# The Cost of Monitoring Alone*

Luca Aceto[1,2][0000−0002−2197−3018], Antonis Achilleos[2][0000−0002−1314−333X], Adrian Francalanza[3][0000−0003−3829−7391], Anna Ingólfsdóttir[2][0000−0001−8362−3075], and Karoliina Lehtinen[4][0000−0003−1171−8790]

[1] Gran Sasso Science Institute luca.aceto@gssi.it
[2] Reykjavik University {luca,antonios,annai}@ru.is
[3] University of Malta adrian.francalanza@um.edu.mt
[4] University of Liverpool karoliina.lehtinen@liverpool.ac.uk

**Abstract.** We compare the succinctness of two monitoring systems for properties of infinite traces, namely parallel and regular monitors. Although a parallel monitor can be turned into an equivalent regular monitor, the cost of this transformation is a double-exponential blowup in the syntactic size of the monitors, and a triple-exponential blowup when the goal is a deterministic monitor. We show that these bounds are tight and that they also hold for translations between corresponding fragments of Hennessy-Milner logic with recursion over infinite traces.

## 1 Introduction

Runtime Verification is a lightweight verification technique where a computational entity that we call a monitor is used to observe a system run in order to verify a given property. That property, which we choose to formalize in Hennessy-Milner logic with recursion (recHML) [13], can be a potential property of either the system [1, 12], or of the current system run, encoded as a trace of events [4] — see also, for example, [8, 10, 14] for earlier work on the monitoring of trace properties, mainly formalized on LTL.

To address the case of verifying trace properties, the authors introduced in [4] a class of monitors that can generate multiple parallel components that analyse the same system trace. These were called parallel monitors. When some of them reach a verdict, they can combine these verdicts into one. In the same paper, it was determined that this monitoring system has the same monitoring power as its restriction to a single monitoring component, as it was used in [1, 12], called regular monitors. However, the cost of the translation from the more general monitoring system to this fragment, as given in [4], is doubly exponential with respect to the syntactic size of the monitors. Furthermore, if the goal is a

* This research was partially supported by the projects "TheoFoMon: Theoretical Foundations for Monitorability" (grant number: 163406-051) and "Epistemic Logic for Distributed Runtime Monitoring" (grant number: 184940-051) of the Icelandic Research Fund, by the BMBF project "Aramis II" (project number: 01IS160253) and the EPSRC project "Solving parity games in theory and practice" (project number: EP/P020909/1).

deterministic regular monitor [2, 3], then the resulting monitor is quadruply-exponentially larger than the original, parallel one, in [4].

In this paper, we show that the double-exponential cost for translating from parallel to equivalent regular monitors is tight. Furthermore, we improve the translation cost from parallel monitors to equivalent deterministic monitors to a triple exponential, and we show that this bound is tight. We define monitor equivalence in two ways, the first one stricter than the second. For the first definition, two monitors are equivalent when they reach the same verdicts for the same finite traces, while for the second one it suffices to reach the same verdicts for the same infinite traces. We prove the upper bounds for a transformation that gives monitors that are equivalent with respect to the stricter definition, while we prove the lower bounds with respect to transformations that satisfy the coarser definition. Therefore, our bounds hold for both definitions of monitor equivalence. This treatment allows us to derive stronger results, which yield similar bounds for the case of logical formulae, as well.

In [4], we show that, when interpreted over traces, MXHML, the fragment of RECHML that does not use least fixed points, is equivalent to the syntactically smaller safety fragment SHML. That is, every MXHML formula can be translated to a logically equivalent SHML formula. Similarly to the aforementioned translation of monitors, this translation of formulae results in a formula that that is syntactically at most doubly-exponentially larger than the original formula. We show that this upper bound is tight.

The first four authors have worked on the complexity of monitor transformations before in [2, 3], where the cost of determinizing monitors is examined. Similarly to [2, 3], in [4], but also in this paper, we use results and techniques from Automata Theory and specifically about alternating automata [9, 11].

In Sec. 2, we introduce the necessary background on monitors and RECHML on infinite traces, as these were used in [4]. In Sec. 3, we describe the monitor translations that we mentioned above, and we provide upper bounds for these, which we prove to be tight in Sec. 4. In Sec. 5, we extrapolate these bounds to the case where we translate logical formulae, from MXHML to SHML. In Sec. 6, we conclude the paper. Omitted proofs can be found in the appendix.

## 2 Preliminaries

Monitors are expected to monitor for a specification, which, in our case, is written in RECHML. We use the linear-time interpretation of the logic RECHML, as it was given in [4]. According to that interpretation, formulae are interpreted over infinite *traces*.

### 2.1 The model and the logic

We assume a finite set of actions $\alpha, \beta, \ldots \in \text{Act}$ with distinguished silent action $\tau$. We also assume that $\tau \notin \text{Act}$ and that $\mu \in \text{Act} \cup \{\tau\}$, and refer to the actions in Act as *visible* actions (as opposed to the silent action $\tau$). The metavariables

$t, u \in \textsc{Trc} = \textsc{Act}^{\omega}$ range over (infinite) sequences of visible actions, which abstractly represent system runs. We also use the metavariable $T \subseteq \textsc{Trc}$ to range over *sets of traces*. We often need to refer to *finite traces*, denoted as $s, r \in \textsc{Act}^{*}$, to represent objects such as a finite prefix of a system run, or to traces that may be finite or infinite (*finfinite traces*, as they were called in [4]), denoted as $g, h \in \textsc{Act}^{*} \cup \textsc{Act}^{\omega}$. A trace (*resp.*, finite trace, *resp.*, finfinite trace) with action $\alpha$ at its head is denoted as $\alpha t$ (*resp.*, $\alpha s$, *resp.*, $\alpha g$). Similarly a trace with a prefix $s$ is written $st$.

**Syntax**

$$\varphi, \psi \in \textsc{recHML} ::= \mathsf{tt} \mid \mathsf{ff} \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \langle \alpha \rangle \varphi \mid [\alpha]\varphi \mid \mathsf{min}\, X.\varphi \mid \mathsf{max}\, X.\varphi \mid X$$

**Linear-Time Semantics**

$$\begin{array}{rcl@{\qquad}rcl}
[\![\mathsf{tt}, \rho]\!] & \stackrel{\text{def}}{=} & \textsc{Trc} & [\![\mathsf{ff}, \rho]\!] & \stackrel{\text{def}}{=} & \emptyset \\
[\![\varphi_1 \wedge \varphi_2, \rho]\!] & \stackrel{\text{def}}{=} & [\![\varphi_1, \rho]\!] \cap [\![\varphi_2, \rho]\!] & [\![\varphi_1 \vee \varphi_2, \rho]\!] & \stackrel{\text{def}}{=} & [\![\varphi_1, \rho]\!] \cup [\![\varphi_2, \rho]\!] \\
[\![[\alpha]\varphi, \rho]\!] & \stackrel{\text{def}}{=} & \{\beta t \mid \beta \neq \alpha \text{ or } t \in [\![\varphi, \rho]\!]\} & [\![\langle \alpha \rangle \varphi, \rho]\!] & \stackrel{\text{def}}{=} & \{\alpha t \mid t \in [\![\varphi, \rho]\!]\} \\
[\![\mathsf{min} X.\varphi, \rho]\!] & \stackrel{\text{def}}{=} & \bigcap \{T \mid [\![\varphi, \rho[X \mapsto T]]\!] \subseteq T\} & & & \\
[\![\mathsf{max} X.\varphi, \rho]\!] & \stackrel{\text{def}}{=} & \bigcup \{T \mid T \subseteq [\![\varphi, \rho[X \mapsto T]]\!]\} & [\![X, \rho]\!] & \stackrel{\text{def}}{=} & \rho(X)
\end{array}$$

**Fig. 1.** recHML Syntax and Linear-Time Semantics

The logic recHML [7, 13] assumes a countable set LVar (with $X \in \textsc{LVar}$) of logical variables, and is defined as the set of *closed* formulae generated by the grammar of Fig. 1. Apart from the standard constructs for truth, falsehood, conjunction and disjunction, the logic is equipped with possibility and necessity modal operators labelled by visible actions, together with recursive formulae expressing least or greatest fixpoints; formulae $\mathsf{min}\, X.\varphi$ and $\mathsf{max}\, X.\varphi$ bind free instances of the logical variable $X$ in $\varphi$, inducing the usual notions of open/closed formulae and formula equality up to alpha-conversion.

We interpret recHML formulae over traces, using an interpretation function $[\![-]\!]$ that maps formulae to sets of traces, relative to an environment $\rho : \textsc{LVar} \to 2^{\textsc{Trc}}$, which intuitively assigns to each variable $X$ the set of traces that are assumed to satisfy it, as defined in Fig. 1. The semantics of a closed formula $\varphi$ is independent of the environment $\rho$ and is simply written $[\![\varphi]\!]$. Intuitively, $[\![\varphi]\!]$ denotes the set of traces satisfying $\varphi$. For a formula $\varphi$, we use $l(\varphi)$ to denote the length of $\varphi$ as a string of symbols.

**Syntax**

$$m, n \in \textsc{Mon} ::= v \mid \alpha.m \mid m + n \mid \mathsf{rec}\, x.m \mid x \mid m \otimes n \mid m \oplus n$$
$$v \in \textsc{Verd} ::= \mathsf{no} \mid \mathsf{yes} \mid \mathsf{end}$$

**Dynamics**

`Regular monitor rules:`

$$\textsc{Act}\ \frac{}{\alpha.m \xrightarrow{\alpha} m} \qquad \textsc{RecF}\ \frac{}{\mathsf{rec}\, x.m \xrightarrow{\tau} m} \qquad \textsc{RecB}\ \frac{}{x \xrightarrow{\tau} p_x}$$
$$\textsc{SelL}\ \frac{m \xrightarrow{\mu} m'}{m + n \xrightarrow{\mu} m'} \qquad \textsc{SelR}\ \frac{n \xrightarrow{\mu} n'}{m + n \xrightarrow{\mu} n'} \qquad \textsc{Ver}\ \frac{}{v \xrightarrow{\alpha} v}$$

`Parallel tracing rules:`

$$\textsc{Par}\ \frac{m \xrightarrow{\alpha} m' \quad n \xrightarrow{\alpha} n'}{m \odot n \xrightarrow{\alpha} m' \odot n'} \qquad \textsc{TauL}\ \frac{m \xrightarrow{\tau} m'}{m \odot n \xrightarrow{\tau} m' \odot n}$$

`Parallel evaluation rules:`

$$\textsc{VrE}\ \frac{}{\mathsf{end} \odot \mathsf{end} \xrightarrow{\tau} \mathsf{end}} \qquad \textsc{VrC1}\ \frac{}{\mathsf{yes} \otimes m \xrightarrow{\tau} m} \qquad \textsc{VrC2}\ \frac{}{\mathsf{no} \otimes m \xrightarrow{\tau} \mathsf{no}}$$
$$\textsc{VrD1}\ \frac{}{\mathsf{no} \oplus m \xrightarrow{\tau} m} \qquad \textsc{VrD2}\ \frac{}{\mathsf{yes} \oplus m \xrightarrow{\tau} \mathsf{yes}}$$

**Fig. 2.** Monitor Syntax and Semantics

### 2.2 Two monitoring systems

We now present two monitoring systems, parallel and regular monitors, that were introduced in [1,4,12]. A monitoring system is a *Labelled Transition System (LTS)* based on Act, the set of actions, that is comprised of the monitor states, or monitors, and a transition relation. The set of monitor states, Mon, and the monitor transition relation, $\longrightarrow \subseteq (\textsc{Mon} \times (\textsc{Act} \cup \{\tau\}) \times \textsc{Mon})$, are defined in Fig. 2. There and elsewhere, $\odot$ ranges over both parallel operators $\oplus$ and $\otimes$. When discussing a monitor with free variables (an *open monitor*) $m$, we assume it is part of a larger monitor $m'$ without free variables (a *closed monitor*), where every variable $x$ appears at most once in a recursive operator. Therefore, we assume an injective mapping from each monitor variable $x$ to a unique monitor $p_x$, of the form $\mathsf{rec}\, x.m$ that is a submonitor of $m'$.

The suggestive notation $m \xrightarrow{\mu} n$ denotes $(m, \mu, n) \in \longrightarrow$; we also write $m \overset{\mu}{\not\longrightarrow}$ to denote $\neg(\exists n.\ m \xrightarrow{\mu} n)$. We employ the usual notation for weak transitions and write $m \Longrightarrow n$ in lieu of $m(\xrightarrow{\tau})^* n$ and $m \overset{\mu}{\Longrightarrow} n$ for $m \Longrightarrow \cdot \xrightarrow{\mu} \cdot \Longrightarrow n$. We write sequences of transitions $m \overset{\alpha_1}{\Longrightarrow} \cdots \overset{\alpha_k}{\Longrightarrow} m_k$ as $m \overset{s}{\Longrightarrow} m_k$, where $s = \alpha_1 \cdots \alpha_k$. The monitoring system of *parallel monitors* is defined using the full syntax and all the rules from Fig. 2; *regular monitors* are parallel monitors that do not use the parallel operators $\otimes$ and $\oplus$. Regular monitors were defined and used already in [1] and [12], while parallel monitors were defined in [4]. We observe that the

rules RecF and RecB are not the standard recursion rules from [1] and [12], but they are equivalent to those rules [4, 5] and more convenient for our arguments.

A transition $m \xrightarrow{\alpha} n$ denotes that the monitor in state $m$ can *analyse* the (visible) action $\alpha$ and transition to state $n$. Monitors may reach any one of *three* verdicts after analysing a finite trace: *acceptance*, $\mathsf{yes}$, *rejection*, $\mathsf{no}$, and the *inconclusive* verdict $\mathsf{end}$. We highlight the transition rule for verdicts in Fig. 2, describing the fact that from a verdict state any action can be analysed by transitioning to the same state; verdicts are thus *irrevocable*. Rule Par states that *both* submonitors need to be able to analyse an external action $\alpha$ for their parallel composition to transition with that action. The rules in Fig. 2 also allow $\tau$-transitions for the reconfiguration of parallel compositions of monitors. For instance, rules VrC1 and VrC2 describe the fact that, in conjunctive parallel compositions, whereas $\mathsf{yes}$ verdicts are uninfluential, $\mathsf{no}$ verdicts supersede the verdicts of other monitors (Fig. 2 omits the obvious symmetric rules). The dual applies for $\mathsf{yes}$ and $\mathsf{no}$ verdicts in a disjunctive parallel composition, as described by rules VrD1 and VrD2 (again, we omit the obvious symmetric rules). Rule VrE applies to both forms of parallel composition and consolidates multiple inconclusive verdicts. Finally, rules TauL and its omitted dual TauR are contextual rules for these monitor reconfiguration steps.

**Definition 1 (Acceptance and Rejection).** *We say that $m$* rejects *(resp.,* accepts*) $s \in \text{Act}^*$ when $m \stackrel{s}{\Longrightarrow} \mathsf{no}$ (resp., $m \stackrel{s}{\Longrightarrow} \mathsf{yes}$). We similarly say that $m$ rejects (resp., accepts) $t \in \text{Act}^\omega$ if $m$ rejects (resp., accepts) some prefix of $t$.*

Just like for formulae, we use $l(m)$ to denote the length of $m$ as a string of symbols. In the sequel, for a finite nonempty set of indices $I$, we use $\sum_{i \in I} m_i$ to denote any combination of the monitors in $\{m_i \mid i \in I\}$ using the operator $+$. The notation is justified, because $+$ is commutative and associative with respect to the transitions that a resulting monitor can exhibit. For each $j \in I$, $m_j$ is called a summand of $\sum_{i \in I} m_i$ (and the term $\sum_{i \in I} m_i$ is called a sum of $m_j$). The regular monitors in Fig. 2 have an important property, namely that their state space, *i.e.,* the set of reachable states, is finite (see Remark 1). On the other hand, parallel monitors can be infinite-state, but they are convenient when one synthesizes monitors. However, the two monitoring systems are equivalent (see Prop. 2). For a monitor $m$, $\mathsf{reach}(m)$ is the set of monitor states reachable through a transition sequence from $m$.

**Lemma 1 (Verdict Persistence, [4, 12]).** *$v \stackrel{s}{\Longrightarrow} m$ implies $m = v$.*

**Lemma 2.** *Every submonitor of a closed regular monitor $m$ can only transition to submonitors of $m$.*

*Remark 1.* An immediate consequence of Lem. 2 is that regular monitors are finite-state. This is not the case for parallel monitors, in general. For example, consider parallel monitor $m_\tau = \mathsf{rec}\, x.(x \otimes (a.\mathsf{yes} + b.\mathsf{yes}))$. We can see that there

is a unique sequence of transitions that can be made from $m_\tau$:

$$\begin{aligned} m_\tau &\xrightarrow{\tau} x \otimes (a.\mathsf{yes} + b.\mathsf{yes}) \xrightarrow{\tau} m_\tau \otimes (a.\mathsf{yes} + b.\mathsf{yes}) \\ &\xrightarrow{\tau} (x \otimes (a.\mathsf{yes} + b.\mathsf{yes})) \otimes (a.\mathsf{yes} + b.\mathsf{yes}) \\ &\xrightarrow{\tau} (m_\tau \otimes (a.\mathsf{yes} + b.\mathsf{yes})) \otimes (a.\mathsf{yes} + b.\mathsf{yes}) \longrightarrow \cdots \end{aligned}$$ □

One basic requirement that we maintain on monitors is that they are not allowed to give conflicting verdicts for the same trace.

**Definition 2 (Monitor Consistency).** *A monitor $m$ is* consistent *when there is* no *finite trace $s$ such that $m \stackrel{s}{\Longrightarrow} \mathsf{yes}$ and $m \stackrel{s}{\Longrightarrow} \mathsf{no}$.*

We identify a useful monitor predicate that allows us to neatly decompose the behaviour of a parallel monitor in terms of its constituent sub-monitors.

**Definition 3 (Monitor Reactivity).** *We call a monitor $m$* reactive *when for every $n \in \mathsf{reach}(m)$ and $\alpha \in \textsc{Act}$, there is some $n'$ such that $n \stackrel{\alpha}{\Longrightarrow} n'$.*

The following lemma states that parallel monitors behave as expected with respect to the acceptance and rejection of traces as long as the constituent sub-monitors are reactive.

**Lemma 3 ([4]).** *For reactive $m_1$ and $m_2$:*

- *$m_1 \otimes m_2$ rejects $t$ if and only if either $m_1$ or $m_2$ rejects $t$.*
- *$m_1 \otimes m_2$ accepts $t$ if and only if both $m_1$ and $m_2$ accept $t$.*
- *$m_1 \oplus m_2$ rejects $t$ if and only if both $m_1$ and $m_2$ reject $t$.*
- *$m_1 \oplus m_2$ accepts $t$ if and only if either $m_1$ or $m_2$ accepts $t$.*

The following example, which stems from [4], indicates why the assumption that $m_1$ and $m_2$ are reactive is needed in Lem. 3.

*Example 1.* Assume that $\textsc{Act} = \{a, b\}$. The monitors $a.\mathsf{yes} + b.\mathsf{no}$ and $\mathsf{rec}\, x.(a.x + b.\mathsf{yes})$ are both reactive. The monitor $m = a.\mathsf{yes} \otimes b.\mathsf{no}$, however, is *not* reactive. Since the submonitor $a.\mathsf{yes}$ can only transition with $a$, according to the rules of Fig. 2, $m$ cannot transition with any action that is not $a$. Similarly, as the submonitor $b.\mathsf{no}$ can only transition with $b$, $m$ cannot transition with any action that is not $b$. Thus, $m$ cannot transition to any monitor, and therefore it cannot reject or accept any trace.

In general, we are interested in reactive parallel monitors, and the parallel monitors that we use will have this property.

### 2.3 Automata, Languages, Equivalence

In [4], we describe how to transform a parallel monitor to a verdict equivalent regular one. This transformation goes through alternating automata [9,11]. For our purposes, we only need to define nondeterministic and deterministic automata.

**Definition 4 (Finite Automata).** *A nondeterminitic finite automaton (NFA) is a quintuple $A = (Q, \text{Act}, q_0, \delta, F)$, where $Q$ is a finite set of states,* Act *is a finite alphabet (here it coincides with the set of actions), $q_0$ is the starting state, $F \subseteq Q$ is the set of accepting, or final states, and $\delta \subseteq Q \times \text{Act} \times Q$ is the transition relation. An NFA is deterministic (DFA) if $\delta$ is a function from $Q \times \text{Act}$ to $Q$.*

Given a state $q \in Q$ and a symbol $\alpha \in \text{Act}$, $\delta$ returns a set of possible states where the NFA can transition, and we typically use $q' \in \delta(q, \alpha)$ instead of $(q, \alpha, q') \in \delta$. We extend the transition relation to $\delta^* : Q \times \text{Act}^* \to 2^Q$, so that $\delta^*(q, \varepsilon) = \{q\}$ and $\delta^*(q, \alpha s) = \bigcup\{\delta^*(q', s) \mid q' \in \delta(q, \alpha)\}$. We say that the automaton *accepts* $s \in \text{Act}^*$ when $\delta^*(q_0, s) \cap F \neq \emptyset$, and that it recognizes $L \subseteq \text{Act}^*$ when $L$ is the set of strings accepted by the automaton.

**Definition 5 (Monitor Language Recognition).** *A monitor $m$ recognizes positively (resp., negatively) a set of finite traces (i.e., a language) $L \subseteq \text{Act}^*$ when for every $s \in \text{Act}^*$, $s \in L$ if and only if $m$ accepts (resp., rejects) $s$. We call the set that $m$ recognizes positively (resp., negatively) $L_a(m)$ (resp., $L_r(m)$). Similarly, we say that $m$ recognizes positively (resp., negatively) a set of infinite traces $L \subseteq \text{Act}^\omega$ when for every $t \in \text{Act}^\omega$, $t \in L$ if and only if $m$ accepts (resp., rejects) $t$.*

Observe that, by Lem. 1, $L_a(m)$ and $L_r(m)$ are closed under finite extensions.

**Lemma 4.** *The set of* infinite *traces that is recognized positively (resp., negatively) by $m$ is exactly $L_a(m) \cdot \text{Act}^\omega$ (resp., $L_r(m) \cdot \text{Act}^\omega$).*

*Proof.* The lemma is a consequence of verdict persistence (Lem. 1). □

To compare different monitors, we use a notion of monitor equivalence from [3] that focusses on how monitors can reach verdicts.

**Definition 6 (Verdict Equivalence).** *Monitors $m$ and $n$ are* verdict equivalent*, denoted as $m \simeq_v n$, if $L_a(m) = L_a(n)$ and $L_r(m) = L_r(n)$.*

One may consider the notion of verdict equivalence, as defined in Def. 6, to be too strict. After all, verdict equivalence is defined with respect to finite traces, so if we want to turn a parallel monitor into a regular or deterministic monitor, the resulting monitor not only needs to accept and reject the same infinite traces, but it is required to do so at the same time the original parallel monitor does. However, one may prefer to have a smaller, but not tight monitor, if possible, as long as it accepts the same infinite traces.

**Definition 7 ($\omega$-Verdict Equivalence).** *Monitors $m$ and $n$ are* $\omega$-verdict equivalent*, denoted as $m \simeq_\omega n$, if $L_a(m) \cdot \text{Act}^\omega = L_a(n) \cdot \text{Act}^\omega$ and $L_r(m) \cdot \text{Act}^\omega = L_r(n) \cdot \text{Act}^\omega$.*

From Lem. 4 we observe that verdict equivalence implies $\omega$-verdict equivalence. The converse does not hold, because $\mathsf{no} \simeq_\omega \sum_{\alpha \in \text{Act}} \alpha.\mathsf{no}$, but $\mathsf{no} \not\simeq_\text{v} \sum_{\alpha \in \text{Act}} \alpha.\mathsf{no}$.

**Definition 8 ([2]).** *A closed regular monitor $m$ is* deterministic *iff every sum of at least two summands that appears in $m$ is of the form $\sum_{\alpha \in A} \alpha.m_\alpha$, where $A \subseteq \text{Act}$.*

*Example 2.* The monitor $a.b.\mathsf{yes} + a.a.\mathsf{no}$ is not deterministic while the verdict equivalent monitor $a.(b.\mathsf{yes} + a.\mathsf{no})$ is deterministic.

### 2.4 Synthesis

There is a tight connection between the logic from Sec. 2.1 and the monitoring systems from Sec. 2.2. Ideally, we would want to be able to synthesize a monitor from any formula $\varphi$, such that the monitor recognizes $[\![\varphi]\!]$ positively and $\text{Trc} \setminus [\![\varphi]\!]$ negatively. However, as shown in [4], neither goal is possible for all formulae. Instead, we identify the following fragments of recHML.

**Definition 9 (MAX and MIN Fragments of recHML).** *The greatest-fixed-point and least-fixed-point fragments of* recHML *are, respectively, defined as:*

$$\varphi, \psi \in \text{mxHML} ::= \mathsf{tt} \mid \mathsf{ff} \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \langle \alpha \rangle \varphi \mid [\alpha]\varphi \mid \mathsf{max}\, X.\varphi \quad \textit{and}$$
$$\varphi, \psi \in \text{mnHML} ::= \mathsf{tt} \mid \mathsf{ff} \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \langle \alpha \rangle \varphi \mid [\alpha]\varphi \mid \mathsf{min}\, X.\varphi.$$

**Definition 10 (Safety and co-Safety Fragments of recHML).** *The safety and co-safety fragments of* recHML *are, respectively, defined as:*

$$\varphi, \psi \in \text{sHML} ::= \mathsf{tt} \mid \mathsf{ff} \mid \varphi \wedge \psi \mid [\alpha]\varphi \mid \mathsf{max}\, X.\varphi \quad \textit{and}$$
$$\varphi, \psi \in \text{cHML} ::= \mathsf{tt} \mid \mathsf{ff} \mid \varphi \vee \psi \mid \langle \alpha \rangle \varphi \mid \mathsf{min}\, X.\varphi.$$

**Theorem 1 (Monitorability and Maximality, [4]).**

1. *For every $\varphi \in$ mxHML (resp., $\varphi \in$ mnHML), there is a reactive parallel monitor $m$, such that $l(m) = O(l(\varphi))$ and $L_r(m) \cdot \text{Act}^\omega = \text{Act}^\omega \setminus [\![\varphi]\!]$ (resp., $L_a(m) \cdot \text{Act}^\omega = [\![\varphi]\!]$).*
2. *For every reactive parallel monitor $m$, there are $\varphi \in$ mxHML and $\psi \in$ mnHML, such that $l(\varphi), l(\psi) = O(l(m))$, $L_r(m) \cdot \text{Act}^\omega = \text{Act}^\omega \setminus [\![\varphi]\!]$, and $L_a(m) \cdot \text{Act}^\omega = [\![\psi]\!]$.*
3. *For every $\varphi \in$ sHML (resp., $\varphi \in$ cHML), there is a regular monitor $m$, such that $l(m) = O(l(\varphi))$ and $L_r(m) \cdot \text{Act}^\omega = \text{Act}^\omega \setminus [\![\varphi]\!]$ (resp., $L_a(m) \cdot \text{Act}^\omega = [\![\varphi]\!]$).*
4. *For every regular monitor $m$, there are $\varphi \in$ sHML and $\psi \in$ cHML, such that $l(\varphi), l(\psi) = O(m)$, $L_r(m) \cdot \text{Act}^\omega = \text{Act}^\omega \setminus [\![\varphi]\!]$, and $L_a(m) \cdot \text{Act}^\omega = [\![\psi]\!]$.*

We say that a logical fragment is monitorable for a monitoring system, such as parallel or regular monitors, when for each of the fragment's formulae there is a monitor that detects exactly the satisfying or violating traces for that formula. One of the consequences of Thm. 1 is that the fragments defined in Defs. 9 and 10 are semantically the largest monitorable fragments of recHML for parallel and

regular monitors, respectively. As we will see in Sec. 3, every parallel monitor has a verdict equivalent regular monitor (Props. 3 and 4), and therefore all formulae in MNHML and MXHML can be translated into equivalent CHML and SHML formulae respectively, as Thm. 4 later on demonstrates. However, Thm. 5 to follow shows that the cost of this translation is significant.

## 3 Monitor Transformations: Upper Bounds

In this section we explain how to transform a parallel monitor into a regular or deterministic monitor, and what is the cost, in monitor size, of this transformation. The various relevant transformations, including some from [2] and [9, 11], are summarized in Fig. 3, where each edge is labelled with the best-known worst-case upper bounds for the cost of the corresponding transformation in Fig. 3 (AFA abbreviates alternating finite automaton [9]). As we see in [2, 3] and in Sec. 4, these bounds cannot be improved significantly.

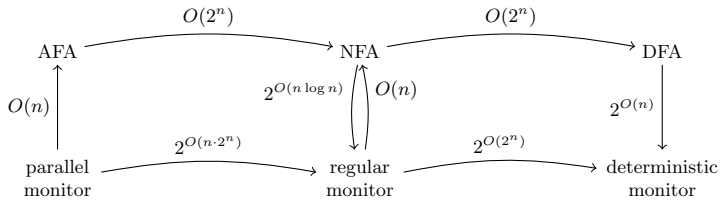


**Fig. 3.** Monitor Transformations and Costs

**Proposition 1 ([4]).** *For every reactive parallel monitor $m$, there are an alternating automaton that recognizes $L_a(m)$ and one that recognises $L_r(m)$, with $O(l(m))$ states.*

*Proof.* The construction from [4, Proposition 3.6] gives an automaton that has the submonitors of $m$ as states. The automaton's transition function corresponds to the semantics of the monitor. □

**Corollary 1 (Corollary 3.7 of [4])** *For every reactive and closed parallel monitor $m$, there are an NFA that recognises $L_a(m)$ and an NFA that recognises $L_r(m)$, and each has at most $2^{l(m)}$ states.*

**Proposition 2.** *For every reactive and closed parallel monitor $m$, there exists a verdict equivalent regular monitor $n$ such that $l(n) = 2^{O\left(l(m) \cdot 2^{l(m)}\right)}$.*

*Proof.* Let $A^a_m$ be an NFA for $L_a(m)$ with at most $2^{l(m)}$ states, and let $A^r_m$ be an NFA for $L_r(m)$ with at most $2^{l(m)}$ states, which exist by Cor. 1. From

these NFAs, we can construct regular monitors $m_R^a$ and $m_R^r$, such that $m_R^a$ recognizes $L_a(m)$ positively and $m_R^r$ recognizes $L_r(m)$ negatively, and $l(m_R^a), l(m_R^r) = 2^{O\left(l(m)\cdot 2^{l(m)}\right)}$ [3, Theorem 2]. Therefore, $m_R^a + m_R^r$ is regular and verdict equivalent to $m$, and $l(m_R^a + m_R^r) = 2^{O\left(l(m)\cdot 2^{l(m)}\right)}$. □

The constructions from [4] include a determinization result, based on [2].

**Theorem 2 (Corollary 3 of [3]).** *For every consistent closed regular monitor $m$, there is a deterministic monitor $n$ such that $n \simeq_v m$ and $l(n) = 2^{2^{O(l(m))}}$.*

**Proposition 3 (Proposition 3.11 of [4]).** *For every consistent reactive and closed parallel monitor $m$, there is a verdict equivalent deterministic regular monitor $n$ such that $l(n) = 2^{2^{2^{O\left(l(m)\cdot 2^{l(m)}\right)}}}$.*

However, the bound given by Prop. 3 for the construction of deterministic regular monitors from parallel ones is not optimal, as we observe below.

**Proposition 4.** *For every consistent reactive and closed parallel monitor $m$, there is a deterministic monitor $n$ such that $n \simeq_v m$ and $l(n) = 2^{2^{2^{O(l(m))}}}$.*

In the following Sec. 4, we see that the upper bounds of Props. 2 and 4 are tight, even for monitors that can only accept or reject, and even when the constructed regular or deterministic monitor is only required to be $\omega$-verdict equivalent to the starting one, and not necessarily verdict equivalent, to the original parallel monitor. As we only need to focus on acceptance monitors, in the following we say that a monitor recognizes a language to mean that it recognizes the language positively.

## 4 Lower Bounds

We now prove that the transformations of Sec. 3 are optimal, by establishing the corresponding lower bounds. To this end, we introduce a family of suffix-closed languages $L_A^k \subseteq \{0, 1, e, \$, \#\}^*$. Each $L_A^k$ is a variation of a language introduced in [9] to prove the $2^{2^{o(n)}}$ lower bound for the transformation from an alternating automaton to a deterministic one. In this section, we only need to consider closed monitors, and as such, all monitors are assumed to be closed.

A variation of the language that was introduced in [9] is the following:

$$L_V^k = \{u\#w\#v\$w \mid u, v \in \{0, 1, \#\}^*,\ w \in \{0, 1\}^k\}.$$

An alternating automaton that recognizes $L_V^k$ can nondeterministically skip to the first occurrence of $w$ and then verify that, for every number $i$ between 0 and $k-1$, the $i$'th bit matches the $i$'th bit after the \$ symbol. This verification can be done using up to $O(k)$ states, to count the position $i$ of the bit that is checked. On the other hand, a DFA that recognizes $L_V^k$ must remember all possible candidates for $w$ that have appeared before \$, and hence requires $2^{2^k}$ states. We can also conclude that any NFA for $L_V^k$ must have at least $2^k$ states, because a smaller NFA could be determinized to a smaller DFA.

*A gap language* For our purposes, we use a similar family $L_A^k$ of suffix-closed languages, which are designed to be recognized by small parallel monitors, but such that each regular monitor recognizing $L_A^k$ must be "large". We fix two numbers $l, k \geq 0$, such that $k = 2^l$. First, we observe that we can encode every string $w \in \{0,1\}^k$ as a string $a_1\alpha_1 a_2\alpha_2 \cdots a_k\alpha_k \in \{0,1\}^{(l+1)\cdot k}$, where $a_1 a_2 \cdots a_k$ is a permutation of $\{0,1\}^l$ and, for all $i$, $\alpha_i \in \{0,1\}$. Then, $a_i\alpha_i$ gives the information that, for $j$ being the number with binary representation $a_i$, the $j$'th position of $w$ holds bit $\alpha_i$. Let

$$W = \left\{ a_1\alpha_1 \cdots a_k\alpha_k \in \{0,1\}^{(l+1)\cdot k} \;\middle|\; \begin{array}{l} \text{for all } 1 \leq i \leq k,\ \alpha_i \in \{0,1\} \text{ and} \\ a_i \in \{0,1\}^l, \text{ and } a_1 \cdots a_k \in \{0,1\}^l! \end{array} \right\}.$$

Let $w, w' \in W$, where $w = a_1\alpha_1 a_2\alpha_2 \cdots a_k\alpha_k$ and $w' = b_1\beta_1 b_2\beta_2 \cdots b_k\beta_k$, and for $1 \leq i \leq k$, $a_i, b_i \in \{0,1\}^l$ and $\alpha_i, \beta_i \in \{0,1\}$. We define $w \equiv w'$ to mean that for every $i, j \leq k$, $a_i = b_j$ implies $\alpha_i = \beta_j$. Let $a = \alpha_0 \ldots \alpha_{l-1} \in \{0,1\}^l$; then, $enc(a) = bin(0)\alpha_0 \ldots bin(l-1)\alpha_{l-1}$ is the ordered encoding of $a$, where $bin(i)$ is the binary encoding of $i$. Then, $w \in W$ is called an encoding of $a$ if $w \equiv enc(a)$.

Let $\Sigma = \{0, 1, \#\}$ and $\Sigma_\$ = \Sigma \cup \{\$\}$. Then,

$$L_A^k = \{u\#w\#v\$u'\#w'\#\$v' \mid u, v' \in \Sigma_\$^*, u', v \in \Sigma^*,\ w, w' \in W, \text{ and } w \equiv w'\}.$$

In other words, a finite trace is in $L_A^k$ exactly when it has a substring of the form $\#w\#v\$u'\#w'\#\$$, where $w$ and $w'$ are encodings of the same string and there is only one \$ between them. Intuitively, # is there to delimit bit-strings that may be elements of $W$, and \$ delimits sequences of such bit-strings. So, the language asks if there are two such consecutive sequences where the last bit-string of the second sequence comes from $W$ and matches an element from the first sequence. We observe that $L_A^k$ is suffix-closed.

**Lemma 5.** *$s \in L_A^k$ if and only if $\forall t.\ st \in L_A^k \cdot \Sigma_\$^\omega$.*

*Conventions* For the conclusions of Lem. 3 to hold, monitors need to be reactive. However, a reactive monitor can have a larger syntactic description than an equivalent non-reactive one, *e.g.*, $\alpha.\mathsf{yes}$ *vs.* $\alpha.\mathsf{yes} + \beta.\mathsf{end} + \gamma.\mathsf{end}$, when $\text{Act} = \{\alpha, \beta, \gamma\}$. This last monitor is also verdict equivalent to $\alpha.\mathsf{yes} + \mathsf{end}$. In what follows, for brevity and clarity, whenever we write a sum $s$ of a monitor of the form $\alpha.m$, we will mean $s + \mathsf{end}$, which is reactive, so it can be safely used with a parallel operator, and is verdict equivalent to $s$. We use set-notation for monitors: for $A \subseteq \text{Act}$, $A.m$ stands for $\sum_{\alpha \in A} \alpha.m$ (or $\sum_{\alpha \in A} \alpha.m + \mathsf{end}$ under the above convention). Furthermore, we define $\{0,1\}^0.m = m$ and $\{0,1\}^{i+1}.m = 0.\{0,1\}^i.m + 1.\{0,1\}^i.m$. Notice that $l(\{0,1\}^i.m) = 2^i \cdot l(m) + 5 \cdot (2^i - 1)$. We can also similarly define $T.m$ for $T \subseteq \{0,1\}^i$.

*Auxiliary monitors* We start by defining auxiliary monitors. Given a (closed) monitor $m$, let

$$\begin{aligned} skip_{\#}(m) &:= \mathsf{rec}\, x.((0.x + 1.x + \#.x) \oplus \#.m), \\ next_{\#}(m) &:= \mathsf{rec}\, x.(0.x + 1.x + \#.m), \\ next_{\$}(m) &:= \mathsf{rec}\, x.(0.x + 1.x + \#.x + \$.m), \quad \text{and} \\ skip\_last(m) &:= \mathsf{rec}\, x.(0.x + 1.x + \#.x + \#.(m \otimes \mathsf{rec}\, y.(0.y + 1.y + \#.\$.\mathsf{yes}))). \end{aligned}$$

These monitors read the trace until they reach a certain symbol, and then they activate submonitor $m$. We can think that $skip_{\#}(m)$ nondeterministically skips to some occurrence of # that comes before the first occurrence of \$; $next_{\#}(m)$ and $next_{\$}(m)$ respectively skip to the next occurrence of # and \$; and $skip\_last(m)$ skips to the last occurrence of # before the next occurrence of #\$.

**Lemma 6.** *$skip_{\#}(m)$ accepts $g$ iff there are $s$ and $h$, such that $s\#h = g$, $m$ accepts $h$, and $s \in \{0, 1, \#\}^*$.*

The following lemmata are straightforward and explain how the remaining monitors defined above are used.

**Lemma 7.** *$next_{\#}(m)$ accepts $g$ iff there are $s$ and $h$, such that $s\#h = g$, $m$ accepts $h$, and $s \in \{0, 1\}^*$.*

**Lemma 8.** *$next_{\$}(m)$ accepts $g$ iff there are $s$ and $h$, such that $s\$h = g$, $m$ accepts $h$, and $s \in \{0, 1, \#\}^*$.*

**Lemma 9.** *$skip\_last(m)$ accepts $g$ iff there are $s, r$, and $h$, such that $s\#r\#\$h = g$, $m$ accepts $r\#\$h$, $r \in \{0, 1\}^*$, and $s \in \{0, 1, \#\}^*$.*

The following monitors help us ensure that a bit-string from $\{0, 1\}^{(l+1)\cdot k}$ is actually a member of $W$. Monitor *all* ensures that all bit positions appear in the bit-string; $no\_more(s)$ assures us that the bit position $s$ does not appear any more in the remainder of the bit-string; and *unique* guarantees that each bit position appears at most once. Monitor *perm* combines these monitors together.

$$all := \otimes\{\mathsf{rec}\, x.(\{0, 1\}^{l+1}.x \oplus s.\{0, 1\}.\mathsf{yes}) \mid s \in \{0, 1\}^l\};$$

for $s \in \{0, 1\}^l$,

$$no\_more(s) := \mathsf{rec}\, x.\left(\#.\mathsf{yes} + (\{0, 1\}^l \setminus \{s\}).\{0, 1\}.x\right);$$

$$unique := \mathsf{rec}\, x.\left(\#.\mathsf{yes} + \left(\{0, 1\}^{l+1}.x \otimes \sum_{s \in \{0,1\}^l} s.\{0, 1\}.no\_more(s)\right)\right); \text{ and}$$

$$perm := all \otimes unique.$$

The purpose of *perm* is to ensure that a certain block of bits before the appearance of the # symbol is a member of the set $W$: it accepts $w\#$ exactly

when $w$ is a sequence of blocks of bits with length exactly $l+1$ (by *unique*) and for every $a \in \{0,1\}^l$ there is some $\alpha \in \{0,1\}$ such that $a\alpha$ is one of these blocks (by *all*), and that for each such $a$ only one block is of the form $a\alpha'$ (by *unique*).

**Lemma 10.** *perm accepts $g$ iff $w\#$ is a prefix of $g$, for some $w \in W$.*

**Lemma 11.** *$l(perm) = O(k^2)$.*

Given a block $s$ of $l+1$ bits, monitor $find(s)$ accepts a sequence of blocks of $l+1$ bits $w$ exactly when $s$ is one of the blocks of $w$:

$$find(s) := \mathsf{rec}\, x.(s.\mathsf{yes} + (\{0,1\}^{l+1} \setminus \{s\}).x).$$

**Lemma 12.** *For $s \in \{0,1\}^{l+1}$, $find(s)$ accepts $g$ if and only if there is some $r \in (\{0,1\}^{l+1})^*$, such that $rs$ is a prefix of $g$.*

For $s \in \{0,1\}^{l+1}$, $match(s)$ ensures that right before the second occurrence of \$, there is a $\#w\#$, where $w \in (\{0,1\}^{l+1})^+$ and $s$ is a $(l+1)$-bit block in $w$.

$$match(s) := next_{\$}(skip\_last(find(s))).$$

**Lemma 13.** *For $s \in \{0,1\}^{l+1}$, $match(s)$ accepts $g$ if and only if there are $r\$r'\#w\#\$h = g$, such that $r, r' \in \Sigma^*$, $w \in \{0,1\}^*$, and there is a prefix $w's$ of $w$, such that $w' \in (\{0,1\}^{l+1})^*$.*

*Recognizing $L_A^k$ with a parallel monitor* We can now define a parallel acceptance-monitor of length $O(k^2)$ that recognizes $L_A^k$. Monitor *matching* ensures that every one of the consecutive blocks of $l+1$ bits that follow, also appears in the block of bits that appears right before the occurrence of \#\$ that follows the next \$ (and that there is no other \$ between these \$ and \#\$). Therefore, if what follows from the current position in the trace and what appears right before that occurrence of \#\$ are elements $w, w'$ of $W$, *matching* ensures that $w \equiv w$. Then, $m_A^k$ nondeterministically chooses an occurrence of \#, it verifies that the block of bits that follows is an element $w$ of $W$ that ends with \#, and that what follows is of the form $v\$u'\#w'\#\$v'$, where $u', v \in \Sigma^*$, $w' \in W$, and $w \equiv w'$, which matches the description of $L_A^k$.

$$matching := \mathsf{rec}\, x. \sum_{s \in \{0,1\}^{l+1}} s.(x \otimes match(s))$$

$$m_A^k := skip_{\#}\,(perm \otimes next_{\$}(skip\_last(perm)) \otimes matching)$$

**Lemma 14.** *$m_A^k$ recognizes $L_A^k$ and $l(m_A^k) = O(k^2)$.*

*Proof.* The lemma follows from this section's previous lemmata and from counting the sizes of the various monitors that we have constructed. □

**Lemma 15.** *If $m$ is a deterministic monitor that recognizes $L_A^k \cdot \Sigma_{\$}^{\omega}$, then*

$$|m| \geq \left((2^{2^{k-1}} - 1)!\right)^2 = 2^{\Omega\left(2^{2^{k-1}+k}\right)}.$$

Thm. 3 gathers our conclusions about $L_A^k$.

**Theorem 3.** *For every $k > 0$, $L_A^k$ is recognized by an alternating automaton of $O(k^2)$ states and a parallel monitor of length $O(k^2)$, but by no DFA with $2^{o(2^k)}$ states and no deterministic monitor of length $2^{o(2^{2^{k-1}+k})}$. $L_A^k \cdot \Sigma_\$^\omega$ is recognized by a parallel monitor of length $O(k^2)$, but by no deterministic monitor of length $2^{o(2^{2^{k-1}+k})}$.*

*Proof.* Lem. 14 informs us that there is a parallel monitor $m$ of length $O(k^2)$ that recognizes $L_A^k$. Therefore, it also recognizes $L_A^k \cdot \Sigma_\$^\omega$ by Lem. 1. Prop. 1 tells us that $m$ can be turned into an alternating automaton with $l(m) = O(k^2)$ states that recognizes $L_A^k$. Lem. 15 yields that there is no deterministic monitor of length $2^{o(2^{2^{k-1}+k})}$ that recognizes that language. From [3], we know that if there were a DFA with $2^{o(2^k)}$ states that recognizes $L_A^k$, then there would be a deterministic monitor of length $2^{2^{o(2^k)}}$ that recognizes $L_A^k$, which, as we argued, cannot exist. □

*Hardness for regular monitors* Proposition 3 does not guarantee that the $2^{O(n \cdot 2^n)}$ upper bound for the transformation from a parallel monitor to a nondeterministic regular monitor is tight. To prove a tighter lower bound, let $L_U^k$ be the language that includes all strings of the form $\#w_1\#w_2\#\cdots\#w_n\$w$ where for $i = 1, \ldots, n$, $w_i \in W$, and $w \in \{0, 1, \#, \$\}^*$, and for every $i < n$, $w_i$ encodes a string that is smaller than the string encoded by $w_{i+1}$, in the lexicographic order.

**Lemma 16.** *$s \in L_U^k$ if and only if $\forall t.\ st \in L_U^k \cdot \Sigma_\$^\omega$.*

We describe how $L_U^k$ can be recognized by a parallel monitor of size $O(k^2)$. The idea is that we need to compare the encodings of two consecutive blocks of $l + 1$ bits. Furthermore, a string is smaller than another if there is a position in these strings, where the first string has the value 0 and the second 1, and for every position that comes before that position, the bits of the two strings are the same. We define the following monitors:

$$smaller = \sum_{s \in \{0,1\}^l} \left( \begin{array}{c} find(s0) \otimes next_\#(find(s1)) \\ \otimes \\ \bigotimes_{r<s} \sum_{b \in \{0,1\}} (find(rb) \otimes next_\#(find(rb))) \end{array} \right)$$

$$last = \mathsf{rec}\, x.(0.x + 1.x + \$.\mathsf{yes})$$

$$m_U = \mathsf{rec}\, x.(next_\#(perm \otimes (last \oplus (smaller \otimes x))))$$

**Proposition 5.** *$m_U$ recognizes $L_U^k$ (and $L_U^k \cdot \Sigma_\$^\omega$) and $|m_U| = O(k^2)$. Every regular monitor that recognizes $L_U^k$ or $L_U^k \cdot \Sigma_\$^\omega$ must be of length $2^{\Omega(2^k)}$.*

## 5 Logical Consequences

We now turn our attention back from the two monitoring systems to the corresponding logical fragments. We observe that the bounds that we have proved in the previous sections also apply when we discuss formula translations. A version of Thm. 4 was proven in [4], but without complexity bounds.

**Theorem 4.** *For every $\varphi \in$ mnHML (resp., $\varphi \in$ mxHML), there is some $\psi \in$ cHML (resp., $\psi \in$ sHML), such that $l(\psi) = 2^{O(l(m)\cdot 2^{l(m)})}$ and $[\![\varphi]\!] = [\![\psi]\!]$.*

*Proof.* We prove the case for $\varphi \in$ mnHML, as the case for $\varphi \in$ mxHML is similar. By Thm. 1, we know that there is a reactive parallel monitor $m$, such that $L_a(m) \cdot \text{Act}^\omega = [\![\varphi]\!]$ and $l(m) = O(l(\varphi))$. By Prop. 2, we know that there is a regular monitor $n$, such that $L_a(n) = L_a(m)$ and $l(n) = 2^{O(l(m)\cdot 2^{l(m)})}$. We can then see that $l(n) = 2^{O(l(\varphi)\cdot 2^{l(\varphi)})}$. According to Thm. 1, there is a formula $\psi \in$ cHML, such that $[\![\psi]\!] = L_a(n) \cdot \text{Act}^\omega = L_a(m) \cdot \text{Act}^\omega = [\![\varphi]\!]$, and $l(\psi) = O(l(n))$, yielding that $l(\psi) = 2^{O(l(\varphi)\cdot 2^{l(\varphi)})}$. □

The cost of the construction in the proof of Thm. 4 is due to the regularization of the monitor. Our lower bounds — and specifically Prop. 5 — demonstrate that this construction is optimal, because a better construction of $\psi$ from $\varphi$ would lower the cost of regularization via the synthesis functions.

**Theorem 5.** *There is some $\varphi \in$ mxHML, such that for every $\psi \in$ sHML, if $[\![\varphi]\!] = [\![\psi]\!]$, then $l(\psi) = 2^{\Omega\left(2^{\sqrt{l(\varphi)}}\right)}$.*

*Proof (Sketch).* Otherwise, we could regularize $m_U$ from Sec. 4 more efficiently than Prop. 5 allows, by first turning $m_U$ to $\varphi \in$ mxHML, then to $\psi \in$ sHML, and finally to a regular monitor $m$. The full proof is in the appendix. □

*Remark 2.* We observe that to prove Thm. 5, it was necessary to prove Prop. 5 for regular monitors that are *$\omega$-verdict equivalent*, and not just verdict equivalent, to $m_U$. The reason is that, in the proof of Thm. 5, the monitor $m$ that monitors for $\psi$ is $\omega$-verdict equivalent to $m_U$ and there is no guarantee that it is, in fact, verdict equivalent to $m_U$. □

*Remark 3.* In [2], the authors define a deterministic fragment of sHML, which they then show to be equivalent to the full sHML. We can claim analogous bounds for translating formulae into this smaller fragment, using similar arguments to those used above. We omit a full exposition of this claim. □

## 6 Conclusion

We determined the cost of turning a parallel monitor into an equivalent regular, or deterministic, monitor. As a result, we saw that, over infinite traces, mxHML is doubly-exponentially more succinct than sHML.

Regular monitors were introduced in [12] to monitor for sHML over processes. The cost of determinization of regular monitors was examined in [2, 3]. Aceto *et al.* in [6] used a similar determinization process on formulae in the context of enforcement.

In [4], we also synthesized *tight* monitors, which are monitors that reach a verdict as soon as they have analysed enough information from the trace, and not later. It is often important to reach a verdict as soon as possible, but it is also important to avoid burdening a monitored system with a very large monitor. Therefore, it would also be of interest to determine how much it costs to turn a parallel or regular monitor into a verdict-equivalent tight monitor. This is a topic that we leave for future work.

## References


1. Luca Aceto, Antonis Achilleos, Adrian Francalanza, and Anna Ingólfsdóttir. Monitoring for silent actions. In Satya Lokam and R. Ramanujam, editors, *FSTTCS*, volume 93 of *LIPIcs*, pages 7:1–7:14, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
2. Luca Aceto, Antonis Achilleos, Adrian Francalanza, Anna Ingólfsdóttir, and Sævar Örn Kjartansson. Determinizing monitors for HML with recursion. *CoRR*, abs/1611.10212, 2016.
3. Luca Aceto, Antonis Achilleos, Adrian Francalanza, Anna Ingólfsdóttir, and Sævar Örn Kjartansson. On the complexity of determinizing monitors. In Arnaud Carayol and Cyril Nicaud, editors, *Implementation and Application of Automata - 22nd International Conference, CIAA 2017*, volume 10329 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2017.
4. Luca Aceto, Antonis Achilleos, Adrian Francalanza, Anna Ingólfsdóttir, and Karoliina Lehtinen. Adventures in monitorability: From branching to linear time and back again. *Proceedings of the ACM on Programming Languages*, 3(POPL):52:1–52:29, 2019.
5. Luca Aceto, Antonis Achilleos, Adrian Francalanza, Anna Ingólfsdóttir, and Karoliina Lehtinen. Adventures in monitorability: From branching to linear time and back again. *CoRR*, abs/1902.00435, 2019.
6. Luca Aceto, Ian Cassar, Adrian Francalanza, and Anna Ingólfsdóttir. On Runtime Enforcement via Suppressions. In Sven Schewe and Lijun Zhang, editors, *29th International Conference on Concurrency Theory (CONCUR 2018)*, volume 118 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
7. Luca Aceto, Anna Ingólfsdóttir, Kim Guldstrand Larsen, and Jiri Srba. *Reactive Systems: Modelling, Specification and Verification*. Cambridge Univ. Press, New York, NY, USA, 2007.
8. Andreas Bauer, Martin Leucker, and Christian Schallhart. Runtime verification for LTL and TLTL. *ACM Transactions on Software Engineering and Methodology*, 20(4):14, 2011.
9. Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. Alternation. *Journal of the ACM*, 28(1):114–133, jan 1981.
10. Yliès Falcone, Jean-Claude Fernandez, and Laurent Mounier. What can you verify and enforce at runtime? *International Journal on Software Tools for Technology Transfer*, 14(3):349–382, Jun 2012.

11. A. Fellah, H. Jürgensen, and S. Yu. Constructions for alternating finite automata∗. *International Journal of Computer Mathematics*, 35(1-4):117–132, jan 1990.
12. Adrian Francalanza, Luca Aceto, and Anna Ingólfsdóttir. Monitorability for the Hennessy-Milner logic with recursion. *Formal Methods in System Design*, 51(1):87–116, 2017.
13. Kim G. Larsen. Proof Systems for Satisfiability in Hennessy-Milner Logic with recursion. *Theoretical Computer Science*, 72(2):265 – 288, 1990.
14. Amir Pnueli and Aleksandr Zaks. PSL model checking and run-time verification via testers. In Jayadev Misra, Tobias Nipkow, and Emil Sekerinski, editors, *FM 2006: Formal Methods, 14th International Symposium on Formal Methods*, volume 4085 of *Lecture Notes in Computer Science*, pages 573–586. Springer, 2006.

## Appendix

This appendix gathers the omitted proofs from the main text of this paper.

### Omitted Proofs from Sec. 3

*Proof (of Cor. 1).* The alternating automaton that is constructed in the proof of Prop. 1 has at most as many states as there are submonitors in $m$ which, in turn, are no more than $l(m)$. To conclude, every alternating automaton with $k$ states can be converted into an NFA with at most $2^k$ states that recognises the same language [9, 11]. □

The following technical lemmata help in proving the improved upper bound in Prop. 4.

**Lemma 17.** *Let $m$ be a regular monitor and $n$ a submonitor of $m$, such that $n$ is a sum of $\alpha.\mathsf{yes}$ and of $\beta.\mathsf{yes}$. Then, there is some $s \in \text{Act}^*$, such that $m$ accepts both $s\alpha$ and $s\beta$.*

*Proof.* By straightforward induction on the construction of $m$ from $n$. □

*Remark 4.* Monitor $m_\tau$ from Remark 1 demonstrates that Lem. 17 does not hold for parallel monitors, and so does $\mathsf{no} \otimes (a.\mathsf{yes} + b.\mathsf{yes})$. □

**Lemma 18.** *Let $m$ be a closed regular monitor that does not have any submonitor of the form $\mathsf{rec}\, x.\mathsf{yes}$, and let $m'$ be a submonitor of $m$. If $m' \Rightarrow \mathsf{yes}$, then $m' = \mathsf{yes}$.*

*Proof.* If $m' \Rightarrow \mathsf{yes}$, then $m' \xrightarrow{\tau}{}^{k} \mathsf{yes}$ for some $k \geq 0$. We prove that if $k > 0$, then $\mathsf{rec}\, x.\mathsf{yes}$ is a submonitor of $m$. If $k = 1$, then $m'$ is a sum of $\mathsf{rec}\, x.\mathsf{yes}$, and so $\mathsf{rec}\, x.\mathsf{yes}$ is a submonitor of $m'$, and therefore of $m$. If $m' \xrightarrow{\tau} n$ and $n \xrightarrow{\tau}{}^{k-1} \mathsf{yes}$, then by Lem. 2, $n$ is a submonitor of $m$, so $\mathsf{rec}\, x.\mathsf{yes}$ is a submonitor of $n$ by induction, and therefore of $m$. □

*Proof (of Prop. 4).* We recall that, in the proof of Prop. 3 given in [4], the construction starts from a consistent reactive parallel monitor $m$ and turns it into two NFAs $A_a$ and $A_r$, such that $L(A_a) = L_a(m)$, $L(A_r) = L_r(m)$, and the number of states in $A_a$ and in $A_r$ is at most $2^{O(l(m))}$ (Cor. 1). Then, these automata can be turned into regular monitors $m_a$ and $m_r$, and we can conclude by determinizing $m_R^a + m_R^r$. To improve the upper bound for the construction, we determinize starting from the NFAs. Let $A_a = (Q_a, \text{Act}, \delta_a, q_0^a, F_a)$ and $A_r = (Q_r, \text{Act}, \delta_r, q_0^r, F_r)$.

We first construct a new NFA[5], $N = (Q, \textsc{Act} \cup \{\heartsuit, \spadesuit, \varepsilon\}, \delta, q_0, F)$, such that: $Q = \{q_0, f\} \sqcup Q_a \sqcup Q_r$, where $\sqcup$ stands for disjoint union; $F = \{f\}$; and

$$\delta(q,\alpha) = \begin{cases} \{q_0^a, q_0^r\} & \text{for } q = q_0,\ \alpha = \varepsilon, \\ \delta_o(q,\alpha) & \text{for } q \in Q_o,\ \alpha \in \textsc{Act} \cup \{\varepsilon\},\ o \in \{a, r\}, \\ \{f\} & \text{for } q \in F_a,\ \alpha = \heartsuit, \\ \{f\} & \text{for } q \in F_r,\ \alpha = \spadesuit, \\ \{f\} & \text{for } q = f, \\ \emptyset & \text{otherwise.} \end{cases}$$

It is not hard to see that the set of minimal traces in $L(N)$, with respect to the prefix order, is $\min L(N) = L(A_a) \cdot \{\heartsuit\} \ \cup \ L(A_r) \cdot \{\spadesuit\}$.

From [3, Theorem 2], there is a deterministic (closed) monitor $n'$, such that $L_a(n') = L(N)$ and $l(n') = 2^{2^{O(|Q|)}} = 2^{2^{2^{O(l(m))}}}$. We observe that there cannot be a submonitor of $n'$ that is a sum both of $\heartsuit.\mathsf{yes}$ and of $\spadesuit.\mathsf{yes}$ — otherwise, by Lem. 17, $L_a(m) \cap L_r(m) \neq \emptyset$, which is a contradiction, because $m$ is consistent. We further assume that $n'$ has no submonitors of the form $\mathsf{rec}\, x.\mathsf{yes}$, $\mathsf{rec}\, x.\mathsf{no}$, as these can be replaced by $\mathsf{yes}$ and $\mathsf{no}$, respectively, yielding an equivalent monitor. Let $n$ be the result of replacing in $n'$ all maximal sums of $\heartsuit.\mathsf{yes}$ by $\mathsf{yes}$ and of $\spadesuit.\mathsf{yes}$ by $\mathsf{no}$. In particular, there are monitor variables $x_1, \ldots, x_k, y_1, \ldots, y_l$ that do not appear in $n'$ and an open monitor $n_o(x_1, \ldots, x_k, y_1, \ldots, y_l)$, such that $n' = n_o(s_1, \ldots, s_k, s'_1, \ldots, s'_l)$, where $s_1, \ldots, s_k$ are (all the occurrences of) sums of $\heartsuit.\mathsf{yes}$ in $n'$, $s'_1, \ldots, s'_l$ are (all the occurrences of) sums of $\spadesuit.\mathsf{yes}$ in $n'$, and $n = n_o(\mathsf{yes}, \ldots, \mathsf{yes})$. We prove that $\min L_a(n') = L_a(n) \cdot \{\heartsuit\} \ \cup \ L_r(n) \cdot \{\spadesuit\}$.

Let $w \in \min L_a(n')$. Because $L_a(n') = L(N)$, we know that $w \in L(N)$, and therefore $w \in L(A_a) \cdot \{\heartsuit\} \ \cup \ L(A_r) \cdot \{\spadesuit\}$. So, it is either the case that $w = w'\heartsuit$ for $w' \in L_a(A_a)$, or that $w = w'\spadesuit$ for $w' \in L_a(A_a)$ — w.l.o.g. we assume the first case. Therefore, $n' \xRightarrow{w'\heartsuit} \mathsf{yes}$ and since $w \in \min L_a(n')$, $n' \xRightarrow{w'} \not\Rightarrow \mathsf{yes}$. We demonstrate that $n \xRightarrow{w'} \mathsf{yes}$. Specifically, *we prove by induction on an arbitrary submonitor* $n'_o = n'_o(x_1, \ldots, x_k, y_1, \ldots, y_l)$ *of* $n_o(x_1, \ldots, x_k, y_1, \ldots, y_l)$ *that if* $n_s = n'_o(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \xRightarrow{w'\heartsuit} \mathit{yes}$ *and* $n_s \xRightarrow{w'} \not\Rightarrow \mathit{yes}$ *for some* $w' \in \textsc{Act}^*$, *then* $n_y = n'_o(\mathsf{yes}, \ldots, \mathsf{yes}) \xRightarrow{w'} \mathit{yes}$.

**The base cases are:** $n'_o$ is a verdict, which is immediate, because the claim's assumptions cannot both hold; or $n'_o = x_i$, in which case $n_y = \mathsf{yes}$.

**If for some** $\alpha \in \mathbf{Act} \cup \{\heartsuit, \spadesuit\}$, $n'_o = \alpha.n''$ and $n_s \xRightarrow{w'\heartsuit} \mathsf{yes}$ and $n_s \xRightarrow{w'} \not\Rightarrow \mathsf{yes}$, then there are two cases. In the first case, $\alpha = \heartsuit$. Then, as $w' \in \textsc{Act}^*$, we have that $w' = \varepsilon$, so $n''(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \Rightarrow \mathsf{yes}$, and by Lem. 18 ($n_s$ is a submonitor of $n'$), $n''(s_1, \ldots, s_k, s'_1, \ldots, s'_l) = \mathsf{yes}$, and therefore $n'_o = \heartsuit.\mathsf{yes}$, which contradicts our definitions. Otherwise, for $\alpha \neq \heartsuit$, $w' = \alpha\, w''$, and $n''(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \xRightarrow{w''\heartsuit} \mathsf{yes}$ and $n''(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \xRightarrow{w''} \not\Rightarrow \mathsf{yes}$. By the inductive hypothesis, $n''(\mathsf{yes}, \ldots, \mathsf{yes}) \xRightarrow{w''} \mathsf{yes}$, and therefore $n_y \xRightarrow{w'} \mathsf{yes}$.

[5] For clarity, the construction of the NFA $N$ makes a mild use of $\varepsilon$-transitions.

**If** $n'_o = n_1 + n_2$ and $n_s \stackrel{w'\heartsuit}{\Longrightarrow} \mathsf{yes}$ and $n_s \stackrel{w'}{\not\Longrightarrow} \mathsf{yes}$, then we know that $n_s$ is not a sum of $\heartsuit.\mathsf{yes}$. Therefore, w.l.o.g. $n_1(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \stackrel{w'\heartsuit}{\Longrightarrow} \mathsf{yes}$ and $n_1(s_1, \ldots, s_k, s'_1, \ldots, s'_l) \stackrel{w'}{\not\Longrightarrow} \mathsf{yes}$, and we are done by the inductive hypothesis.

**The case for** $n'_o = \mathbf{rec}\, x.n''$ is more straightforward, as $n'_o$ and $n''$ have the same weak transitions.

We now prove that if $w \in L_a(n) \cdot \{\heartsuit\} \cup L_r(n) \cdot \{\spadesuit\}$, then $w \in \min L_a(n')$. Let $w \in L_a(n) \cdot \{\heartsuit\}$ (the case for $w \in L_r(n) \cdot \{\spadesuit\}$ is symmetric). Then there is a $w' \in L_a(n)$, such that $w = w'.\heartsuit$. Let $w''$ be the shortest prefix of $w'$, such that $w'' \in L_a(n)$. We observe that if $w''.\heartsuit \in \min L_a(n')$, then $w''.\heartsuit \in \min L(N)$, and therefore $w'' \in L(A_a)$, yielding that $w' \in L(A_a)$, because $A_a$ is suffix-closed (by verdict-persistence, Lem. 1), and therefore $w = w'.\heartsuit \in \min L_a(n')$. Furthermore, if $w''.\heartsuit \in L_a(n')$, we can see that $w''.\heartsuit \in \min L_a(n')$, as $w'' \in \textsc{Act}^*$. Thus, it suffices to prove that $w''.\heartsuit \in L_a(n')$. By induction on the derivation, and due to the minimality of $w''$, we can see that either $n_o(x_1, \ldots, x_k, y_1, \ldots, y_l) \stackrel{w''}{\Longrightarrow} \mathsf{yes}$, in which case $n' \stackrel{w''}{\Longrightarrow} \mathsf{yes}$, and therefore $n' \stackrel{w''.\heartsuit}{\Longrightarrow} \mathsf{yes}$, or $n_o(x_1, \ldots, x_k, y_1, \ldots, y_l) \stackrel{w''}{\Longrightarrow} x_i$ for some $x_i$, in which case $n' \stackrel{w''}{\Longrightarrow} s'_i$, where $s'_i$ is $s_i$ after applying a number of steps substituting variables for monitors. As $s_i$ is a sum of $\heartsuit.\mathsf{yes}$, so is $s'_i$, therefore $s'_i \stackrel{\heartsuit}{\longrightarrow} \mathsf{yes}$.

We just demonstrated that $\min L_a(n') = L_a(n) \cdot \{\heartsuit\} \cup L_r(n) \cdot \{\spadesuit\}$. But we have also seen that $L_a(n') = L(N)$ and that $\min L(N) = L(A_a) \cdot \{\heartsuit\} \cup L(A_r) \cdot \{\spadesuit\}$, and therefore $L_a(n) = L(A_a) = L_a(m)$ and $L_r(n) = L(A_r) = L_r(m)$, which is what we wanted to show. □

### Omitted Proofs from Sec. 4

*Proof (of Lem. 5).* The "only if" direction is immediate, and therefore we only show the "if" direction. Let $s$ be such that $\forall t.\ st \in L^k_A \cdot \Sigma^\omega_\$$. From our assumptions about $s$, $s\#^\omega \in L^k_A \cdot \Sigma^\omega_\$$. According to the definition of $L^k_A$, there must be a finite prefix $r$ of $s\#^\omega$, such that $r \in L^k_A$ and ends with \$. Since the only symbol in $\#^\omega$ is $\#$, $r$ is a prefix of $s$. Therefore, since $L^k_A$ is suffix-closed, we conclude that $s \in L^k_A$. □

*Proof (of Lem. 6).* The "if" direction is straightforward, using Lem. 3. For the "only if" direction, notice that the $\#.m$ component of the monitor is the only component that can produce a $\mathsf{yes}$ verdict, and it activates at the occurrences of the $\#$ symbol. Therefore, if $skip_\#(m)$ accepts $g$, there must be some $s\#h = g$, such that $m$ accepts $h$. Let $s$ be minimal for this to happen.

We now prove by contradiction that $s \in \{0, 1, \#\}^*$. If that is not the case, \$ appears in $s$, so there are $s_1\$s_2 = s$. When reading $s_1\$$, all monitor components of $skip_\#(m)$ must fail (reach $\mathsf{end}$), except the ones that have come from $\#.m$. Therefore, we can split $g$ as $s_1\#h' = g$, such that $m$ accepts $h'$, which is a contradiction, due to the assumed minimality of $s$. Therefore, $s \in \{0, 1, \#\}^*$. □

*Hardness of $L_A^k$ for deterministic monitors* We introduce the notion of a simple trace. As Lem. 19 reveals, the cardinality of a set of simple traces gives a lower bound on the size of a regular monitor.

**Definition 11.** *We call a derivation $m \stackrel{s}{\Rightarrow} m'$ simple, if rules* RecB *and* Ver *are not used in the proof of any transition of the derivation. We say that a trace $s \in \Sigma_{\$}^*$ is simple for monitor $m$ if there is a simple derivation $m \stackrel{s}{\Rightarrow} m'$. We say that a set of simple traces for $m$ is simple for $m$.* ■

**Lemma 19 (Lemma 20 from [2]).** *Let $m$ be a regular monitor and $G$ a (finite) simple set of traces for $m$. Then, $l(m) \geq |G|$.* □

The following Lems. 20 and 21 are variations of the Pumping Lemma for monitors.

**Lemma 20 (Lemma 22 from [2]).** *Let $m \stackrel{s}{\Rightarrow} n$, such that $m$ is regular and $s$ is not simple for $m$. Then, there are $s = s_1 s_2 s_3$, such that $|s_2| > 0$ and for every $i \geq 0$, $m \stackrel{s_1 s_2^i s_3}{\Longrightarrow} n$.* □

**Lemma 21 (Lemma 27 from [2]).** *Let $m \stackrel{s}{\Rightarrow} m'$, such that $s$ is not simple for $m$ and $m$ is deterministic (and regular). Then, there are $s = s_1 s_2 s_3$ and monitor $n$, such that $|s_2| > 0$ and $m \stackrel{s_1}{\Longrightarrow} n \stackrel{s_2}{\Longrightarrow} n \stackrel{s_3}{\Longrightarrow} m'$.*

Fix a partition $\{C, D\}$ of $\{0,1\}^k$, such that $|C| = |D| = 2^{k-1}$; let $K = 2^{2^{k-1}}$. Let $P_C = C_0 \cdots C_{K-2}$ be a permutation of $2^C \setminus \{\emptyset\}$, and $P_D = D_0 \cdots D_{K-2}$ a permutation of $2^D \setminus \{\emptyset\}$. For every $0 \leq i < K-1$, let

$$t_i = \#enc(w_0)\#enc(w_1)\#\cdots\#enc(w_{|C_i|})\#,$$

where $w_0 w_1 \cdots w_{|C_i|}$ is a permutation of $C_i$ and

$$s_i = \#enc(w'_0)\#enc(w'_1)\#\cdots\#enc(w'_{|D_i|})\#,$$

where $w'_0 w'_1 \cdots w'_{|D_i|}$ is a permutation of $D_i$. Thus, each $t_i$ and $s_i$ encodes a permutation of a distinct subset of $\{0,1\}^k$. Let

$$t_K(P_C, P_D) = t_0\$s_0\$t_1\$s_1\$\cdots\$t_{K-1}\$s_{K-1}$$

and let

$$T = \{t_K(P_C, P_D) \mid P_C \in (2^C \setminus \{\emptyset\})! \text{ and } P_D \in (2^D \setminus \{\emptyset\})!\}.$$

Notice that $|T| = ((K-1)!)^2$.

The following Lem. 22 characterizes trace $t_K$, as defined in Sec. 4, and is useful in the proof of Cor. 2, which follows immediately from it.

**Lemma 22.** *If $f \neq g$ are prefixes of $t_K$, then there is some $h \in \{0,1,\#,\$\}^*$, such that $fh \in L_A^k$ iff $gh \notin L_A^k$.*

*Proof.* We can assume that $f$ is a proper prefix of $g$. We have the following cases:

**The symbol \$ appears more times in $g$ than in $f$.** Then, for some $f', g' \in \Sigma^*$ and $0 \leq i \leq j$:

- $f = f'$ and ($g = t_0\$s_0\$\cdots\$s_j\$g'$ or $g = t_0\$s_0\$\cdots\$t_j\$g'$); or
- $f = t_0\$s_0\$\cdots\$t_i\$f'$ and $g = t_0\$s_0\$\cdots\$s_j\$g'$; or
- $f = t_0\$s_0\$\cdots\$s_i\$f'$ and $g = t_0\$s_0\$\cdots\$t_j\$g'$; or
- $f = t_0\$s_0\$\cdots\$t_i\$f'$ and $g = t_0\$s_0\$\cdots\$t_j\$g'$ and $i < j$; or
- $f = t_0\$s_0\$\cdots\$s_i\$f'$ and $g = t_0\$s_0\$\cdots\$s_j\$g'$ and $i < j$.

Then, for all these cases we can immediately see that there is some $w \in W$ that appears in one of $t_i, s_i, t_j, s_j$, such that for $h = \#w\#\$$, $fh \in L_A^k$ iff $gh \notin L_A^k$.

**The symbol # appears more times in $g$ than in $f$, but \$ does not.** Then, for some $f', g' \in \{0,1\}^*$,

$$f = t'\$\#enc(w_0)\#enc(w_1)\#\cdots\#enc(w_i)\#f' \quad \text{and}$$
$$g = t'\$\#enc(w_0)\#enc(w_1)\#\cdots\#enc(w_j)\#g'$$

Then, for $h = \$\#enc(w_j)\#\$$, $gh \in L_A^k$ and $fh \notin L_A^k$.

**The symbols \$, # appear the same number of times in $g$ and in $f$.** Then, for some $f', g' \in \{0,1\}^*$,

$$f = t'\$\#enc(w_0)\#enc(w_1)\#\cdots\#enc(w_i)\#f' \quad \text{and}$$
$$g = t'\$\#enc(w_0)\#enc(w_1)\#\cdots\#enc(w_i)\#f'g'$$

and $|g'| > 0$. Therefore, $f'g'$ is a prefix of $enc(w_{i+1})$. Let $h_0$ be such that $f'g'h_0 = enc(w_{i+1})$. Then, for $h = h_0\#\$\#enc(w_{i+1})\#\$$, $gh \in L_A^k$ and $fh \notin L_A^k$. □

**Corollary 2** *If $s \neq r$ are prefixes of $t_K$, then there is some $t \in \Sigma_\$^\omega$, such that $st \in L_A^k \cdot \Sigma_\$^\omega$ iff $rt \notin L_A^k \cdot \Sigma_\$^\omega$.*

*Proof.* By Lems. 5 and 22. □

**Proposition 6.** *If $m$ is a deterministic regular monitor that recognizes $L_A^k \cdot \Sigma_\$^\omega$, then $t_K$ is a simple trace for $m$.*

*Proof.* If $t_K$ is not simple, then by Lem. 21, $t_K = s_1s_2s_3$, such that $|s_2| > 0$ and there is a monitor $n$, such that $m \xRightarrow{s_1} n \xRightarrow{s_2} n$. But according to Cor. 2, there is some $t$, such that $s_1t \in L_A^k \cdot \Sigma_\$^\omega$ iff $s_1s_2t \notin L_A^k \cdot \Sigma_\$^\omega$. This is a contradiction, because $s_1t \in L_A^k \cdot \Sigma_\$^\omega$ iff $n$ accepts $t \cdot \Sigma_\$^\omega$ iff $s_1s_2t \in L_A^k \cdot \Sigma_\$^\omega$. □

*Proof (of Lem. 15).* An immediate consequence of Prop. 6 and Lem. 19. □

*Proof (of Prop. 5).* First, we sketch the proof of the first part of the proposition. We observe that *smaller* accepts a trace $w_1\#w_2t$, where $w_1, w_2 \in W$, exactly when $w_1$ encodes a smaller sequence than $w_2$, with respect to the lexicographic ordering. Then, $m_U$ keeps reading blocks of bits between the # separators, while ensuring that each of these is an element of $W$ (using monitor *perm*), and that it either is the last such block of bits (using monitor *last*), or that it encodes a smaller sequence than the next one (using monitor *smaller*).

It now suffices to prove that, for any ordered sequence $a_1a_2\cdots a_c$ of strings from $\{0,1\}^k$, the finite trace $s = \#enc(a_1)\#enc(a_2)\#\cdots\#enc(a_c)$ is simple for any regular monitor that accepts exactly the infinite extensions of $L_U^k$. If $s$ is not simple for $m$, then, according to Lemma 20, for every $m \overset{s}{\Rightarrow} n$, there are $s = s_1s_2s_3$, such that $|s_2| > 0$ and for every $i \geq 0$, $m \xrightarrow{s_1s_2^is_3} n$. Therefore, if $m \overset{s}{\Rightarrow} n \xrightarrow{\$0^j}$ $\mathsf{yes}$ for some $j \geq 0$, then $m \xrightarrow{s_1s_2^2s_3} n \xrightarrow{\$0^j}$ $\mathsf{yes}$ and $m \xrightarrow{s_1s_2^3s_3} n \xrightarrow{\$0^j}$ $\mathsf{yes}$. If $|s_2|$ is not a multiple of $k(l+1)+1$, then it is not hard to see that we reach a contradiction, as $s_1s_2^2s_3\$0^j \notin L_U^k$, because it includes a block of bits that is not in $W$. Otherwise, $s_2$ is of the form $f\#g$ or $f\#enc(a_i)\#\cdots\#enc(a_{i+j})\#g$, where $fg \in \{0,1\}^{k(l+1)}$ and $j \geq 0$. If $gf \notin W$, then $s_1s_2^2s_3\$0^j \notin L_U^k$, and again, we have a contradiction. But even if $gf \in W$, we see that $gf$ appears twice in $s_1s_2^3s_3$, and again we have a contradiction, as $s_1s_2^3s_3\$0^j \notin L_U^k$, because there appear two elements of $W$ in $s_1s_2^3s_3$ that encode the same string. □

## Omitted Proofs from Sec. 5

*Proof (of Thm. 5).* Let $k$ and $m_U$ be as defined in Sec. 4. Thm. 1 asserts the existence of a formula $\varphi$, such that $l(\varphi) = O(l(m_U))$ and $[\![\varphi]\!] = L_a(m_U)\cdot\Sigma_\$^\omega = L_U^k\cdot\Sigma_\$^\omega$. Now, let $\psi \in$ sHML, where $[\![\varphi]\!] = [\![\psi]\!]$. According to Thm. 1, there is a regular monitor $m$, such that $l(m) = O(l(\psi))$, and $[\![\varphi]\!] = [\![\psi]\!] = L_a(m)\cdot\Sigma_\$^\omega$. Therefore, $m$ recognizes $L_U^k\cdot\Sigma_\$^\omega$, and by Prop. 5, $l(m) = 2^{\Omega(2^k)} = 2^{\Omega(2^{\sqrt{l(\varphi)}})}$. This yields that $l(\psi) = 2^{\Omega(2^k)} = 2^{\Omega(2^{\sqrt{l(\varphi)}})}$. □