



HAL
open science

A Myhill-Nerode Theorem for Higher-Dimensional Automata

Uli Fahrenberg, Krzysztof Ziemiański

► **To cite this version:**

Uli Fahrenberg, Krzysztof Ziemiański. A Myhill-Nerode Theorem for Higher-Dimensional Automata. Application and Theory of Petri Nets and Concurrency (PETRI NETS), Jun 2023, Caparica, Portugal. 10.1007/978-3-031-33620-1_9 . hal-04580371

HAL Id: hal-04580371

<https://hal.science/hal-04580371v1>

Submitted on 19 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Myhill-Nerode Theorem for Higher-Dimensional Automata

Uli Fahrenberg¹ and Krzysztof Ziemiański²

¹ EPITA Research Laboratory (LRE), France

² University of Warsaw, Poland

Abstract. We establish a Myhill-Nerode type theorem for higher-dimensional automata (HDAs), stating that a language is regular precisely if it has finite prefix quotient. HDAs extend standard automata with additional structure, making it possible to distinguish between interleavings and concurrency. We also introduce deterministic HDAs and show that not all HDAs are determinizable, that is, there exist regular languages that cannot be recognised by a deterministic HDA. Using our theorem, we develop an internal characterisation of deterministic languages.

Keywords: higher-dimensional automata; Myhill-Nerode theorem; concurrency theory; determinism

1 Introduction

Higher-dimensional automata (HDAs), introduced by Pratt and van Glabbeek [23, 27, 28], extend standard automata with additional structure that makes it possible to distinguish between interleavings and concurrency. That puts them in a class with other non-interleaving models for concurrency such as Petri nets [22], event structures [21], configuration structures [31, 32], asynchronous transition systems [3, 26], and similar approaches [19, 24, 25, 30], while retaining some of the properties and intuition of automata-like models. As an example, Fig. 1 shows Petri net and HDA models for a system with two events, labeled a and b . The Petri net and HDA on the left side model the (mutually exclusive) interleaving of a and b as either $a.b$ or $b.a$; those to the right model concurrent execution of a and b . In the HDA, this independence is indicated by a filled-in square.

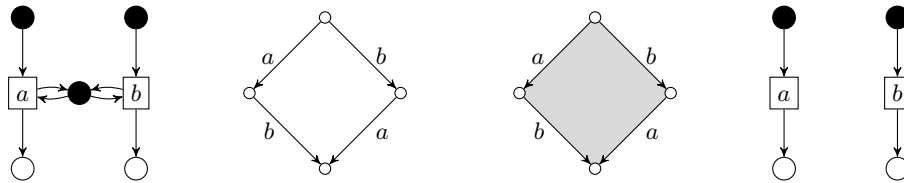


Fig. 1. Petri net and HDA models distinguishing interleaving (left) from non-interleaving (right) concurrency. Left: Petri net and HDA models for $a.b + b.a$; right: HDA and Petri net models for $a \parallel b$.

We have recently introduced languages of HDAs [6], which consist of partially ordered multisets with interfaces (ipomsets), and shown a Kleene theorem for them [7]. Here we continue to develop the language theory of HDAs. Our first contribution is a Myhill-Nerode type theorem for HDAs, stating that a language is regular iff it has finite prefix quotient. This provides a necessary and sufficient condition for regularity. Our proof is inspired by the standard proofs of the Myhill-Nerode theorem, but the higher-dimensional structure introduces some difficulties. For example, we cannot use the standard prefix quotient relation but need to develop a stronger one which takes concurrency of events into account.

As a second contribution, we give a precise definition of deterministic HDAs and show that there exist regular languages that cannot be recognised by deterministic HDAs. Our Myhill-Nerode construction will produce a deterministic HDA for such deterministic languages, and a non-deterministic HDA otherwise. (We make no claim as to minimality of our Myhill-Nerode HDAs.) Our definition of determinism is more subtle than for standard automata as it is not always possible to remove non-accessible parts of HDAs. We develop a language-internal characterisation of deterministic languages.

2 Pomsets with interfaces

HDAs model systems in which labelled events have duration and may happen concurrently. Every event has a time interval during which it is active: it starts at some point, then remains active until its termination and never reappears. Events may be concurrent, that is, their activity intervals may overlap; otherwise, one of the events precedes the other. We also need to consider executions in which some events are already active at the beginning (*source events*) or are still active at the end (*target events*).

At any moment of an execution we observe a list of currently active events (such lists are called *losets* below). The relative position of any two concurrent events on these lists remains the same, regardless of the point in time. This provides a secondary relation between events, which we call *event order*.

To make the above precise, let Σ be a finite alphabet. An *loset*³ $(U, \dashrightarrow, \lambda)$ is a finite set U with a total order \dashrightarrow called the *event order* and a labelling function $\lambda : U \rightarrow \Sigma$. Losets (or rather their isomorphism classes) are effectively strings but consist of concurrent, not subsequent, events.

A *labelled poset with event order (lposet)* $(P, <, \dashrightarrow, \lambda)$ consists of a finite set P with two relations: *precedence* $<$ and *event order* \dashrightarrow , together with a labelling function $\lambda : P \rightarrow \Sigma$. Note that different events may carry the same label: we do *not* exclude autoconcurrency. We require that both $<$ and \dashrightarrow are strict partial orders, that is, they are irreflexive and transitive (and thus asymmetric). We also require that for each $x \neq y$ in P , at least one of $x < y$ or $y < x$ or $x \dashrightarrow y$ or $y \dashrightarrow x$ must hold; that is, if x and y are concurrent, then they must be related by \dashrightarrow .

³ Pronunciation: ell-oh-set

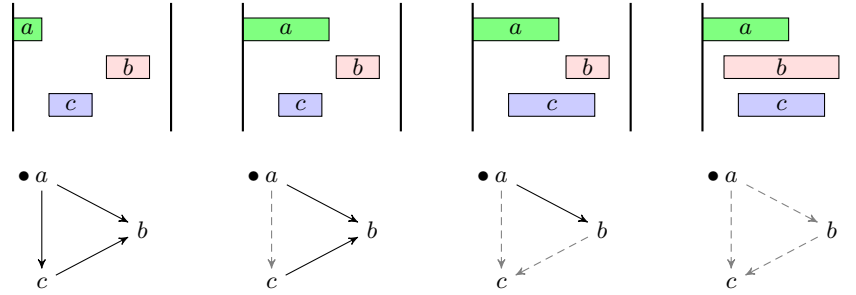


Fig. 2. Activity intervals (top) and corresponding iposets (bottom), see Example 1. Full arrows indicate precedence order; dashed arrows indicate event order; bullets indicate interfaces.

Losets may be regarded as lposets with empty precedence relation; the last condition enforces that their elements are totally ordered by $-\rightarrow$. A temporary state of an execution is described by an loset, while the whole execution provides an lposet of its events. The precedence order expresses that one event terminates before the other starts. The execution starts at the loset of $<$ -minimal elements and finishes with the loset of $<$ -maximal elements. The event order of an lposet is generated by the event orders of temporary losets. Hence any two events which are active concurrently are unrelated by $<$ but related by $-\rightarrow$.

In order to accommodate source and target events, we need to introduce lposets with interfaces (iposets). An *iposet* $(P, <, -\rightarrow, S, T, \lambda)$ consists of an lposet $(P, <, -\rightarrow, \lambda)$ together with subsets $S, T \subseteq P$ of source and target *interfaces*. Elements of S must be $<$ -minimal and those of T $<$ -maximal; hence both S and T are losets. We often denote an iposet as above by ${}_S P_T$, ignoring the orders and labelling, or use $S_P = S$ and $T_P = T$ if convenient. Source and target events will be marked by “•” at the left or right side, and if the event order is not shown, we assume that it goes downwards.

Example 1. Figure 2 shows some simple examples of activity intervals of events and the corresponding iposets. The left iposet consists of three totally ordered events, given that the intervals do not overlap; the event a is already active at the beginning and hence in the source interface. In the other iposets, the activity intervals do overlap and hence the precedence order is partial (and the event order non-trivial).

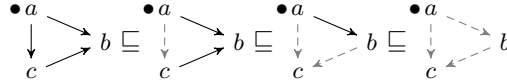
Given that the precedence relation $<$ of an iposet represents activity intervals of events, it is an *interval order* [12]. In other words, any of the iposets we will encounter admits an *interval representation*: functions b and e from P to real numbers such that $b(x) \leq e(x)$ for all $x \in P$ and $x <_P y$ iff $e(x) < b(y)$ for all $x, y \in P$. We will only consider interval iposets in this paper and omit the qualification “interval”. This is *not* a restriction, but rather induced by the semantics.

Iposets may be refined by shortening the activity intervals of events, so that some events stop being concurrent. This corresponds to expanding the precedence relation $<$ (and, potentially, removing event order). The inverse to refinement is called subsumption and defined as follows. For iposets P and Q , we say that Q *subsumes* P (or that P is a *refinement* of Q) and write $P \sqsubseteq Q$ if there exists a bijection $f : P \rightarrow Q$ (a *subsumption*) which

- respects interfaces and labels: $f(S_P) = S_Q$, $f(T_P) = T_Q$, and $\lambda_Q \circ f = \lambda_P$;
- reflects precedence: $f(x) <_Q f(y)$ implies $x <_P y$; and
- preserves essential event order: $x \dashrightarrow_P y$ implies $f(x) \dashrightarrow_Q f(y)$ whenever x and y are concurrent (that is, $x \not<_P y$ and $y \not<_P x$).

(Event order is essential for concurrent events, but by transitivity, it also appears between non-concurrent events; subsumptions may ignore such non-essential event order.)

Example 2. In Fig. 2, there is a sequence of refinements from right to left, each time shortening some activity intervals. Conversely, there is a sequence of subsumptions from left to right:



Interfaces need to be preserved across subsumptions, so in our example, the left endpoint of the a -interval must stay at the boundary.

Iposets and subsumptions form a category. The isomorphisms in that category are invertible subsumptions, and isomorphism classes of iposets are called *ipomsets*. Concretely, an isomorphism $f : P \rightarrow Q$ of iposets is a bijection which

- respects interfaces and labels: $f(S_P) = S_Q$, $f(T_P) = T_Q$, and $\lambda_Q \circ f = \lambda_P$;
- respects precedence: $x <_P y$ iff $f(x) <_Q f(y)$; and
- respects essential event order: $x \dashrightarrow_P y$ iff $f(x) \dashrightarrow_Q f(y)$ whenever $x \not<_P y$ and $y \not<_P x$.

Isomorphisms between iposets are unique (because of the requirement that all elements be ordered by $<$ or \dashrightarrow), hence we may switch freely between ipomsets and concrete representations, see [7] for details. We write $P \cong Q$ if iposets P and Q are isomorphic and let iiPoms denote the set of ipomsets.

Ipomsets may be *glued*, using a generalisation of the standard serial composition of pomsets [13]. For ipomsets P and Q , their *gluing* $P * Q$ is defined if the targets of P match the sources of Q : $T_P \cong S_Q$. In that case, its carrier set is the quotient $(P \sqcup Q)_{/x \equiv f(x)}$, where $f : T_P \rightarrow S_Q$ is the unique isomorphism, the interfaces are $S_{P*Q} = S_P$ and $T_{P*Q} = T_Q$, \dashrightarrow_{P*Q} is the transitive closure of $\dashrightarrow_P \cup \dashrightarrow_Q$, and $x <_{P*Q} y$ iff $x <_P y$, or $x <_Q y$, or $x \in P - T_P$ and $y \in Q - S_Q$. We will often omit the “*” in gluing compositions. For ipomsets with empty interfaces, $*$ is serial pomset composition; in the general case, matching interface points are glued, see [6, 8] or below for examples.

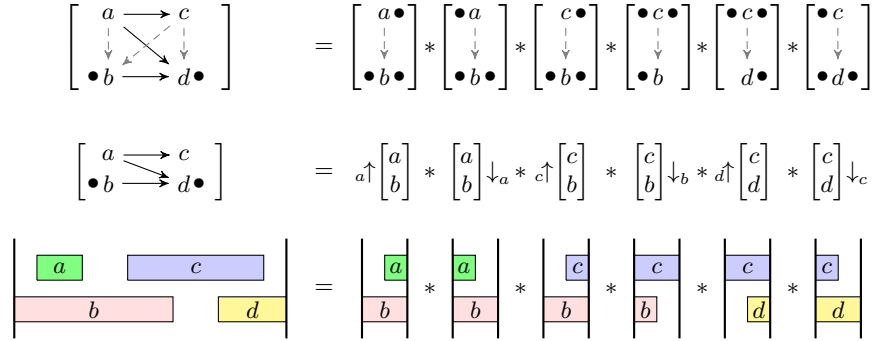


Fig. 3. Sparse decomposition of ipomset into starters and terminators.

An ipomset P is *discrete* if \langle_P is empty and \dashrightarrow_P total. Losets are discrete ipomsets with empty interfaces. Discrete ipomsets $U U_U$ are identities for gluing composition and written id_U . A *starter* is an ipomset $U \dashrightarrow_A U_U$, a *terminator* is $U U_{U \dashrightarrow A}$; these will be written $A \uparrow U$ and $U \downarrow_A$, respectively.

Any ipomset can be presented as a gluing of starters and terminators [8, Prop. 21]. (This is related to the fact that a partial order is interval iff its antichain order is total, see [12, 17, 18]). Such a presentation we call a *step decomposition*; if starters and terminators are alternating, the decomposition is *sparse*.

Example 3. Figure 3 shows a sparse decomposition of an ipomset into starters and terminators. The top line shows the graphical representation, in the middle the representation using the notation we have introduced for starters and terminators, and the bottom line shows activity intervals.

Proposition 4. *Every ipomset P has a unique sparse step decomposition.*

A *language* is, a priori, a set of ipomsets $L \subseteq \text{iiPoms}$. However, we will assume that languages are closed under refinement (inverse subsumption), so that refinements of any ipomset in L are also in L :

Definition 5. *A language is a subset $L \subseteq \text{iiPoms}$ such that $P \sqsubseteq Q$ and $Q \in L$ imply $P \in L$.*

Using interval representations, this means that languages are closed under shortening activity intervals of events. The set of all languages is denoted $\mathcal{L} \subseteq 2^{\text{iiPoms}}$.

For $X \subseteq \text{iiPoms}$ an arbitrary set of ipomsets, we denote by

$$X \downarrow = \{P \in \text{iiPoms} \mid \exists Q \in X : P \sqsubseteq Q\}$$

its downward subsumption closure, that is, the smallest language which contains X . Then

$$\mathcal{L} = \{X \subseteq \text{iiPoms} \mid X \downarrow = X\}.$$

3 HDAs and their languages

An HDA is a collection of *cells* which are connected according to specified *face maps*. Each cell has an associated list of *labelled events* which are interpreted as being executed in that cell, and the face maps may terminate some events or, inversely, indicate cells in which some of the current events were not yet started. Additionally, some cells are designated *start* cells and some others *accept* cells; computations of an HDA begin in a start cell and proceed by starting and terminating events until they reach an accept cell.

To make the above precise, let \square denote the set of losets. A *precubical set* consists of a set of cells X together with a mapping $\text{ev} : X \rightarrow \square$ which to every cell assigns its list of active events. For an loset U we write $X[U] = \{x \in X \mid \text{ev}(x) = U\}$ for the cells of type U . Further, for every $U \in \square$ and subset $A \subseteq U$ there are *face maps* $\delta_A^0, \delta_A^1 : X[U] \rightarrow X[U - A]$. The *upper* face maps δ_A^1 terminate the events in A , whereas the *lower* face maps δ_A^0 “unstart” these events: they map cells $x \in X[U]$ to cells $\delta_A^0(x) \in X[U - A]$ where the events in A are not yet active.

If $A, B \subseteq U$ are disjoint, then the order in which events in A and B are terminated or unstarted should not matter, so we require that $\delta_A^\nu \delta_B^\mu = \delta_B^\mu \delta_A^\nu$ for $\nu, \mu \in \{0, 1\}$: the *precubical identities*. A *higher-dimensional automaton (HDA)* is a precubical set together with subsets $\perp_X, \top_X \subseteq X$ of *start* and *accept* cells. For a precubical set X and subsets $Y, Z \subseteq X$ we denote by X_{\perp}^Z the HDA with precubical set X , start cells Y and accept cells Z . We do *not* generally assume that precubical sets or HDAs are finite. The *dimension* of an HDA X is $\dim(X) = \sup\{|\text{ev}(x)| \mid x \in X\} \in \mathbb{N} \cup \{\infty\}$.

Example 6. One-dimensional HDAs X are standard automata. Cells in $X[\emptyset]$ are states, cells in $X[a]$ for $a \in \Sigma$ are a -labelled transitions. Face maps δ_a^0 and δ_a^1 attach source and target states to transitions. In contrast to ordinary automata we allow start and accept *transitions* instead of merely states, so languages of such automata may contain not only words but also “words with interfaces”. In any case, at most one event is active at any point in time, so the event order is unnecessary.

Example 7. Figure 4 shows an HDA both as a combinatorial object (left) and in a more geometric realisation (right). We write isomorphism classes of losets as lists of labels and omit the set braces in $\delta_{\{a\}}^0$ etc.

An *HDA-map* between HDAs X and Y is a function $f : X \rightarrow Y$ that preserves structure: types of cells ($\text{ev}_Y \circ f = \text{ev}_X$), face maps ($f(\delta_A^\nu(x)) = \delta_A^\nu(f(x))$) and start/accept cells ($f(\perp_X) \subseteq \perp_Y, f(\top_X) \subseteq \top_Y$). Similarly, a precubical map is a function that preserves the first two of these three. HDAs and HDA-maps form a category, as do precubical sets and precubical maps.

Computations of HDAs are paths: sequences of cells connected by face maps. A *path* in X is, thus, a sequence

$$\alpha = (x_0, \varphi_1, x_1, \dots, x_{n-1}, \varphi_n, x_n), \quad (1)$$

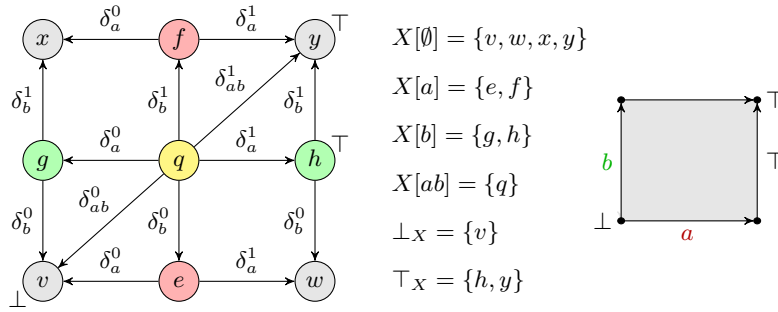


Fig. 4. A two-dimensional HDA X on $\Sigma = \{a, b\}$, see Example 7.

where the x_i are cells of X and the φ_i indicate types of face maps: for every i , $(x_{i-1}, \varphi_i, x_i)$ is either

- $(\delta_A^0(x_i), \nearrow^A, x_i)$ for $A \subseteq \text{ev}(x_i)$ (an *upstep*)
- or $(x_{i-1}, \searrow_B, \delta_B^1(x_{i-1}))$ for $B \subseteq \text{ev}(x_{i-1})$ (a *downstep*).

Upsteps start events in A while downsteps terminate events in B . The *source* and *target* of α as in (1) are $\text{src}(\alpha) = x_0$ and $\text{tgt}(\alpha) = x_n$.

The set of all paths in X starting at $Y \subseteq X$ and terminating in $Z \subseteq X$ is denoted by $\text{Path}(X)_Y^Z$; we write $\text{Path}(X)_Y = \text{Path}(X)_Y^X$, $\text{Path}(X)^Z = \text{Path}(X)_X^Z$, and $\text{Path}(X) = \text{Path}(X)_X^X$. A path α is *accepting* if $\text{src}(\alpha) \in \perp_X$ and $\text{tgt}(\alpha) \in \top_X$. Paths α and β may be concatenated if $\text{tgt}(\alpha) = \text{src}(\beta)$; their concatenation is written $\alpha * \beta$, and we omit the “*” in concatenations if convenient.

Path equivalence is the congruence \simeq generated by $(z \nearrow^A y \nearrow^B x) \simeq (z \nearrow^{A \cup B} x)$, $(x \searrow_A y \searrow_B z) \simeq (x \searrow_{A \cup B} z)$, and $\gamma \alpha \delta \simeq \gamma \beta \delta$ whenever $\alpha \simeq \beta$. Intuitively, this relation allows to assemble subsequent upsteps or downsteps into one “bigger” step. A path is *sparse* if its upsteps and downsteps are alternating, so that no more such assembling may take place. Every equivalence class of paths contains a unique sparse path.

Example 8. Paths in one-dimensional HDAs are standard paths, *i.e.*, sequences of transitions connected at states. Path equivalence is a trivial relation, and all paths are sparse.

Example 9. The HDA X of Fig. 4 admits five sparse accepting paths:

$$\begin{aligned}
 & v \nearrow^a e \searrow_a w \nearrow^b h, & v \nearrow^a e \searrow_a w \nearrow^b h \searrow_b y, \\
 & v \nearrow^{ab} q \searrow_a h, & v \nearrow^{ab} q \searrow_{ab} y, & v \nearrow^b g \searrow_b x \nearrow^a f \searrow_a y.
 \end{aligned}$$

The observable content or *event ipomset* $\text{ev}(\alpha)$ of a path α is defined recursively as follows:

- If $\alpha = (x)$, then $\text{ev}(\alpha) = \text{id}_{\text{ev}(x)}$.
- If $\alpha = (y \nearrow^A x)$, then $\text{ev}(\alpha) = A \uparrow \text{ev}(x)$.

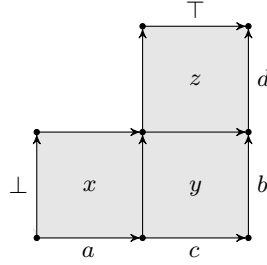


Fig. 5. HDA Y consisting of three squares glued along common faces.

- If $\alpha = (x \searrow_B y)$, then $\text{ev}(\alpha) = \text{ev}(x) \downarrow_B$.
- If $\alpha = \alpha_1 * \dots * \alpha_n$ is a concatenation, then $\text{ev}(\alpha) = \text{ev}(\alpha_1) * \dots * \text{ev}(\alpha_n)$.

[7, Lemma 8] shows that $\alpha \simeq \beta$ implies $\text{ev}(\alpha) = \text{ev}(\beta)$. Further, if $\alpha = \alpha_1 * \dots * \alpha_n$ is a sparse path, then $\text{ev}(\alpha) = \text{ev}(\alpha_1) * \dots * \text{ev}(\alpha_n)$ is a sparse step decomposition.

Example 10. Event ipomsets of paths in one-dimensional HDAs are words, possibly with interfaces. Sparse step decompositions of words are obtained by splitting symbols into starts and terminations, for example, $\bullet ab = \bullet a * b \bullet * \bullet b$.

Example 11. The event ipomsets of the five sparse accepting paths in the HDA X of Fig. 4 are $ab \bullet$, ab , $\begin{bmatrix} a \\ b \bullet \end{bmatrix}$, $\begin{bmatrix} a \\ b \end{bmatrix}$, and ba . Figure 5 shows another HDA which admits an accepting path $(\delta_a^0 x \nearrow^a x \searrow_a \delta_a^1 x \nearrow^c y \searrow_b \delta_b^1 y \nearrow^d z \searrow_d \delta_d^1 z)$. Its event ipomset is precisely the ipomset of Fig. 3, with the indicated sparse step decomposition arising directly from the sparse presentation above.

The *language* of an HDA X is

$$\text{Lang}(X) = \{\text{ev}(\alpha) \mid \alpha \text{ accepting path in } X\}.$$

[7, Prop. 10] shows that languages of HDAs are sets of ipomsets which are closed under subsumption, *i.e.*, languages in the sense of Def. 5.

A language is *regular* if it is the language of a finite HDA.

Example 12. The languages of our example HDAs are $\text{Lang}(X) = \{\begin{bmatrix} a \\ b \bullet \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix}\} \downarrow = \{\begin{bmatrix} a \\ b \bullet \end{bmatrix}, ab \bullet, \begin{bmatrix} a \\ b \end{bmatrix}, ab, ba\}$ and

$$\text{Lang}(Y) = \left\{ \left[\begin{array}{ccc} a & \rightarrow & c \bullet \\ & \searrow & \\ \bullet b & \rightarrow & d \end{array} \right] \right\} \downarrow.$$

We say that a cell $x \in X$ in an HDA X is

- *accessible* if $\text{Path}(X)_{\perp}^x \neq \emptyset$, *i.e.*, x can be reached by a path from a start cell;
- *coaccessible* if $\text{Path}(X)_x^{\top} \neq \emptyset$, *i.e.*, there is a path from x to an accept cell;
- *essential* if it is both accessible and coaccessible.

A path is *essential* if its source and target cells are essential. This implies that all its cells are essential. Segments of accepting paths are always essential.

The set of essential cells of X is denoted by $\text{ess}(X)$; this is not necessarily a sub-HDA of X given that faces of essential cells may be non-essential. For example, all bottom cells of the HDA Y in Fig. 5 are inaccessible and hence non-essential.

Lemma 13. *Let X be an HDA. There exists a smallest sub-HDA $X^{\text{ess}} \subseteq X$ that contains all essential cells, and $\text{Lang}(X^{\text{ess}}) = \text{Lang}(X)$. If $\text{ess}(X)$ is finite, then X^{ess} is also finite.*

Proof. The set of all faces of essential cells

$$X^{\text{ess}} = \{\delta_A^0 \delta_B^1(x) \mid x \in \text{ess}(X), A, B \subseteq \text{ev}(x), A \cap B = \emptyset\}$$

is a sub-HDA of X . Clearly every sub-HDA of X that contains $\text{ess}(X)$ must also contain X^{ess} . Since all accepting paths are essential, $\text{Lang}(X^{\text{ess}}) = \text{Lang}(X)$. If $|\text{ess}(X)| = n$ and $|\text{ev}(x)| \leq d$ for all $x \in \text{ess}(X)$, then $|X^{\text{ess}}| \leq n \cdot 3^d$. \square

Track objects, introduced in [6], provide a mapping from ipomsets to HDAs and are a powerful tool for reasoning about languages. We only need some of their properties in proofs, so we do not give a definition here but instead refer to [6, Sect. 5.3]. Let \square^P denote the track object of an ipomset P ; this is an HDA with one start cell c_{\perp}^P and one accept cell c_P^{\top} . Below we list properties of track objects needed in the paper.

Lemma 14. *Let X be an HDA, $x, y \in X$ and $P \in \text{iiPoms}$. The following conditions are equivalent:*

1. *There exists a path $\alpha \in \text{Path}(X)_x^y$ such that $\text{ev}(\alpha) = P$.*
2. *There is an HDA-map $f : \square^P \rightarrow X_x^y$ (i.e., $f(c_{\perp}^P) = x$ and $f(c_P^{\top}) = y$).*

Proof. This is an immediate consequence of [6, Prop. 89]. \square

Lemma 15. *Let X be an HDA, $x, y \in X$ and $\gamma \in \text{Path}(X)_x^y$. Assume that $\text{ev}(\gamma) = P * Q$ for ipomsets P and Q . Then there exist paths $\alpha \in \text{Path}(X)_x$ and $\beta \in \text{Path}(X)_y$ such that $\text{ev}(\alpha) = P$, $\text{ev}(\beta) = Q$ and $\text{tgt}(\alpha) = \text{src}(\beta)$.*

Proof. By Lemma 14, there is an HDA-map $f : \square^{PQ} \rightarrow X_x^y$. By [6, Lem. 65], there exist precubical maps $j_P : \square^P \rightarrow \square^{PQ}$, $j_Q : \square^Q \rightarrow \square^{PQ}$ such that $j_P(c_{\perp}^P) = c_{\perp}^{PQ}$, $j_P(c_P^{\top}) = j_Q(c_{\perp}^Q)$ and $j_Q(c_Q^{\top}) = c_P^{\top}$. Let $z = f(j_P(c_{\perp}^P))$, then $f \circ j_P : \square^P \rightarrow X_x^z$ and $f \circ j_Q : \square^Q \rightarrow X_z^y$ are HDA-maps, and by applying Lemma 14 again to j_P and j_Q we obtain α and β . \square

4 Myhill-Nerode theorem

The *prefix quotient* of a language $L \in \mathcal{L}$ by an ipomset P is the language

$$P \setminus L = \{Q \in \text{iiPoms} \mid PQ \in L\}.$$

Similarly, the *suffix quotient* of L by P is $L/P = \{Q \in \text{iiPoms} \mid QP \in L\}$. Denote

$$\text{suff}(L) = \{P \setminus L \mid P \in \text{iiPoms}\}, \quad \text{pref}(L) = \{L/P \mid P \in \text{iiPoms}\}.$$

We record the following property of quotient languages.

Lemma 16. *If L is a language and $P \sqsubseteq Q$, then $Q \setminus L \subseteq P \setminus L$.*

Proof. If $P \sqsubseteq Q$, then $PR \sqsubseteq QR$. Thus,

$$R \in Q \setminus L \iff QR \in L \implies PR \in L \iff R \in P \setminus L. \quad \square$$

The main goal of this section is to show the following.

Theorem 17. *For a language $L \in \mathcal{L}$ the following conditions are equivalent.*

- (a) L is regular.
- (b) The set $\text{suff}(L) \subseteq \mathcal{L}$ is finite.
- (c) The set $\text{pref}(L) \subseteq \mathcal{L}$ is finite.

We prove only the equivalence between (a) and (b); equivalence between (a) and (c) is symmetric. First we prove the implication (a) \implies (b). Let X be an HDA with $\text{Lang}(X) = L$. For $x \in X$ define languages $\text{Pre}(x) = \text{Lang}(X_{\perp}^x)$ and $\text{Post}(x) = \text{Lang}(X_x^{\top})$.

Lemma 18. *For every $P \in \text{iiPoms}$, $P \setminus L = \bigcup \{\text{Post}(x) \mid x \in X, P \in \text{Pre}(x)\}$.*

Proof. We have

$$\begin{aligned} Q \in P \setminus L &\iff PQ \in L \iff \exists f : \square^{PQ} \rightarrow X = X_{\perp}^{\top} \\ &\iff \exists x \in X, g : \square^P \rightarrow X_{\perp}^x, h : \square^Q \rightarrow X_x^{\top} \\ &\iff \exists x \in X : P \in \text{Lang}(X_{\perp}^x), Q \in \text{Lang}(X_x^{\top}) \\ &\iff \exists x \in X : P \in \text{Pre}(x), Q \in \text{Post}(x). \end{aligned}$$

The last condition says that Q belongs to the right-hand side of the equation. \square

Proof of Thm. 17, (a) \implies (b). The family of languages $\{P \setminus L \mid P \in \text{iiPoms}\}$ is a subfamily of $\{\bigcup_{x \in Y} \text{Post}(x) \mid Y \subseteq X\}$ which is finite. \square

HDA construction. Now we show that (b) implies (a). Fix a language $L \in \mathcal{L}$, with $\text{suff}(L)$ finite or infinite. We will construct an HDA $\text{MN}(L)$ that recognises L and show that if $\text{suff}(L)$ is finite, then its essential part $\text{MN}(L)^{\text{ess}}$ is finite. The cells of $\text{MN}(L)$ are equivalence classes of ipomsets under a relation \approx_L induced by L which we will introduce below. The relation \approx_L is defined using prefix quotients, but needs to be stronger than prefix quotient equivalence. This is because events may be concurrent and because ipomsets have interfaces. We give examples just after the construction.

For an ipomset ${}_S P_T$ define its (*target*) *signature* to be the starter $\text{fin}(P) = T - S \uparrow T$. Thus $\text{fin}(P)$ collects all target events of P , and its source interface contains those events that are also in the source interface of P . We also write $\text{rfin}(P) = T - S \subseteq \text{fin}(P)$: the set of all target events of P that are not source events. An important property is that removing elements of $\text{rfin}(P)$ does not change the source interface of P . For example,

$$\text{fin}\left(\begin{bmatrix} \bullet^a \bullet \\ \bullet^a \bullet \\ \bullet^c \bullet \end{bmatrix}\right) = [\bullet^a \bullet], \quad \text{fin}\left(\begin{bmatrix} \bullet^a \bullet \\ \bullet^c \bullet \\ \bullet^c \bullet \end{bmatrix}\right) = [\bullet^c \bullet], \quad \text{fin}\left(\begin{bmatrix} \bullet^a \bullet \\ \bullet^c \bullet \\ \bullet^c \bullet \end{bmatrix}\right) = [\bullet^c \bullet];$$

rfin is $\{c\}$ in the first two examples and equal to $\{b\}$ in the last.

We define two equivalence relations on iiPoms induced by L :

- Ipomsets P and Q are *weakly equivalent* ($P \sim_L Q$) if $\text{fin}(P) \cong \text{fin}(Q)$ and $P \setminus L = Q \setminus L$. Obviously, $P \sim_L Q$ implies $T_P \cong T_Q$ and $\text{rfin}(P) \cong \text{rfin}(Q)$.
- Ipomsets P and Q are *strongly equivalent* ($P \approx_L Q$) if $P \sim_L Q$ and for all $A \subseteq \text{rfin}(P) \cong \text{rfin}(Q)$ we have $(P - A) \setminus L = (Q - A) \setminus L$.

Evidently $P \approx_L Q$ implies $P \sim_L Q$, but the inverse does not always hold. We explain in Example 21 below why \approx_L , and not \sim_L , is the proper relation to use for constructing $\text{MN}(L)$.

Lemma 19. *If $P \approx_L Q$, then $P - A \approx_L Q - A$ for all $A \subseteq \text{rfin}(P) \cong \text{rfin}(Q)$.*

Proof. For every A we have $(P - A) \setminus L = (Q - A) \setminus L$, and

$$\text{fin}(P - A) = \text{fin}(P) - A \cong \text{fin}(Q) - A = \text{fin}(Q - A),$$

Thus, $P - A \sim_L Q - A$. Further, for every $B \subseteq \text{rfin}(P - A) \cong \text{rfin}(Q - A)$,

$$((P - A) - B) \setminus L = (P - (A \cup B)) \setminus L = (Q - (A \cup B)) \setminus L = ((Q - A) - B) \setminus L,$$

which shows that $P - A \approx_L Q - A$. \square

Now define an HDA $\text{MN}(L)$ as follows. For $U \in \square$,

$$\text{MN}(L)[U] = (\text{iiPoms}_U / \approx_L) \cup \{w_U\},$$

where the w_U are new *subsidiary* cells which are introduced solely to define some lower faces. (They will not affect the language of $\text{MN}(L)$).

The \approx_L -equivalence class of P will be denoted by $\langle P \rangle$ (but often just by P in examples). Face maps are defined as follows, for $A \subseteq U \in \square$ and $P \in \text{iiPoms}_U$:

$$\delta_A^0(\langle P \rangle) = \begin{cases} \langle P - A \rangle & \text{if } A \subseteq \text{rfin}(P), \\ w_{U-A} & \text{otherwise,} \end{cases} \quad \delta_A^1(\langle P \rangle) = \langle P * U \downarrow_A \rangle, \quad (2)$$

$$\delta_A^0(w_U) = \delta_A^1(w_U) = w_{U-A}.$$

In other words, if A has no source events of P , then δ_A^0 removes A from P (the source interface of P is unchanged). If A contains any source event, then $\delta_A^0(P)$ is a subsidiary cell.

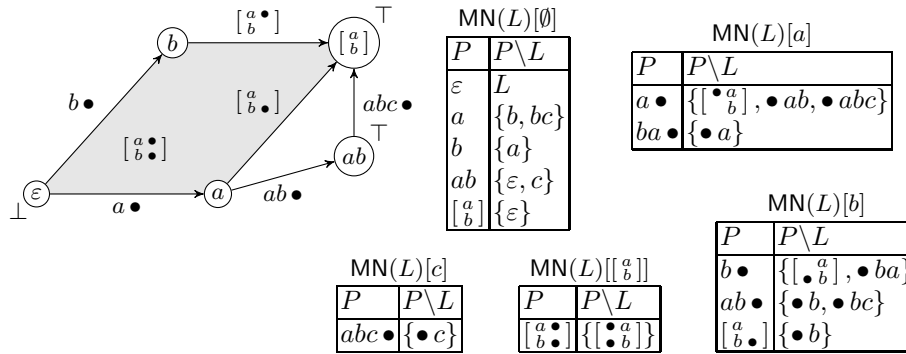


Fig. 6. HDA $MN(L)$ of Example 20, showing names of cells instead of labels (labels are target interfaces of names). Tables show essential cells together with prefix quotients.

Finally, start and accept cells are given by

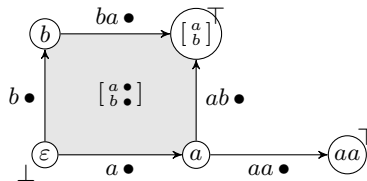
$$\perp_{MN(L)} = \{\langle id_U \rangle\}_{U \in \square}, \quad \top_{MN(L)} = \{\langle P \rangle \mid P \in L\}.$$

The cells $\langle P \rangle$ will be called *regular*. They are \approx_L -equivalence classes of ipomsets, lower face maps unstart events, and upper face maps terminate events. All faces of subsidiary cells w_U are subsidiary, and upper faces of regular cells are regular. Below we present several examples, in which we show only the essential part $MN(L)^{ess}$ of $MN(L)$.

Example 20. Let $L = \{[a/b], abc\} \downarrow = \{[a/b], ab, ba, abc\}$. Figure 6 shows the HDA $MN(L)^{ess}$ together with a list of essential cells of $M(L)$ and their prefix quotients in L . Note that the state $\langle a \rangle$ has *two* outgoing b -labelled edges: $\langle ab \bullet \rangle$ and $\langle [a/b] \bullet \rangle$. The generating ipomsets have different prefix quotients because of $\{[a/b], abc\} \subseteq L$ but the same lower face $\langle a \rangle$.

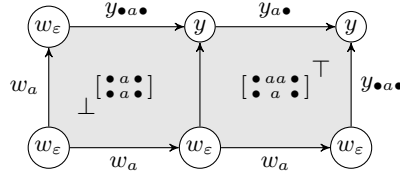
Intuitively, $MN(L)^{ess}$ is thus *non-deterministic*; this is interesting because the standard Myhill-Nerode theorem for finite automata constructs deterministic automata. We will give a precise definition of determinism for HDAs in the next section and show in Example 42 that no deterministic HDA X exists with $\text{Lang}(X) = L$.

Example 21. Here we explain why we need to use \approx_L -equivalence classes and not \sim_L -equivalence classes. Let $L = \{[a/b], aa\} \downarrow$. Then $MN(L)^{ess}$ is as below.



Note that $(aa\bullet)\setminus L = (ba\bullet)\setminus L = \{\bullet a\}$, thus $aa\bullet \sim_L ba\bullet$. Yet $aa\bullet$ and $ba\bullet$ are not strongly equivalent, because $a\setminus L = \{a, b\} \neq \{a\} = b\setminus L$. This provides an example of weakly equivalent ipomsets whose lower faces are not weakly equivalent and shows why we cannot use \sim_L to construct $\text{MN}(L)$.

Example 22. The language $L = \{[\begin{smallmatrix} \bullet & aa & \bullet \\ & a & \bullet \end{smallmatrix}]\}$ is recognised by the HDA $\text{MN}(L)^{\text{ess}}$ below:



Cells with the same names are identified. Here we see subsidiary cells w_ε and w_a , and regular cells (denoted by y indexed with their signature) that are not coaccessible. The middle vertical edge is $\langle [\begin{smallmatrix} \bullet & a & \bullet \\ & a & \bullet \end{smallmatrix}] \rangle$, $y_{\bullet a \bullet} = \langle [\begin{smallmatrix} \bullet & a & \bullet \\ & a & \bullet \end{smallmatrix}] \rangle = \langle [\begin{smallmatrix} \bullet & aa & \bullet \\ & a & \bullet \end{smallmatrix}] \rangle$, $y_{a\bullet} = \langle [\begin{smallmatrix} \bullet & aa & \bullet \\ & a & \bullet \end{smallmatrix}] \rangle$, and $y = \langle [\begin{smallmatrix} \bullet & a & \bullet \\ & a & \bullet \end{smallmatrix}] \rangle$.

MN(L) is well-defined. We need to show that the formulas (2) do not depend on the choice of a representative in $\langle P \rangle$ and that the precubical identities are satisfied.

Lemma 23. *Let P, Q and R be ipomsets with $T_P = T_Q = S_R$. Then*

$$P\setminus L \subseteq Q\setminus L \implies (PR)\setminus L \subseteq (QR)\setminus L.$$

In particular, $P\setminus L = Q\setminus L$ implies $(PR)\setminus L = (QR)\setminus L$.

Proof. For $N \in \text{iiPoms}$ we have

$$\begin{aligned} N \in (PR)\setminus L &\iff PRN \in L \iff RN \in P\setminus L \\ &\implies RN \in Q\setminus L \iff QRN \in L \iff N \in (QR)\setminus L. \quad \square \end{aligned}$$

The next lemma shows an operation to “add order” to an ipomset P . This is done by first removing some points $A \subseteq T_P$ and then adding them back in, forcing arrows from all other points in P . The result is obviously subsumed by P .

Lemma 24. *For $P \in \text{iiPoms}$ and $A \subseteq \text{rfin}(P)$, $(P - A) * {}_A\uparrow T_P \sqsubseteq P$. \square*

The next two lemmas, whose proofs are again obvious, state that events may be unstarted or terminated in any order.

Lemma 25. *Let U be an loset and $A, B \subseteq U$ disjoint subsets. Then*

$$U\downarrow_B * (U - B)\downarrow_A = U\downarrow_{A \cup B} = U\downarrow_A * (U - A)\downarrow_B. \quad \square$$

Lemma 26. *Let $P \in \text{iiPoms}$ and $A, B \subseteq T_P$ disjoint subsets. Then*

$$(P * T_P\downarrow_B) - A = (P - A) * (T_P - A)\downarrow_B. \quad \square$$

Lemma 27. *Assume that $P \approx_L Q$ for $P, Q \in \text{iiPoms}_U$. Then $P * U \downarrow_B \approx_L Q * U \downarrow_B$ for every $B \subseteq U$.*

Proof. Obviously $\text{fin}(P * U \downarrow_B) = \text{fin}(P) - B \cong \text{fin}(Q) - B = \text{fin}(Q * U \downarrow_B)$. For every $A \subseteq \text{rfin}(P) - B \simeq \text{rfin}(Q) - B$ we have

$$((P - A) * (U - A) \downarrow_B) \setminus L = ((Q - A) * (U - A) \downarrow_B) \setminus L$$

by assumption and Lemma 23. But $(P * U \downarrow_B) - A = (P - A) * (U - A) \downarrow_B$ and $(Q * U \downarrow_B) - A = (Q - A) * (U - A) \downarrow_B$ by Lemma 26. \square

Proposition 28. *$\text{MN}(L)$ is a well-defined HDA.*

Proof. The face maps are well-defined: for δ_A^0 this follows from Lemma 19, for δ_B^1 from Lemma 27. The precubical identities $\delta_A^\nu \delta_B^\mu = \delta_B^\mu \delta_A^\nu$ are clear for $\nu = \mu = 0$, follow from Lemma 25 for $\nu = \mu = 1$, and from Lemma 26 for $\{\nu, \mu\} = \{0, 1\}$. \square

Paths and essential cells of $\text{MN}(L)$. The next lemma provides paths in $\text{MN}(L)$.

Lemma 29. *For every $N, P \in \text{iiPoms}$ such that $T_N \cong S_P$ there exists a path $\alpha \in \text{Path}(\text{MN}(L))_{\langle N \rangle}^{\langle NP \rangle}$ such that $\text{ev}(\alpha) = P$.*

Proof. Choose a decomposition $P = Q_1 * \dots * Q_n$ into starters and terminators. Denote $U_k = T_{Q_k} = S_{Q_{k+1}}$ and define

$$x_k = \langle N * Q_1 * \dots * Q_k \rangle, \quad \varphi_k = \begin{cases} d_A^0 & \text{if } Q_k = A \uparrow U_k, \\ d_B^1 & \text{if } Q_k = U_{k-1} \downarrow_B \end{cases}$$

for $k = 1, \dots, n$. If $\varphi_k = d_A^0$ and $Q_k = A \uparrow U_k$, then

$$\begin{aligned} \delta_A^0(x_k) &= \langle N * Q_1 * \dots * Q_{k-1} * A \uparrow U_k - A \rangle \\ &= \langle N * Q_1 * \dots * Q_{k-1} * \text{id}_{U_k - A} \rangle = x_{k-1}. \end{aligned}$$

If $\varphi_k = d_B^1$ and $Q_k = U_{k-1} \downarrow_B$, then

$$\delta_B^1(x_{k-1}) = \langle N * Q_1 * \dots * Q_{k-1} * U_{k-1} \downarrow_B \rangle = x_k.$$

Thus, $\alpha = (x_0, \varphi_1, x_1, \dots, \varphi_n, x_n)$ is a path with $\text{ev}(\alpha) = P$, $\text{src}(\alpha) = \langle N \rangle$ and $\text{tgt}(\alpha) = \langle N * P \rangle$. \square

Our goal is now to describe essential cells of $\text{MN}(L)$.

Lemma 30. *All regular cells of $\text{MN}(L)$ are accessible. If $P \setminus L \neq \emptyset$, then $\langle P \rangle$ is coaccessible.*

Proof. Both claims follow from Lemma 29. For every P there exists a path from $\langle \text{id}_{S_P} \rangle$ to $\langle \text{id}_{S_P} * P \rangle = \langle P \rangle$. If $Q \in P \setminus L$, then there exists a path $\alpha \in \text{Path}(\text{MN}(L))_{\langle P \rangle}^{\langle PQ \rangle}$, and $PQ \in L$ entails that $\langle PQ \rangle \in \top_{\text{MN}(L)}$. \square

Lemma 31. *Subsidiary cells of $\text{MN}(L)$ are not accessible. If $P \setminus L = \emptyset$, then $\langle P \rangle$ is not coaccessible.*

Proof. If $\alpha \in \text{Path}(\text{MN}(L))_{\perp}^{wU}$, then it contains a step β from a regular cell to a subsidiary cell (since all start cells are regular). Yet β can be neither an upstep (since lower faces of subsidiary cells are subsidiary) nor a downstep (since upper faces of regular cells are regular). This contradiction proves the first claim.

To prove the second part we use a similar argument. If $P \setminus L = \emptyset$, then a path $\alpha \in \text{Path}(\text{MN}(L))_{\langle P \rangle}^{\top}$ contains only regular cells (as shown above). Given that $R \setminus L \neq \emptyset$ for all $\langle R \rangle \in \top_{\text{MN}(L)}$, α must contain a step β from $\langle Q \rangle$ to $\langle R \rangle$ such that $Q \setminus L = \emptyset$ and $R \setminus L \neq \emptyset$. If β is a downstep, i.e., $\beta = (\langle Q \rangle \searrow_A \langle Q * U \downarrow_A \rangle)$, and $N \in R \setminus L = (Q * U \downarrow_A) \setminus L$, then $U \downarrow_A * N \in Q \setminus L \neq \emptyset$: a contradiction. If $\beta = (\langle R - A \rangle \nearrow^A \langle R \rangle)$ is an upstep and $N \in R \setminus L$, then, by Lemma 24,

$$(R - A) * {}_A \uparrow U * N \subseteq R * N \in L,$$

implying that $Q \setminus L = (R - A) \setminus L \neq \emptyset$ by Lemma 16: another contradiction. \square

Lemmas 30 and 31 together immediately imply the following.

Proposition 32. $\text{ess}(\text{MN}(L)) = \{\langle P \rangle \mid P \setminus L \neq \emptyset\}$. \square

MN(L) recognises L. One inclusion follows immediately from Lemma 29:

Lemma 33. $L \subseteq \text{Lang}(\text{MN}(L))$.

Proof. For every $P \in \text{iiPoms}$ there exists a path $\alpha \in \text{Path}(\text{MN}(L))_{\langle \text{id}_{S_P} \rangle}^{\langle P \rangle}$ such that $\text{ev}(\alpha) = P$. If $P \in L$, then $\varepsilon \in P \setminus L$, i.e., $\langle P \rangle$ is an accept cell. Thus α is accepting and $P = \text{ev}(\alpha) \in \text{Lang}(\text{MN}(L))$. \square

The converse inclusion requires more work. For a regular cell $\langle P \rangle$ of $\text{MN}(L)$ denote $\langle P \rangle \setminus L = P \setminus L$ (this obviously does not depend on the choice of P).

Lemma 34. *If $S \in \square$ and $\alpha \in \text{Path}(\text{MN}(L))_{\langle \text{id}_S \rangle}$, then $\text{tgt}(\alpha) \setminus L \subseteq \text{ev}(\alpha) \setminus L$.*

Proof. By Lemma 31, all cells appearing along α are regular. We proceed by induction on the length of α . For $\alpha = (\langle \text{id}_S \rangle)$ the claim is obvious. If α is non-trivial, we have two cases.

- $\alpha = \beta * (\delta_A^0(\langle P \rangle) \nearrow^A \langle P \rangle)$, where $\langle P \rangle \in \text{MN}(L)[U]$ and $A \subseteq \text{rfin}(P) \subseteq U \cong T_P$. By the induction hypothesis,

$$(P - A) \setminus L = \delta_A^0(\langle P \rangle) \setminus L = \text{tgt}(\beta) \setminus L \subseteq \text{ev}(\beta) \setminus L.$$

For $Q \in \text{iiPoms}$ we have

$$\begin{aligned} Q \in P \setminus L &\iff PQ \in L \implies (P - A) * {}_A \uparrow U * Q \in L && \text{(Lemma 24)} \\ &\iff {}_A \uparrow U * Q \in (P - A) \setminus L \\ &\implies {}_A \uparrow U * Q \in \text{ev}(\beta) \setminus L && \text{(induction hypothesis)} \\ &\iff \text{ev}(\beta) * {}_A \uparrow U * Q \in L \\ &\iff \text{ev}(\alpha) * Q \in L \iff Q \in \text{ev}(\alpha) \setminus L. \end{aligned}$$

Thus, $\langle P \rangle \setminus L = P \setminus L \subseteq \text{ev}(\alpha) \setminus L$.

- $\alpha = \beta * (\langle P \rangle \searrow_B \delta_B^1(\langle P \rangle))$, where $\langle P \rangle \in \mathbf{MN}(L)[U]$ and $B \subseteq U \cong T_P$. By inductive assumption, $P \setminus L = \mathbf{tgt}(\beta) \setminus L \subseteq \mathbf{ev}(\beta) \setminus L$. Thus,

$$\mathbf{tgt}(\alpha) \setminus L = \delta_B^1(\langle P \rangle) \setminus L = \langle P * U \downarrow_B \rangle \setminus L \subseteq (\mathbf{ev}(\beta) * U \downarrow_B) \setminus L = \mathbf{ev}(\alpha) \setminus L.$$

The inclusion above follows from Lemma 23. \square

Proposition 35. $\mathbf{Lang}(\mathbf{MN}(L)) = L$.

Proof. The inclusion $L \subseteq \mathbf{Lang}(\mathbf{MN}(L))$ is shown in Lemma 33. For the converse, let $S \in \square$ and $\alpha \in \mathbf{Path}(\mathbf{MN}(L))_{(\text{id}_S)}$, then Lemma 34 implies

$$\mathbf{tgt}(\alpha) \in \top_{\mathbf{MN}(L)} \iff \varepsilon \in \mathbf{tgt}(\alpha) \setminus L \implies \varepsilon \in \mathbf{ev}(\alpha) \setminus L \iff \mathbf{ev}(\alpha) \in L,$$

that is, if α is accepting, then $\mathbf{ev}(\alpha) \in L$. \square

Finiteness of $\mathbf{MN}(L)$. The HDA $\mathbf{MN}(L)$ is not finite, since it contains infinitely many subsidiary cells w_U . Below we show that its essential part $\mathbf{MN}(L)^{\text{ess}}$ is finite if L has finitely many prefix quotients.

Lemma 36. *If $\text{suff}(L)$ is finite, then $\text{ess}(\mathbf{MN}(L))$ is finite.*

Proof. For $\langle P \rangle, \langle Q \rangle \in \text{ess}(L)$, we have $\langle P \rangle \approx_L \langle Q \rangle$ iff $f(\langle P \rangle) = f(\langle Q \rangle)$, where

$$f(\langle P \rangle) = (P \setminus L, \text{fin}(P), ((P - A) \setminus L)_{A \subseteq \text{rfin}(P)}).$$

We will show that f takes only finitely many values on $\text{ess}(L)$. Indeed, $P \setminus L$ belongs to the finite set $\text{suff}(L)$. Further, all ipomsets in $P \setminus L$ have source interfaces equal to T_P . Since $P \setminus L$ is non-empty, $\text{fin}(P)$ is a starter with T_P as underlying loset. Yet, there are only finitely many starters on any loset. The last coordinate also may take only finitely many values, since $\text{rfin}(P)$ is finite and $(P - A) \setminus L \in \text{suff}(L)$. \square

Proof of Thm. 17, (b) \implies (a). From Lemma 36 and Lemma 13, $\mathbf{MN}(L)^{\text{ess}}$ is a finite HDA. By Prop. 35 we have $\mathbf{Lang}(\mathbf{MN}(L)^{\text{ess}}) = \mathbf{Lang}(\mathbf{MN}(L)) = L$. \square

Example 37. We finish this section with another example, which shows some subtleties related to higher-dimensional loops. Let L be the language of the HDA shown to the left of Fig. 7 (a looping version of the HDA of Fig. 5), then

$$L = \{\bullet a \bullet\} \cup \{[\bullet \begin{smallmatrix} aa \\ b \end{smallmatrix} \bullet]^n \mid n \geq 1\} \downarrow.$$

Our construction yields $\mathbf{MN}(L)^{\text{ess}}$ as shown on the right of the figure. Here, $e = \langle [\bullet \begin{smallmatrix} a \\ b \end{smallmatrix} \bullet] \rangle$, and cells with the same names are identified. These identifications follow from the fact that $[\bullet \begin{smallmatrix} aa \\ bb \end{smallmatrix} \bullet] \approx_L [\bullet \begin{smallmatrix} a \\ b \end{smallmatrix} \bullet]$, $[\bullet \begin{smallmatrix} aa \\ bb \end{smallmatrix}] \approx_L [\bullet \begin{smallmatrix} a \\ b \end{smallmatrix}]$, and $[\bullet \begin{smallmatrix} aa \\ b \end{smallmatrix}] \approx_L \bullet a$. Note that $[\bullet \begin{smallmatrix} a \\ b \end{smallmatrix} \bullet]$ and $[\bullet \begin{smallmatrix} aa \\ bb \end{smallmatrix} \bullet]$ are not strongly equivalent, since they have different signatures: $[\bullet \begin{smallmatrix} a \\ b \end{smallmatrix} \bullet]$ and $[\begin{smallmatrix} a \\ b \end{smallmatrix} \bullet]$, respectively.

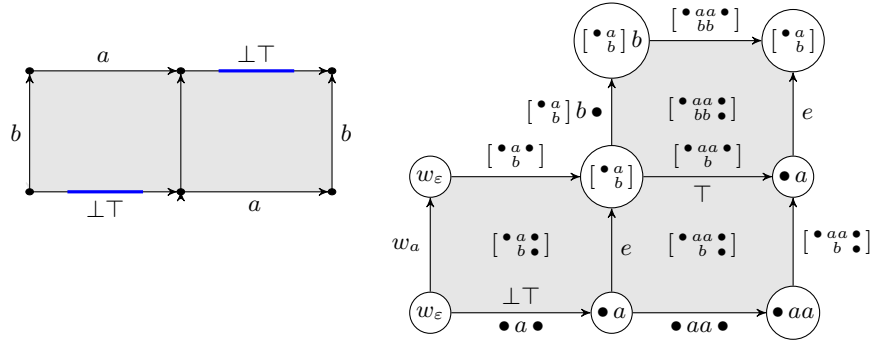


Fig. 7. Two HDAs recognising the language of Example 37. On the left side, start/accept edges are identified.

5 Determinism

We now make precise our notion of determinism and show that not all HDAs may be determinised. Recall that we do not assume finiteness.

Definition 38. An HDA X is deterministic if

1. for every $U \in \square$ there is at most one initial cell in $X[U]$, and
2. for all $V \in \square$, $A \subseteq V$ and an essential cell $x \in X[V - A]$ there exists at most one essential cell $y \in X[V]$ such that $\delta_A^0(y) = x$.

That is, in any essential cell x in a deterministic HDA X and for any set A of events, there is at most one way to start A in x and remain in the essential part of X . We allow multiple initial cells because ipomsets in $\text{Lang}(X)$ may have different source interfaces; for each source interface in $\text{Lang}(X)$, there can be at most one matching start cell in X . Note that we must restrict our definition to essential cells as inessential cells may not always be removed (in contrast to the case of standard automata).

A language is deterministic if it is recognised by a deterministic HDA. We develop a language-internal criterion for being deterministic.

Definition 39. A language L is swap-invariant if it holds for all $P, Q, P', Q' \in \text{iiPoms}$ that $PP' \in L$, $QQ' \in L$ and $P \sqsubseteq Q$ imply $QP' \in L$.

That is, if the P prefix of $PP' \in L$ is subsumed by Q (which is, thus, “more concurrent” than P), and if Q itself may be extended to an ipomset in L , then P may be swapped for Q in the ipomset PP' to yield $QP' \in L$.

Lemma 40. L is swap-invariant iff $P \sqsubseteq Q$ implies $P \setminus L = Q \setminus L$ for all $P, Q \in \text{iiPoms}$, unless $Q \setminus L = \emptyset$.

Proof. Assume that L is swap-invariant and let $P \sqsubseteq Q$. The inclusion $Q \setminus L \subseteq P \setminus L$ follows from Lemma 16, and

$$R \in Q \setminus L, R' \in P \setminus L \iff QR, PR' \in L \implies QR' \in L \iff R' \in Q \setminus L$$

implies that $P \setminus L \subseteq Q \setminus L$. The calculation

$$\begin{aligned} PP', QQ' \in L, P \sqsubseteq Q &\iff P' \in P \setminus L, Q' \in Q \setminus L, P \sqsubseteq Q \implies \\ &P' \in Q \setminus L \iff QP' \in L \end{aligned}$$

shows the converse.

Our main goal is to show the following criterion, which will be implied by Props. 47 and 49 below.

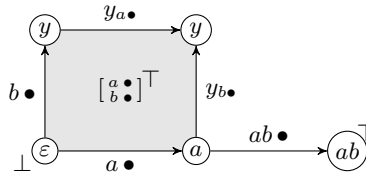
Theorem 41. *A language L is deterministic iff it is swap-invariant.*

Example 42. The regular language $L = \{[a/b], ab, ba, abc\}$ from Example 20 is not swap-invariant: using Lemma 40, $ab \bullet \sqsubseteq [a/b \bullet]$, but $\{ab \bullet\} \setminus L = \{\bullet b, \bullet bc\} \neq \{\bullet b\} = \{[a/b \bullet]\} \setminus L$. Hence L is not deterministic.

The next examples explain why we need to restrict to essential cells in the definition of determinacy.

Example 43. The HDA in Example 22 is deterministic. There are two different a -labelled edges starting at w_ε (w_a and $\langle [a/b \bullet] \rangle$), yet it does not disturb determinism since w_ε is not accessible.

Example 44. Let $L = \{ab, [a/b \bullet]\}$. Then $\text{MN}(L)^{\text{ess}}$ is as follows:



It is deterministic; there are two b -labelled edges leaving a , namely $y_{b \bullet}$ and $ab \bullet$, but only the latter is coaccessible.

Lemma 45. *Let X be a deterministic HDA and $\alpha, \beta \in \text{Path}(X)_\perp$ with $\text{tgt}(\alpha), \text{tgt}(\beta) \in \text{ess}(X)$. If $\text{src}(\alpha) = \text{src}(\beta)$ and $\text{ev}(\alpha) = \text{ev}(\beta)$, then $\text{tgt}(\alpha) = \text{tgt}(\beta)$.*

Proof. We can assume that $\alpha = \alpha_1 * \dots * \alpha_n$ and $\beta = \beta_1 * \dots * \beta_m$ are sparse; note that all of these cells are essential. Denote $P = \text{ev}(\alpha) = \text{ev}(\beta)$, then

$$P = \text{ev}(\alpha) = \text{ev}(\alpha_1) * \dots * \text{ev}(\alpha_n)$$

is a sparse step decomposition of P . Similarly, $P = \text{ev}(\beta_1) * \dots * \text{ev}(\beta_m)$ is a sparse step decomposition. Yet sparse step decompositions are unique by Prop. 4; hence,

$m = n$ and $\text{ev}(\alpha_k) = \text{ev}(\beta_k)$ for every k . We show by induction that $\alpha_k = \beta_k$. Assume that $\alpha_{k-1} = \beta_{k-1}$. Let $x = \text{src}(\alpha_k) = \text{tgt}(\alpha_{k-1}) = \text{tgt}(\beta_{k-1}) = \text{src}(\beta_k)$. If $P_k = \text{ev}(\alpha_k) = \text{ev}(\beta_k)$ is a terminator $U \downarrow_B$, then $\alpha_k = \delta_B^1(x) = \beta_k$. If P_k is a starter $A \uparrow U$, then there are $y, z \in X$ such that $\delta_A^0(y) = \delta_A^0(z) = x$. As y and z are essential and X is deterministic, this implies $y = z$ and $\alpha_k = \beta_k$. \square

Lemma 46. *Let α and β be essential paths on a deterministic HDA X . Assume that $\text{src}(\alpha) = \text{src}(\beta)$ and $\text{ev}(\alpha) \sqsubseteq \text{ev}(\beta)$. Then $\text{tgt}(\alpha) = \text{tgt}(\beta)$.*

Proof. Denote $x \in \text{src}(\alpha) = \text{src}(\beta)$ and $y = \text{tgt}(\beta)$. By Lemma 14 there exists an HDA-map $f : \square^{\text{ev}(\beta)} \rightarrow X_x^y$. By [6, Lemma 63] there is an HDA-map $i : \square^{\text{ev}(\alpha)} \rightarrow \square^{\text{ev}(\beta)}$. We apply Lemma 14 again to the composition $f \circ i$ and obtain that there is a path $\alpha' \in \text{Path}(X)_x^y$ such that $\text{ev}(\alpha') = \text{ev}(\alpha)$. Lemma 45 then implies $\text{tgt}(\alpha) = \text{tgt}(\alpha') = y$. \square

Proposition 47. *If L is deterministic, then L is swap-invariant.*

Proof. Let X be a deterministic automaton that recognises L and fix ipomsets $P \sqsubseteq Q$. From Lemma 16 follows that $Q \setminus L \subseteq P \setminus L$. It remains to prove that if $Q \setminus L \neq \emptyset$, then $P \setminus L \subseteq Q \setminus L$. Denote $U \cong S_P \cong S_Q$.

Let $R \in Q \setminus L$ and let $\omega \in \text{Path}(X)_{\langle \text{id}_U \rangle}^\top$ be an accepting path that recognises QR . By Lemma 15, there exists a path $\beta \in \text{Path}(X)_{\langle \text{id}_U \rangle}$ such that $\text{ev}(\beta) = Q$.

Now assume that $R' \in P \setminus L$, and let $\omega' \in \text{Path}(X)_{\langle \text{id}_U \rangle}^\top$ be a path such that $\text{ev}(\omega') = PR'$. By Lemma 15, there exist paths $\alpha \in \text{Path}(X)_{\langle \text{id}_U \rangle}$ and $\gamma \in \text{Path}(X)^{\text{tgt}(\omega')}$ such that $\text{tgt}(\alpha) = \text{src}(\gamma)$, $\text{ev}(\alpha) = P$ and $\text{ev}(\gamma) = R'$. From Lemma 46 and $P \sqsubseteq Q$ follows that $\text{tgt}(\alpha) = \text{tgt}(\beta)$. Thus, β and γ may be concatenated to an accepting path $\beta * \gamma$. By $\text{ev}(\beta * \gamma) = QR'$ we have $QR' \in L$, i.e., $R' \in Q \setminus L$. \square

Lemma 48. *If $\langle P \rangle \in \text{ess}(\text{MN}(L))$ and $A \subseteq \text{rfin}(P)$, then $\langle P - A \rangle \in \text{ess}(\text{MN}(L))$.*

Proof. By Lemma 33, $\langle P - A \rangle$ is accessible. By assumption, $\langle P \rangle$ is coaccessible and $(\langle P - A \rangle \nearrow^A \langle P \rangle)$ is a path, so $\langle P - A \rangle$ is also coaccessible. \square

Proposition 49. *If L is swap-invariant, then $\text{MN}(L)$ is deterministic.*

Proof. $\text{MN}(L)$ contains only one start cell $\langle \text{id}_U \rangle$ for every $U \in \square$.

Fix $U \in \square$, $P, Q \in \text{iiPoms}_U$ and $A \subseteq U$. Assume that $\delta_A^0(\langle P \rangle) = \delta_A^0(\langle Q \rangle)$, i.e., $\langle P - A \rangle = \langle Q - A \rangle$, and $\langle P \rangle, \langle Q \rangle, \langle P - A \rangle \in \text{ess}(\text{MN}(L))$. We will prove that $\langle P \rangle = \langle Q \rangle$, or equivalently, $P \approx_L Q$.

We have $\text{fin}(P - A) = \text{fin}(Q - A) =: s \uparrow (U - A)$. First, notice that A , regarded as a subset of P (or Q), contains no start events: else, we would have $\delta_A^0(\langle P \rangle) = w_{U-A}$ (or $\delta_A^0(\langle Q \rangle) = w_{U-A}$). As a consequence, $\text{fin}(P) = \text{fin}(Q) = s \uparrow U$.

For every $B \subseteq \text{rfin}(P) = \text{rfin}(Q)$ we have

$$\begin{aligned} (P - A) \approx_L (Q - A) &\implies \\ (P - (A \cup B)) \setminus L = (Q - (A \cup B)) \setminus L &\implies \\ ((P - (A \cup B)) *_{(A-B) \uparrow U} \setminus L) = ((Q - (A \cup B)) *_{(A-B) \uparrow U} \setminus L). \end{aligned}$$

The first implication follows from the definition, and the second from Lemma 23. From Lemma 24 follows that

$$(P - (A \cup B)) *_{(A-B)\uparrow} U \sqsubseteq P - B, \quad (Q - (A \cup B)) *_{(A-B)\uparrow} U \sqsubseteq Q - B.$$

Thus, by swap-invariance we have $(P - B) \setminus L = (Q - B) \setminus L$; note that Lemma 48 guarantees that neither of these languages is empty. \square

6 Conclusion and Further Work

We have proven a Myhill-Nerode type theorem for higher-dimensional automata (HDAs), stating that a language is regular iff it has finite prefix quotient. We have also introduced deterministic HDAs and shown that not all finite HDAs are determinizable.

An obvious follow-up question to ask is whether finite HDAs are *learnable*, that is, whether our Myhill-Nerode construction can be used to introduce a learning procedure for HDAs akin to Angluin’s L^* algorithm [1] or some other recent approaches [2, 15, 16]. (See also [33] which introduces learning for pomset automata.)

Our Myhill-Nerode theorem provides a language-internal criterion for whether a language is regular, and we have developed a similar one to distinguish deterministic languages. Another important aspect is the *decidability* of these questions, together with other standard problems such as membership or language equivalence. We believe that membership of an ipomset in a regular language is decidable, but we are less sure about decidability of the other problems.

Given that we have shown that not all regular languages are deterministic, one might ask for the approximation of deterministic languages by other, less restrictive notions. Preliminary results indicate that *ambiguity* does not buy much, given that we seem to have found a language of unbounded ambiguity; an avenue that remains wide open is the one of *history-determinism* [4, 14, 20].

Lastly, a remark on the fact that we only consider subsumption-closed (or *weak*) languages in this work. While this is quite common in concurrency theory, see for example [10, 11, 13, 34], an extension of our setting to non-weak languages would certainly be interesting. (Note that, for example, languages of Petri nets with inhibitor arcs are non-weak [18].) Such an extension may be obtained by considering *partial* HDAs or HDAs with interfaces, see [5, 7, 9], but this is subject to future work.

Acknowledgement. We are indebted to Amazigh Amrane, Hugo Bazille, Christian Johansen, and Georg Struth for numerous discussions regarding the subjects of this paper; any errors, however, are exclusively ours.

References

1. Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75(2):87–106, 1987.

2. Simone Barlocco, Clemens Kupke, and Jurriaan Rot. Coalgebra learning via duality. In Mikołaj Bojańczyk and Alex Simpson, editors, *FOSSACS*, volume 11425 of *Lecture Notes in Computer Science*, pages 62–79. Springer, 2019.
3. Marek A. Bednarczyk. *Categories of Asynchronous Systems*. PhD thesis, University of Sussex, UK, 1987.
4. Thomas Colcombet. The theory of stabilisation monoids and regular cost functions. In Susanne Albers, Alberto Marchetti-Spaccamela, Yossi Matias, Sotiris E. Nikolettseas, and Wolfgang Thomas, editors, *ICALP*, volume 5556 of *Lecture Notes in Computer Science*, pages 139–150. Springer, 2009.
5. Jérémy Dubut. Trees in partial higher dimensional automata. In Mikołaj Bojańczyk and Alex Simpson, editors, *FOSSACS*, volume 11425 of *Lecture Notes in Computer Science*, pages 224–241. Springer, 2019.
6. Uli Fahrenberg, Christian Johansen, Georg Struth, and Krzysztof Ziemiański. Languages of higher-dimensional automata. *Mathematical Structures in Computer Science*, 31(5):575–613, 2021. <https://arxiv.org/abs/2103.07557>.
7. Uli Fahrenberg, Christian Johansen, Georg Struth, and Krzysztof Ziemiański. A Kleene theorem for higher-dimensional automata. In Bartek Klin, Sławomir Lasota, and Anca Muscholl, editors, *CONCUR*, volume 243 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 29:1–29:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. <https://arxiv.org/abs/2202.03791>.
8. Uli Fahrenberg, Christian Johansen, Georg Struth, and Krzysztof Ziemiański. Posets with interfaces as a model for concurrency. *Information and Computation*, 285(B):104914, 2022. <https://arxiv.org/abs/2106.10895>.
9. Uli Fahrenberg and Axel Legay. Partial higher-dimensional automata. In Lawrence S. Moss and Pawel Sobocinski, editors, *CALCO*, volume 35 of *Leibniz International Proceedings in Informatics*, pages 101–115. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015.
10. Jean Fanchon and Rémi Morin. Regular sets of pomsets with autoconcurrency. In Luboš Brim, Petr Jančar, Mojmir Křetínský, and Antonín Kučera, editors, *CONCUR*, volume 2421 of *Lecture Notes in Computer Science*, pages 402–417. Springer, 2002.
11. Jean Fanchon and Rémi Morin. Pomset languages of finite step transition systems. In Giuliana Franceschinis and Karsten Wolf, editors, *PETRI NETS*, volume 5606 of *Lecture Notes in Computer Science*, pages 83–102. Springer, 2009.
12. Peter C. Fishburn. *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets*. Wiley, 1985.
13. Jan Grabowski. On partial languages. *Fundamentae Informatica*, 4(2):427, 1981.
14. Thomas A. Henzinger and Nir Piterman. Solving games without determinization. In Zoltán Ésik, editor, *CSL*, volume 4207 of *Lecture Notes in Computer Science*, pages 395–410. Springer, 2006.
15. Falk Howar and Bernhard Steffen. Active automata learning as black-box search and lazy partition refinement. In Nils Jansen, Mariëlle Stoelinga, and Petra van den Bos, editors, *A Journey from Process Algebra via Timed Automata to Model Learning - Essays Dedicated to Frits Vaandrager on the Occasion of His 60th Birthday*, volume 13560 of *Lecture Notes in Computer Science*, pages 321–338. Springer, 2022.
16. Malte Isberner, Falk Howar, and Bernhard Steffen. The TTT algorithm: a redundancy-free approach to active automata learning. In Borzoo Bonakdarpour and Scott A. Smolka, editors, *RV*, volume 8734 of *Lecture Notes in Computer Science*, pages 307–322. Springer, 2014.

17. Ryszard Janicki and Maciej Koutny. Structure of concurrency. *Theoretical Computer Science*, 112(1):5–52, 1993.
18. Ryszard Janicki and Maciej Koutny. Operational semantics, interval orders and sequences of antichains. *Fundamentae Informatica*, 169(1-2):31–55, 2019.
19. Christian Johansen. ST-structures. *Journal of Logic and Algebraic Methods in Programming*, 85(6):1201–1233, 2015. <https://arxiv.org/abs/1406.0641>.
20. Orna Kupferman, Shmuel Safra, and Moshe Y. Vardi. Relating word and tree automata. *Annals of Pure and Applied Logic*, 138(1-3):126–146, 2006.
21. Mogens Nielsen, Gordon D. Plotkin, and Glynn Winskel. Petri nets, event structures and domains, part I. *Theoretical Computer Science*, 13:85–108, 1981.
22. Carl A. Petri. *Kommunikation mit Automaten*. Number 2 in Schriften des IIM. Institut für Instrumentelle Mathematik, Bonn, 1962.
23. Vaughan R. Pratt. Modeling concurrency with geometry. In *POPL*, pages 311–322, New York City, 1991. ACM Press.
24. Vaughan R. Pratt. Chu spaces and their interpretation as concurrent objects. In *Computer Science Today: Recent Trends and Developments*, volume 1000 of *Lecture Notes in Computer Science*, pages 392–405. Springer, 1995.
25. Vaughan R. Pratt. Transition and cancellation in concurrency and branching time. *Mathematical Structures in Computer Science*, 13(4):485–529, 2003.
26. Mike W. Shields. Concurrent machines. *Comput. J.*, 28(5):449–465, 1985.
27. Rob J. van Glabbeek. Bisimulations for higher dimensional automata. Email message, June 1991. <http://theory.stanford.edu/~rvg/hda>.
28. Rob J. van Glabbeek. On the expressiveness of higher dimensional automata. *Theoretical Computer Science*, 356(3):265–290, 2006. See also [29].
29. Rob J. van Glabbeek. Erratum to “On the expressiveness of higher dimensional automata”. *Theoretical Computer Science*, 368(1-2):168–194, 2006.
30. Rob J. van Glabbeek and Ursula Goltz. Refinement of actions and equivalence notions for concurrent systems. *Acta Informatica*, 37(4/5):229–327, 2001.
31. Rob J. van Glabbeek and Gordon D. Plotkin. Configuration structures. In *LICS*, pages 199–209. IEEE Computer Society, 1995.
32. Rob J. van Glabbeek and Gordon D. Plotkin. Configuration structures, event structures and Petri nets. *Theoretical Computer Science*, 410(41):4111–4159, 2009.
33. Gerco van Heerdt, Tobias Kappé, Jurriaan Rot, and Alexandra Silva. Learning pomset automata. In Stefan Kiefer and Christine Tasson, editors, *FOSSACS*, volume 12650 of *Lecture Notes in Computer Science*, pages 510–530. Springer, 2021.
34. Walter Vogler. *Modular Construction and Partial Order Semantics of Petri Nets*, volume 625 of *Lecture Notes in Computer Science*. Springer, 1992.