L1) in (W2)(a). Such violations can be removed from a loop-labeling while preserving the other LLEE-witness conditions. To show this, suppose (L1) is violated in some $C_{\hat{C}}(u, \alpha)$. Then $u \to_{[\alpha]}$ but $\neg(u \to_{[\alpha]} \cdot \to^*_{\mathrm{bo}} u)$. Let $\hat{C}_1$ be the result of removing this violation by changing the α-loop-entry transitions from u into body transitions. No new violation of (L1) is introduced in $\hat{C}_1$. (W1) and (W2)(a), (L2), are preserved in $\hat{C}_1$ because an introduced infinite body step path in $\hat{C}_1$ would be a body step cycle that stems from a path $u \to_{[\alpha]} u' \to^*_{\mathrm{bo}} u$ in $\hat{C}$. (W2)(b) might only be violated by a path $w \xrightarrow[\not{t}(w)]{}_{[\beta]} \cdot \xrightarrow[\not{t}(w)]{}^*_{\mathrm{bo}} u \xrightarrow[\not{t}(w,u)]{}_{\mathrm{bo}} u' \xrightarrow[\not{t}(w,u)]{}^*_{\mathrm{bo}} \cdot \to_{[\gamma]}$ with $\beta \leqslant \gamma$ in $\hat{C}_1$ where $u \to_{\mathrm{bo}} u'$ stems from $u \to_{[\alpha]} u'$ in $\hat{C}$; then $\beta > \alpha > \gamma$ by layeredness of $\hat{C}$; so (W2)(b) is preserved. Analogously we find that also (W2)(a), (L3) is preserved, because $\surd$ is never in $C_{\hat{C}}(u, \alpha)$.

To show the correctness of transformation I, consider vertices w_1 and w_2 with (C1) . We show that the result $\hat{C}^{(w_1)}_{w_2}$ of transformation I before the clean-up step satisfies the LLEE-witness properties, except for possible violations of (L1).

To verify (W1) and part (L2) of (W2)(a), it suffices to show that $\hat{C}^{(w_1)}_{w_2}$ does not contain body step cycles. The original loop-labeling $\hat{C}$ is a LLEE-witness, so it does not contain body step cycles. Since the level adaptation step does not turn loop-entry steps into body steps, body step cycles could only arise in the step connecting w_1 through to w_2. Suppose such a body step cycle arises. Then there must be a transition $u \to_{\mathrm{bo}} w_1$ in $\hat{C}$ (which is redirected to w_2 in $\hat{C}^{(w_1)}_{w_2}$) and a path $w_2 \to^*_{\mathrm{bo}} u$ in $\hat{C}$. But then $w_2 \to^*_{\mathrm{bo}} u \to_{\mathrm{bo}} w_1$ in C, which contradicts (C1) that there is no path from w_2 to w_1. Hence (W1) and part (L2) of (W2)(a) hold for $\hat{C}^{(w_1)}_{w_2}$.

Now we verify part (L3) of (W2)(a) in $\hat{C}^{(w_1)}_{w_2}$. Consider a path $u \xrightarrow[\not{t}(u)]{}_{[\alpha]} \cdot \xrightarrow[\not{t}(u)]{}^*_{\mathrm{bo}} w_1$ in $\hat{C}$. Then $u \neq w_1$, and $u \curvearrowright w_1$. It suffices to show that then $\neg(w_2 \to^+ \surd)$ in C. But this is guaranteed, because otherwise w_2 were normed, and due to $u \curvearrowright w_1$ we would have a contradiction with condition (C1) .

Finally we show that (W2)(b) is preserved in $\hat{C}^{(w_1)}_{w_2}$ by both the level adaptation and the connect-through step. First, since in the level adaptation step all adapted loop labels are increased with the same value m, a violation of (W2)(b) would arise by a path $u \to_{[\alpha]} \cdot \to^*_{\mathrm{bo}} \cdot \to_{[\beta]} v$ in $\hat{C}$ where loop label β is increased while α is not. But such a path cannot exist. Since β is increased, there is a path $v \to^* w_1$ in C. But then there is a path $u \to_{[\alpha]} \cdot \to^+ v \to^* w_1$ in $\hat{C}$, which implies that also α is increased in the level adaptation step. Second, a violation of (W2)(b) in the connect-through step would arise from paths $u \to_{[\alpha]} \cdot \to^*_{\mathrm{bo}} w_1$ and $w_2 \to^*_{\mathrm{bo}} \cdot \to_{[\beta]}$ in $\hat{C}'$ with $\alpha \leqslant \beta$. However, in view of the path $u \to_{[\alpha]} \cdot \to^* w_1$, the loop label α was increased with m in the level adaptation step . On the other hand, in view of (C1) that there is no path from w_2 to w_1 in C, w_1 is unreachable at the end of the path $w_2 \to^* \cdot \to_{[\beta]}$. Hence this loop label β was not increased in the level adaptation step. So it is guaranteed that for such a pair of paths in $\hat{C}^{(w_1)}_{w_2}$ always $\alpha > \beta$.

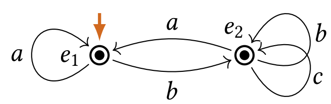
We conclude that the result of transformation I is again a LLEE-witness. □

Theorem 6.9. *The bisimulation collapse of a* LLEE*-chart is again a* LLEE*-chart.*

Proof. Given a LLEE-chart C, repeat the following step: based on a LLEE-witness pick, by Prop. 6.4, bisimilar vertices w_1 and w_2 with (C1), (C2), or (C3), and then connect w_1 through to w_2, obtaining by Prop. 6.8 a LLEE-chart bisimilar to C, due to Lem. 6.2. Hence the bisimulation collapse of C, which is reached eventually, is a LLEE-chart. □

Corollary 6.10. *If a chart is expressible by a star expression modulo bisimilarity, then its collapse is a* LLEE*-chart.*

The converse statement holds as well. But this corollary does not hold for star expressions with 1 and unary star. For example, with respect to the TSS for the process interpretation of star expressions from this class, see e.g. [3], the expression $e_1 := (((1 \cdot a^*) \cdot (b \cdot c^*)) \cdot e$ with $e := (a^* \cdot (b \cdot c^*))^*$ has the following interpretation, where $e_2 := (1 \cdot c^*) \cdot e$:



This is a chart in the extended sense in which immediate termination is permitted at arbitrary vertices. It is a bisimulation collapse that does not satisfy LEE, taking into account that, in the definition of 'loop' for charts in the extended sense, (L3) needs to be changed to exclude immediate termination for vertices in a loop chart other than the start vertex. Note that in the chart above, after elimination of the cycling transitions at e_1 and at e_2, which define loop subcharts, a chart without loop subcharts but with still an infinite behavior is obtained. Therefore this chart does not satisfy LEE.

An idea to overcome this problem is to interpret star expressions with 1 as '1-charts', that is, as charts with special 1-transitions. Such 1-transitions are analogous to ϵ-transitions for finite-state automata. But they have a different interpretation in the context of the process semantics, namely as explicit empty steps. For this purpose, significantly more formal machinery has to be developed (see also [31]). The main challenge for adapting our proof to the setting of star expressions with 1 consists in refining the stepwise collapse procedure in such a way that it can cope with complications that arise in the presence of 1-transitions.

7 The completeness result, and conclusion

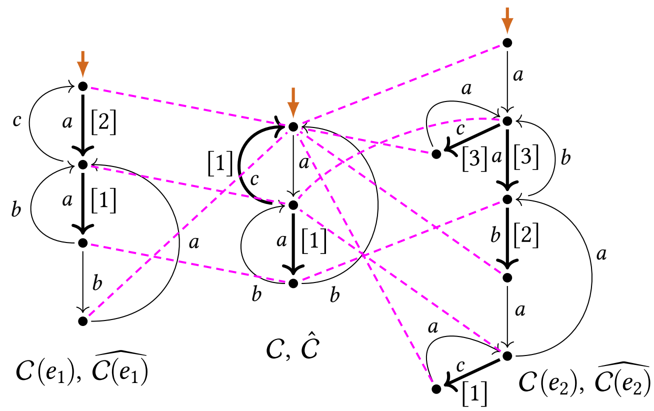
That bisimulation collapse preserves LLEE was the last building block in the proof of the desired completeness result.

Theorem 7.1. *The proof system* BBP *is complete with respect to the bisimulation semantics of star expressions, that is, with respect to bisimilarity of charts that interpret star expressions without* 1 *and with binary Kleene star* $^{\circledast}$.

Proof. The proof steps were already explained in Sect. 4. □

Example 7.2. The bisimilar LLEE-charts C_1 and C_2 in Ex. 4.1 have $(a \cdot (a + b) + b)^{\circledast}0$ and $(b \cdot (a + b) + a)^{\circledast}0$ as their principal solutions. Their bisimulation collapse C_0 has principal solution $(a + b)^{\circledast}0$. Then $(a \cdot (a + b) + b)^{\circledast}0 =_{\mathrm{BBP}} (a + b)^{\circledast}0 =_{\mathrm{BBP}} (b \cdot (a + b) + a)^{\circledast}0$ by Prop. 5.1, Prop. 5.8.

Example 7.3. Revisiting the star expressions e_1, e_2 in Ex. 2.6 with bisimilar chart interpretations $C(e_1)$ and $C(e_2)$, we can apply our proof in order to show that $e_1 =_{\mathrm{BBP}} e_2$. $C(e_1)$ and $C(e_2)$ have provable solutions with principal values e_1 and e_2 by Prop. 2.9. As $C(e_1)$ and $C(e_2)$ are LLEE-charts by Prop. 3.7 with LLEE-witnesses $\widehat{C(e_1)}$ and $\widehat{C(e_2)}$, their bisimulation collapse C is a LLEE-chart by Thm. 6.9. We take here the more familiar $\hat{C}$, but could also take the one obtained in Fig. 4. We saw in Fig. 2 that $\hat{C}$ has a provable solution with principal value $s_{\hat{C}}(v_0) = a \cdot ((c \cdot a + a \cdot (b + b \cdot a))^{\circledast}0)$. Then by Prop. 5.1 and Prop. 5.8 it follows that $e_1 =_{\mathrm{BBP}} s_{\hat{C}}(v_0) =_{\mathrm{BBP}} e_2$.



We have shown that Milner's axiomatization, tailored to star expressions without 1 and with $^{\circledast}$, is complete in bisimulation semantics. At the core of our proof is the graph structure property LLEE, which characterizes the process graphs that can be expressed by star expressions without 1 and with $^{\circledast}$ as charts whose bisimulation collapse is a LLEE-chart.

Completeness of BBP covers completeness of the theory BPA_0^{ω}+RSP^{ω} of perpetual loop iteration $(\cdot)^{\omega}$ [10] in the sense that the latter result can be shown by our means, or by a faithful interpretation $e^{\omega} \mapsto e^{\circledast}0$ of BPA_0^{ω}+RSP^{ω} in BBP.

Completeness of BBP can be extended, also by means of a faithful interpretation, to cover star expressions with 0, 1, and *, but with a syntactic restriction on terms directly under a *: that they can be rewritten to star expressions with only 'harmless' occurrences of 1. This is analogous to the situation that the completeness result from [9, 11] for star expressions without 0 and 1, and with $^{\circledast}$ was extended in [7] to a setting with 1 (but not 0) and *, where a generalized version of the non-empty-word property is disallowed for terms directly under a *. With the interpretation approach, also the result in [7] can be obtained from the one in [9, 11].

The main future goal is to solve Milner's problem entirely by extending our result to the full class of star expressions.

Acknowledgments

We thank Alban Ponse for his suggestion to consider completeness of Milner's axiomatization for the fragment without 1, and Luca Aceto for encouragement, comments on the exposition, and facilitating a visit of the second author to GSSI, from which this paper developed. Also, we thank the reviewers for detailed remarks and suggestions on how to improve the positioning of our completeness result.

References

[1] Stål Aanderaa. 1965. *On the Algebra of Regular Expressions.* Technical Report. Harvard University.

[2] Valentin Antimirov. 1996. Partial Derivatives of Regular Expressions and Finite Automaton Constructions. *Theoretical Computer Science* 155, 2 (1996), 291–319. https://doi.org/10.1016/0304-3975(95)00182-4

[3] Jos Baeten, Flavio Corradini, and Clemens Grabmayer. 2007. A Characterization of Regular Expressions Under Bisimulation. *J. ACM* 54, 2 (2007), 1–28. https://doi.org/10.1145/1219092.1219094

[4] Jan Bergstra, Inge Bethke, and Alban Ponse. 1994. Process Algebra with Iteration and Nesting. *Comput. J.* 37, 4 (1994), 243–258. https://doi.org/10.1093/comjnl/37.4.243

[5] Jan Bergstra, Wan Fokkink, and Alban Ponse. 2001. Process Algebra with Recursive Operations. In *Handbook of Process Algebra*. Elsevier, Amsterdam, 333–389. https://doi.org/10.1016/B978-044482830-9/50023-0

[6] Doeko Bosscher. 1997. *Grammars Modulo Bisimulation.* Ph.D. Dissertation. University of Amsterdam.

[7] Flavio Corradini, Rocco De Nicola, and Anna Labella. 2002. An Equational Axiomatization of Bisimulation over Regular Expressions. *Journal of Logic and Computation* 12, 2 (2002), 301–320. https://doi.org/10.1093/logcom/12.2.301

[8] Wan Fokkink. 1996. *An Axiomatization for the Terminal Cycle.* Technical Report. *Logic Group Preprint Series*, Vol. 167, Utrecht University.

[9] Wan Fokkink. 1996. On the Completeness of the Equations for the Kleene Star in Bisimulation. In *Proc. AMAST'96 (LNCS)*, Vol. 1101. Springer, Berlin, Heidelberg, 180–194.

[10] Wan Fokkink. 1997. Axiomatizations for the Perpetual Loop in Process Algebra. In *Proc. ICALP'97 (LNCS)*, Vol. 1256. Springer, Berlin, Heidelberg, 571–581. https://doi.org/10.1007/3-540-63165-8_212

[11] Wan Fokkink and Hans Zantema. 1994. Basic Process Algebra with Iteration: Completeness of its Equational Axioms. *Comput. J.* 37, 4 (1994), 259–267. https://doi.org/10.1093/comjnl/37.4.259

[12] Clemens Grabmayer. 2005. Using Proofs by Coinduction to Find "Traditional" Proofs. In *Proceedings of CALCO 2005 (LNCS)*, José Luiz Fiadeiro, Neal Harman, Markus Roggenbach, and Jan Rutten (Eds.), Vol. 3629. Springer, Berlin, Heidelberg, 175–193.

[13] Clemens Grabmayer. 2019. Modeling Terms by Graphs with Structure Constraints (Two Illustrations). In *Proc. TERMGRAPH@FSCD'18*, Vol. 288. EPTCS, http://www.eptcs.org/, 1–13. https://doi.org/10.4204/EPTCS.288.1

[14] Clemens Grabmayer and Wan Fokkink. 2020. *A Complete Proof System for 1-Free Regular Expressions Modulo Bisimilarity*. Technical Report. arxiv.org. arXiv:2004.12740

[15] Clemens Grabmayer and Jan Rochel. 2013. Term Graph Representations for Cyclic Lambda-Terms. In *Proc. TERMGRAPH'13*, Vol. 110. EPTCS, http://www.eptcs.org/, 56–73. https://doi.org/10.4204/EPTCS.110.7

[16] Clemens Grabmayer and Jan Rochel. 2014. Maximal Sharing in the Lambda Calculus with Letrec. *SIGPLAN Not.* 49, 9 (aug 2014), 67–80. https://doi.org/10.1145/2692915.2628148

[17] Stephen C. Kleene. 1951. Representation of Events in Nerve Nets and Finite Automata. In *Automata Studies*. Princeton University Press, Princeton, New Jersey, USA, 3–42.

[18] Dexter Kozen. 1994. A Completeness Theorem for Kleene Algebras and the Algebra of Regular Events. *Information and Computation* 110, 2 (1994), 366 – 390. https://doi.org/10.1006/inco.1994.1037

[19] Dexter Kozen and Alexandra Silva. 2020. Left-handed completeness. *Theoretical Computer Science* 807 (2020), 220 – 233. https://doi.org/10.1016/j.tcs.2019.10.040 In memory of Maurice Nivat, a founding father of Theoretical Computer Science - Part II.

[20] Daniel Krob. 1991. Complete systems of B-rational identities. *Theoretical Computer Science* 89, 2 (1991), 207 – 343. https://doi.org/10.1016/0304-3975(91)90395-I

[21] Robin Milner. 1984. A Complete Inference System for a Class of Regular Behaviours. *J. Comput. System Sci.* 28, 3 (1984), 439–466.

[22] David Park. 1981. Concurrency and Automata on Infinite Sequences. In *Proc. 5th GI Conference on Theoretical Computer Science (LNCS)*, Vol. 104. Springer, Berlin, Heidelberg, 167–183. https://doi.org/10.1007/BFb0017309

[23] J.J.M.M. Rutten. 2005. A coinductive calculus of streams. *Mathematical Structures in Computer Science* 15, 1 (2005), 93–147. https://doi.org/10.1017/S0960129504004517

[24] Arto Salomaa. 1966. Two Complete Axiom Systems for the Algebra of Regular Events. *J. ACM* 13, 1 (1966), 158–169. https://doi.org/10.1145/321312.321326

[25] Peter Sewell. 1994. Bisimulation is Not Finitely (First Order) Equationally Axiomatisable. In *Proc. LICS'94*. IEEE, Piscataway, New Jersey, USA, 62–70. https://doi.org/10.1109/LICS.1994.316086

[26] Peter Sewell. 1997. Nonaxiomatisability of Equivalences over Finite State Processes. *Ann. Pure Appl. Logic* 90, 1-3 (1997), 163–191. https://doi.org/10.1016/S0168-0072(97)00036-5

[27] Alexandra Silva. 2010. *Kleene Coalgebra*. Ph.D. Dissertation. University of Nijmengen.

[28] Walter Taylor. 1977. Equational Logic. In *Contributions to Universal Algebra*, B. Csákány and J. Schmidt (Eds.). North-Holland, Amsterdam, 465 – 501. https://doi.org/10.1016/B978-0-7204-0725-9.50040-X

[29] Ken Thompson. 1968. Regular Expression Search Algorithm. *Commun. ACM* 11 (1968), 419–422. https://doi.org/10.1145/363347.363387

[30] Douglas R. Troeger. 1993. Step Bisimulation is Pomset Equivalence on a Parallel Language without Explicit Internal Choice. *Mathematical Structures in Computer Science* 3 (1993), 25–62. https://doi.org/10.1017/S0960129500000116

[31] J.L.M. Vrancken. 1997. The Algebra of Communicating Processes with Empty Process. *Theoretical Computer Science* 177, 2 (1997), 287 – 328. https://doi.org/10.1016/S0304-3975(96)00250-2