Closure and Complexity of Temporal Causality

Mishel Carelli, Bernd Finkbeiner, and Julian Siber
CISPA Helmholtz Center for Information Security
Saarbrücken, Germany
{mishel.carelli, finkbeiner, julian.siber}@cispa.de

***Abstract*—Temporal causality defines what property causes some observed temporal behavior (the effect) in a given computation, based on a counterfactual analysis of similar computations. In this paper, we study its closure properties and the complexity of computing causes. For the former, we establish that safety, reachability, and recurrence properties are all closed under causal inference: If the effect is from one of these property classes, then the cause for this effect is from the same class. We also show that persistence and obligation properties are not closed in this way. These results rest on a topological characterization of causes which makes them applicable to a wide range of similarity relations between computations. Finally, our complexity analysis establishes improved upper bounds for computing causes for safety, reachability, and recurrence properties. We also present the first lower bounds for all of the classes.**

***Index Terms*—Automata, counterfactual reasoning, infinite words, temporal properties, topology**

I. Introduction

Temporal causality is a flavor of counterfactual reasoning that causally relates temporal properties of a given system computation. Given, for instance, the system computation

$$\{start\}\{request_1\}\{request_2\}\{failure\}^\omega \ ,$$

temporal causality can tell us whether the property $\bigcirc request_1$ or $\bigcirc\bigcirc request_2$ is the cause for the property $\Diamond\Box failure$. It generalizes the concept of actual causality [20], [21] to symbolic temporal properties and provides a logician's lens to study reasoning used in a plethora of applications such as explaining verification results [1], [6], [9], attribution of blame in multi-agent systems [12] and explainable AI [5], [35].

According to the theory [10], a causal relationship holds between two properties on a given computation, with respect to a given similarity relation, if both properties are satisfied by the computation, the most similar computations that do not satisfy the cause property do not satisfy the effect property either, and the cause property is the minimal set that qualifies for the previous two conditions. Finkbeiner et al. [15] have recently presented an intuitive order-theoretic reformulation of this: Causes are exactly the largest downward closed set of system computations that satisfy the effect, or – more informally speaking – they describe the set of computations most similar to the given observed computation that continuously satisfy the effect property. With this reformulation, they show that ω-regular effects imply ω-regular causes by giving an algorithm that synthesizes the cause property as a nondeterministic Büchi automaton from a system, computation and effect property with respect to a given (effectively ω-regular) similarity relation. Besides showing that ω-regular properties are in this way *closed under causal inference*, this construction also gives an upper bound on the size of the causal automaton that is roughly exponential in the system size and doubly exponential in the effect size (see Table I for the exact complexity). This blow-up mainly stems from Büchi complementation that is performed twice during the cause synthesis algorithm. In this work, we spin the theoretical aspects of this problem further and conduct a detailed investigation of property classes that are closed under temporal causality, as well as the complexity of constructing temporal causes as automata.

A. Closure Under Temporal Causality

While the closure of ω-regular properties under causal inference is of high practical significance because it suggest a general algorithm for, e.g., constructing explanations of model-checking counterexamples, it also raises intriguing philosophical questions regarding the temporal structure of cause-effect relationships in reactive systems that interact with their environment over a possibly infinite duration. For instance, a fundamental temporal aspect of causal relationships between events is that the cause happens before the effect. A similar trait holds for temporal causality: If the effect is given as a temporal logic formula containing n $\bigcirc$-operators (which is a way of describing a set of concrete events), then the cause can be described by a formula containing at most n $\bigcirc$-operators [3]. Hence, the events described by the cause are guaranteed to happen earlier or at the same time as the events described by the effect. Besides demonstrating this temporal aspect of causal relationships between events, this also gives a complete cause-synthesis algorithm for this fragment based on enumeration [3].

In this paper, we go beyond events as described by the fragment containing only $\bigcirc$-operators to general temporal properties. In this general setting, a comparable notion of temporal precedence does not exist, as general temporal properties may place conditions over a full infinite word: For instance, it is impossible to say that $\Box\Diamond a$ happens before $\Box\Diamond b$. Therefore, we study closure properties along the lines of Manna and Pneuli's hierarchy of temporal properties [31]. The hierarchy organizes temporal properties into classes based on language-theoretic considerations that have a tight connection to concepts in topology, temporal logic, and automata theory. We analyze for which classes membership of the effect property implies membership in the same class for the resulting cause property. Our results and previous ones are illustrated along a recapitulation of the hierarchy of ω-

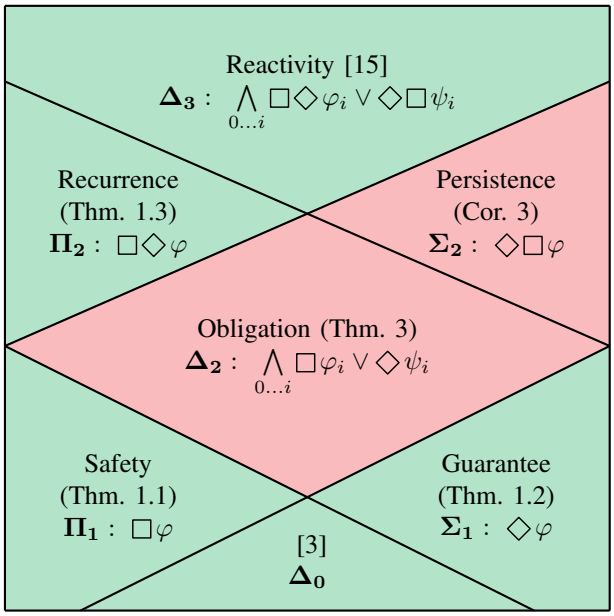


Fig. 1: Results on closure under causality of property classes in Manna and Pneuli's hierarchy of temporal properties [31]. Classes colored green are closed under causal inference, while classes colored in red are not. Note that φ and ψ are formulas containing no future operators.

regular properties in Figure 1. Note that we follow Manna and Pneuli in using LTL operators for the sake of illustration, but mean all ω-regular properties in the respective classes. On the highest level the class of reactivity properties, which is equivalent to the class of ω-regular properties [31], is closed under causality as already shown by Finkbeiner et al. [15]. One level below, things are less clear-cut: While every recurrence effect indeed has a recurrence cause, we show that this is not the case for the class of persistence properties. On the lower levels, we have that safety and guarantee properties are both closed under causality, while obligation properties, which correspond to Boolean combinations of properties from these classes as indicated in the figure, are not closed in the same way. Interestingly, the unnamed class of properties that are both safety and guarantee properties corresponds exactly to the fragment containing only $\bigcirc$-operators for which Beutner et al. have shown closure under causality [3]. Notably, our results require only general assumptions on the similarity relation, such that they can accommodate different relations that may be desirable in different problem settings.

B. Complexity of Temporal Causality

Our study of closure properties along the lines of the hierarchy of temporal properties suggest possible improvements for the synthesis of temporal causes from effects belonging to certain fragments. For instance, since we now know that a cause for a safety effect is itself a safety property, opting for a representation via bad prefixes promises a cheaper complementation operation than in the general case. Therefore, we conduct a detailed inquiry into the complexity of constructing temporal causes for effects from the varying classes.

The main results of this inquiry are shown in Figure I. The exact upper bound obtained from the cause synthesis algorithm of Finkbeiner et al. [15] is also listed in the table. As one of our main results, we can show that there is a family of problems where the cause automaton scales doubly exponential in the size of the effect. While there is still a logarithmic gap between our lower bounds and the upper bound of Finkbeiner et al. [15], this demonstrates that the number of exponents of the upper bounds is already optimal. We have a tight bound for the size of the cause with respect to the size of the system, which is exponential with an additional logarithmic factor in the exponent. These results are particularly valuable because in practical instances such as explaining model-checking results, the system size tends to be much bigger than the effect size.

We can also confirm our initial intuition regarding classes of effects for which cause automata can be synthesized more efficiently than with the general algorithm. As may be expected, this concerns the two classes on the lower level of the hierarchy: guarantee and safety properties. Since properties from these classes can effectively be described by finite word automata for good and bad prefixes, respectively, it is also possible to use the cheaper complementation operations for finite word automata in the general algorithm. We show that this approach results in an upper bound on the size of the cause automaton absent of logarithmic factors, and is mirrored by tight lower bounds in the size of the system and the effect. While the improvement in the upper bound is only by a logarithmic factor, this is still of high practical significance because the classes of guarantee and safety properties are ubiquitous in verification problems.

C. Outline

The paper is structured as follows. We first establish some basic preliminaries (Section II). We then introduce background on temporal causality in Section III. Our contributions are then split into two main parts: In Section IV we establish our results on closure under causal inference using a topological argument, and in Section V we prove lower and upper bounds on the size of causes. We discuss related results in Section VI and end with a short summary and outlook on possible future applications of our results (Section VII).

II. Preliminaries

We recall some general background on infinite words, temporal properties, automata, and temporal logic.

A. Words and Properties

We model computations by words over some *alphabet* Σ. If not discussed explicitly, we assume Σ to be finite. A *word* $\pi = \pi_0\pi_1\ldots$ then is a sequence of letters $\pi_i \in \Sigma$ from the alphabet. For some word π, its prefix of length n is denoted by $\pi_{(n)}$. The set of all prefixes of a given length is defined as $pref(\pi) := \{\pi_{(n)} \mid n \in \mathbb{N}\}$. We denote by Σ^* the set of finite words and by Σ^ω the set of infinite words. $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$

TABLE I: Highlights of our analysis of upper and lower bounds on the size of cause automata synthesized from a system $\mathcal{T}$, computation π, a fixed similarity relation $\leq$ and an effect E, which belongs to a certain class as outlined in the first column.

Effect Class	**Lower bound in E**	**Lower bound in $\mathcal{T}$**	**Upper bound**
Reactivity	$2^{2^{\Omega(\lvert E\rvert)}}$ (Thm. 6)	$2^{\Omega(\lvert\mathcal{T}\rvert\cdot\log(\mathcal{T}\rvert))}$ (Thm. 5.1)	$\lvert\pi\rvert\cdot 2^{2^{\mathcal{O}(\lvert E\rvert\cdot\log(\lvert E\rvert))}\cdot\lvert\mathcal{T}\rvert\cdot\log(\lvert\mathcal{T}\rvert)}$ [15]
Persistence	$2^{2^{\Omega(\lvert E\rvert)}}$ (Thm. 6)	$2^{\Omega(\lvert\mathcal{T}\rvert\cdot\log(\mathcal{T}\rvert))}$ (Thm. 5.1)	$\lvert\pi\rvert\cdot 2^{2^{\mathcal{O}(\lvert E\rvert\cdot\log(\lvert E\rvert))}\cdot\lvert\mathcal{T}\rvert\cdot\log(\lvert\mathcal{T}\rvert)}$ [15]
Recurrence	$2^{2^{\Omega(\lvert E\rvert)}}$ (Thm. 6)	$\Omega(3^{\lvert\mathcal{T}\rvert})$ (Thm. 5.2)	$\lvert\pi\rvert\cdot 3^{2^{\mathcal{O}(\lvert E\rvert\cdot\log(\lvert E\rvert))}\cdot\lvert\mathcal{T}\rvert}$ (Thm. 7)
Safety	$2^{2^{\Omega(\lvert E\rvert)}}$ (Thm. 6)	$\Omega(2^{\lvert\mathcal{T}\rvert})$ (Thm. 5.3)	$\lvert\pi\rvert\cdot 2^{2^{\mathcal{O}(\lvert E\rvert)}\cdot\lvert\mathcal{T}\rvert}$ (Thm. 8)
Guarantee	$2^{2^{\Omega(\lvert E\rvert)}}$ (Thm. 6)	$\Omega(2^{\lvert\mathcal{T}\rvert})$ (Thm. 5.3)	$\lvert\pi\rvert\cdot 2^{2^{\mathcal{O}(\lvert E\rvert)}\cdot\lvert\mathcal{T}\rvert}$ (Thm. 9)

is the set of both finite and infinite words. A *language* L is a subset of Σ^∞. We denote by $\overline{L}$ the complement of a language L. We model *properties* as finite or infinite word languages $L \subset \Sigma^*$ or $L \subset \Sigma^\omega$, respectively. For $L \subseteq \Sigma^\omega$ the set of prefixes of words from L of length n is denoted as $pref_n(L) := \{\pi_{(n)} \mid \pi \in L\}$ and the set of all prefixes of words in L is denoted as: $pref(L) = \{\pi_{(n)} \mid \pi \in L, n \in \mathbb{N}\}$.

We now recall Manna and Pneuli's formal categorization of infinite word languages based on construction rules from finite world languages, which is illustrated in Figure 1.

Definition 1 (Manna and Pneuli [31])**.** *An infinite word language is a safety, guarantee, recurrence or persistence property, if the following holds:*

- *An infinite word language L is a **safety property** if there exists a finite words language Φ such that L consists of all infinite words π such that every prefix of π is in Φ.*
- *An infinite word language L is a **guarantee property** if there exists a finite words language Φ such that L consists of all infinite words π such that there exists a prefix of π that is in Φ.*
- *An infinite word language L is a **recurrence property** if there exists a finite words language Φ such that L consists of all infinite words π such that infinitely many prefixes of π are in Φ.*
- *An infinite word language L is a **persistence property** if there exists a finite words language Φ such that L consists of all infinite words π such that finitely many prefixes of π are in Φ.*
- *An infinite word language L is an **obligation property** if it is a Boolean combination of safety and guarantee properties.*
- *An infinite word language L is a **reactivity property** if it is a Boolean combination of recurrence and persistence properties.*

Moreover, the obligation class is precisely the intersection of the recurrence and persistence classes, and the reactivity class contains all ω-regular properties [31].

B. Automata and Systems

An automaton is a tuple $A = (Q, \Sigma, Q_0, F, \Delta)$, where Q denotes a finite set of *states*, Σ is an alphabet, $Q_0 \subseteq Q$ is a set of *initial states*, $F \subseteq Q$ is the set of *accepting states*, and $\Delta : (Q \times \Sigma) \times Q$ is a *transition relation* that maps a state and a letter to a set of possible successor states. For an automata A the number of states of A is denoted as $|A|$. An automaton $A = (Q, \Sigma, Q_0, F, \Delta)$ is *deterministic* if the transition relation Δ is a function and Q_0 is a singleton, otherwise it is a *nondeterministic* automaton. A *run* of A on a *word* $\pi = \pi_0\pi_1 \ldots \in \Sigma^\infty$ is a sequence $r = q_0q_1 \ldots$ of states $q_i \in Q$ with $q_0 \in Q_0$ and $\big((q_i, w_i), q_{i+1}\big) \in \Delta$ for all i. *Universal automata* are nondeterminstic automata that accept a word π if all of the runs on π fulfill the automatons acceptance condition, for all other automata only one run needs to fulfill this condition. In a *finite word automaton* the acceptance condition for a run $r = q_0q_1 \ldots q_k$ is that $q_k \in F$. We use the shorthands DFW, NFW and UFW for determinstic, nondeterminstic and universal finite word automata, respectively, and analogous shorthands for all other automata types. In a *Büchi word automaton* (DBW, etc.), an infinite run $r = q_0q_1 \ldots$ fulfills the acceptance condition if there exist infinitely many $i \in \mathbb{N}$ such that $q_i \in F$. Lastly, in a *Co-Büchi word automaton* (DCW, etc.), an infinite run $r = q_0q_1 \ldots$ is accepting if all states appearing infinitely often are not in F, i.e., there is an $j \in \mathbb{N}$ such that all $i > j$ have $q_i \notin F$. The *language* $L(A)$ of an automaton A is the set of all words that have an accepting run. An infinite word language $L \subset \Sigma^\omega$ is ω-regular, if it is recognized by a nondeterministic Büchi word automaton (NBW) A such that we have $L(A) = L$.

A *systems* is a tuple $\mathcal{T} = (S, s_0, AP, \delta, l)$ where S is a finite set of *states*, $s_0 \in S$ is the *initial state*, $AP = I \cup O$ consists of *inputs* I and *outputs* O, $\delta : S \times 2^I \to 2^S$ is the *transition function* describing the successor states for some state and input, and $l : S \to 2^O$ is the *labeling function* labeling each state with a set of outputs. A *trace* of $\mathcal{T}$ is an infinite sequence $\pi = \pi_0\pi_1 \ldots \in (2^{AP})^\omega$, with $\pi_i = A \cup l(s_{i+1})$ for some $A \subseteq I$ and $s_{i+1} \in \delta(s_i, A)$ for all $i \geq 0$. Note that the label of the initial state is omitted in the first position. $traces(\mathcal{T})$ is the set of all traces of $\mathcal{T}$. A *zipped trace* of the three traces $\pi^{0,1,2}$ is then defined as $zip(\pi^0, \pi^1, \pi^2)_i = \{a^k \mid a \in \pi^k_i\}$, i.e., we construct the zipped trace from disjoint unions of the positions of the three traces, where inputs and outputs from the

traces $\pi^{0,1,2}$ are distinguished through superscripts. We also define projection and equivalence on traces: for $A, B \subseteq I \cup O$ and traces π, π', let $A|_B = A \cap B$, $\pi|_B = \pi_0|_B \pi_1|_B \ldots$ and $\pi =_A \pi'$ iff $\pi|_A = \pi'|_A$.

C. Linear-time Temporal Logic

We will use *Linear-time Temporal Logic* (LTL) [40] when we want to describe a property more conveniently than with automata in this paper (even though not every ω-regular property can be expressed this way). The grammar for LTL formulas is as follows, where $a \in \Sigma$:

$$\varphi ::= a \mid \neg\varphi \mid \varphi \wedge \varphi \mid \bigcirc \varphi \mid \varphi \mathcal{U} \varphi \mid \bigcirc^- \varphi \mid \varphi \mathcal{U}^- \varphi \ .$$

All temporal operators with a minus superscript are past operators [30], the others are future operators. The semantics of LTL are given as follows.

$$\begin{array}{lll}
\pi, i \vDash a & \text{iff} & a = \pi_i \\
\pi, i \vDash \neg\varphi & \text{iff} & \pi, i \nvDash \varphi \\
\pi, i \vDash \varphi \wedge \psi & \text{iff} & \pi, i \vDash \varphi \text{ and } \pi, i \vDash \psi \\
\pi, i \vDash \bigcirc \varphi & \text{iff} & \pi, i+1 \vDash \varphi \\
\pi, i \vDash \bigcirc^- \varphi & \text{iff} & i > 0 \wedge \pi, i-1 \vDash \varphi \\
\pi, i \vDash \varphi \mathcal{U} \psi & \text{iff} & \exists j \geq i \text{ such that } \pi, j \vDash \psi \text{ and} \\
 & & \forall i \leq k < j. \, \pi, k \vDash \varphi \\
\pi, i \vDash \varphi \mathcal{U}^- \psi & \text{iff} & \exists k \leq i \text{ such that } \pi, k \vDash \psi \text{ and} \\
 & & \forall i \geq j > k : \pi, j \vDash \varphi \ .
\end{array}$$

A word π *satisfies* a formula φ, denoted by $\pi \vDash \varphi$ iff $\pi, 0 \vDash \varphi$, i.e., the formula holds at the first position. The *language* $L(\varphi)$ of a formula φ is the set of all traces that satisfy it. We also use the derived Boolean connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) and temporal operators ($\varphi \mathcal{R} \psi \equiv \neg(\neg\varphi \mathcal{U} \neg\psi)$, $\diamondsuit \varphi \equiv \mathit{true} \mathcal{U} \varphi$, $\square\varphi \equiv \mathit{false} \mathcal{R} \varphi$, $\diamondsuit^- \varphi \equiv \mathit{true} \mathcal{U}^- \varphi$, $\square^- \varphi \equiv \neg \diamondsuit^- \neg\varphi$).

III. Temporal Causality

We now present a comprehensive primer on previous work regarding temporal causality that is relevant to this paper. Temporal causality is concerned with analyzing the cause for some effect emerging on an observed computation of a reactive system. In particular, this observed computation can be infinite, such as obtained as a counterexample from model checking. Informally speaking, the conditions for a causal relationship between two properties are as follows [10]:

- Both the cause property and the effect property are satisfied by the observed computation (**SAT** condition).
- The closest, i.e., most similar traces that do not satisfy the cause also do not satisfy the effect (**CF** condition).
- No subset of the cause property satisfies the previous two conditions (**MIN** condition).

More formally, the definition of such a causal relationship requires fixing a notion of closeness between system computations based on a similarity relation $\leq_\pi \subseteq \Sigma^\omega \times \Sigma^\omega$, which orders two traces $(\pi^1, \pi^2) \in \leq_\pi$ if π^1 is at least as similar to π than π^2. Such a relation can be expressed by a (relational) temporal formula such that, for instance, $\pi^1 \leq_{\pi^0}^{subset} \pi^2$ iff:

$$zip(\pi^0, \pi^1, \pi^2) \vDash \square \bigwedge_{i \in I} \big((i^0 \not\leftrightarrow i^1) \rightarrow (i^0 \not\leftrightarrow i^2) \big) \ .$$

Note that while we allow similarity relations over the full alphabet, we are usually interested in similarity of the input sequences [10], such as defined by $\leq^{subset}$. With these similarity relations at hand, we can now recall the formal definition of temporal causality.

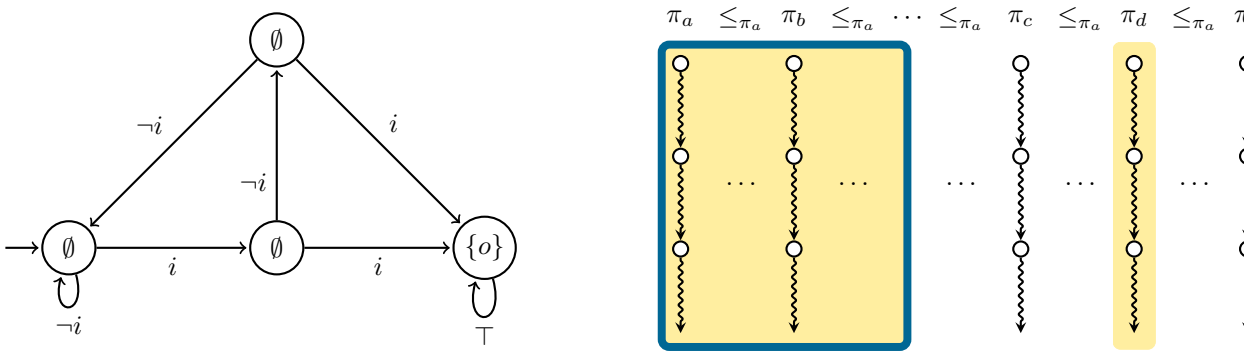
Definition 2 (Temporal Cause [15])**.** *Let $\mathcal{T}$ be a system, $\pi \in traces(\mathcal{T})$ a computation of the system, $\leq_\pi$ a similarity relation, and $E \subseteq (2^{AP})^\omega$ an effect property. We say that $C \subseteq (2^I)^\omega$ is a cause of E on π in $\mathcal{T}$ if the following holds.*

SAT: $\forall \pi' \in traces(\mathcal{T}) : \pi' =_I \pi \rightarrow \pi'|_I \in C \wedge \pi' \in E$.
CF: $\forall \pi' \in \overline{C} : \exists \pi'' \in traces(\mathcal{T}) : \pi'' \leq_\pi \pi' \wedge \pi'' \in \overline{E}$.
MIN: $\nexists C' \subset C : C'$ *satisfies* **SAT** *and* **CF**.

There are two details that stand out over our earlier, informal definition: First, the **SAT** condition requires that all traces that are input-equivalent to the observed trace satisfy cause and effect, which ensures that there is no causal property in the case of nondeterminism on the observed trace. Second, the **CF** condition is realized through a $\forall\exists$-quantifier alternation because there are cases where $(traces(\mathcal{T}), \leq_\pi)$ is not well-founded, such that no "closest" traces exist and the limit assumption for counterfactual reasoning is not met [15], [29].

Example 1. *Consider the reactive system $\mathcal{T}$ illustrated in Figure 2a, the trace $\pi = \{i\}\{i,o\}\{o\}^\omega$, and the effect $E = \square\diamondsuit o$, with similarity relation $\leq^{subset}$. The cause C_π for E on π is characterized by the formula $C_\pi = L(i \wedge \bigcirc i)$. It is easy to see that the **SAT** condition is met, let us take a closer look at the other two. For **CF**, we can find for any $\pi' \notin C_\pi$ either $\{\}\{i\}\{\}^\omega$ or $\{i\}\{\}\{\}^\omega$ as a $\pi'' \notin E$, in particular for traces such as $\pi''' = \{i\}\{\}\{i\}\{i,o\}^\omega$ that satisfy E but are less similar to π, e.g., $\{i\}\{\}\{\}^\omega \leq_\pi^{subset} \pi'''$. For **MIN**, we can see that restricting C_π further in any way that satisfies **SAT** leads to violation of the **CF** condition: For instance, if we had $C'_\pi = L(i \wedge \bigcirc i \wedge \neg\bigcirc\bigcirc i)$, then there would be the trace $\{i\}\{i,o\}\{i,o\}\{o\}^\omega \notin C'_\pi$ for which no at least as similar trace exists that does not satisfy the effect. It gets more complex when we have a trace such as $\sigma = \{i\}\{i,o\}^\omega$ where no finite number of inputs is responsible for obtaining the effect E. We can perform a similar analysis as above to show that $C_\sigma = L\big(\diamondsuit(i \wedge \bigcirc(i \vee \bigcirc i))\big)$ is the cause for E on σ.*

While Definition 2 with its three conditions closely mirrors Halpern and Pearl's definition of actual causality [21] by which it is inspired, Finkbeiner at al. [15] have shown recently that causes can be characterized much more succinctly. This is because they directly correspond to the largest downward closed set of traces in $(traces(\mathcal{T}), \leq_\pi)$ that satisfy the effect E. Vice versa, they are also the complement of the upward closure of $\overline{E}$. This stems from the balance of the **CF** and **MIN** conditions: If there was a trace in the cause that does not satisfy the effect, its upward closure could be removed – including other traces that do in fact satisfy the effect. As pictured in Figure 2b, this results in a cause (framed by the blue border) that, in essence, describes the local continuous neighborhood of traces that are similar to the observed trace



(a) An example reactive system $\mathcal{T}$.

(b) Cause C as the complement of the upward closure of $\overline{E}$ [15].

Fig. 2: Figure 2a illustrated the reactive system with the input i and output o that is used to outline temporal causality in Example 1. The system sets the output o continuously whenever the input i is enabled less than three time units apart. Figure 2b pictures a chain in $(traces(\mathcal{T}), \leq_{\pi_a})$ to illustrate that the cause C (enclosed by the blue frame) on π_a is the largest downward closed set of traces satisfying the effect E (shaded yellow), which is the complement of the upward closure of $\overline{E}$.

π_a and satisfy the effect E colored by the yellow area. More formally, this is captured by the following lemma.

Lemma 1 (Finkbeiner et al. [15])**.** *Let $\mathcal{T}$ be a system, $\pi \in traces(\mathcal{T})$ a computation of the system, $\leq_\pi$ a similarity relation, and $E \subseteq \Sigma^\omega$ an effect property. If there is a cause C of E on π in $\mathcal{T}$ then it is the largest downward closed set of system computations, i.e.,*

$$C = \{\, \rho \in (2^I)^\omega \mid \forall \sigma \in traces(\mathcal{T})\ \sigma \leq_\pi \rho \rightarrow \sigma \in E \,\} \ .$$

Moreover, the above set is empty iff there is no cause

Notably, Lemma 1 also means a cause is unique if it exists. In the following, we use this convenient characterization for a detailed study of the closure of property classes under causal inference and of complexity bounds on the size of automata representations for causes.

IV. Closure Under Causal Inference

In this section, we investigate the closure under causal inference of classes in the temporal hierarchy. As can be seen in Lemma 1, causes can be directly defined via universal quantification over the system traces to express that they are downward closed in $(traces(\mathcal{T}), \leq_\pi)$ and all satisfy the effect E. Speaking more abstractly, we are hence interested in whether properties X for which membership of some word is decided based on whether *all* associated words satisfy some other property Y inherit the hierarchy property classes from Y. We introduce a more convenient yet more general concept to facilitate this abstract analysis: the *universal preimage*. This concept essentially formalizes set membership based on universal quantification and makes our results applicable beyond closure under causal inference, i.e., to quantifier elimination in automata-based model checking of hyperproperties [7], [16] and synthesis from partial information [25].

Definition 3. *Let X and Y be sets, and let $f : X \to \mathcal{P}(Y)$ be a function. The **universal preimage** of S under f, denoted $f_U^{-1}(S)$, is the set of elements whose images under f are subsets of S: $f_U^{-1}(S) := \{x \in X \mid f(x) \subseteq S\}$.*

Informally, for an element $x \in X$ the function f determines the set of values of Y over which our universal quantifier ranges (e.g., all traces at least as similar as ρ in Lemma 1). If all of them are included in the set S, then x is included in the universal preimage $f_U^{-1}(S)$.

Causality can be defined via a universal preimage: The cause of E on π in $\mathcal{T}$ with similarity relation $\leq$ is the universal preimage of the effect under the map that sends a trace π'' to the set of at least as similar traces $\pi' \in \mathcal{T}$, such that $\pi' \leq_\pi \pi''$.

The choice of similarity relation heavily influences the closure of temporal hierarchy classes under causality. Different similarity relations produce different functions f. Instead of proving closure under causal inference for one specific similarity relation, we identify general assumptions on the relation, and hence the function f, that facilitate our closure results, such that they can be transferred to a variety of use cases and similarity relations.

The primary goal for the rest of this section is to formulate and prove the closure of temporal classes under the universal preimage. Section IV-A introduces the *Universal Closure Theorem* and provides a linguistic perspective on the main definitions. Section IV-B then provides a more abstract topological interpretation of the theorem and proves it using results from general topology. The last subsection (Section IV-C) presents the concept of the existential projection and discusses the parts of the temporal hierarchy which are not closed.

A. Linguistic Formulation

In general, temporal classes are not be closed under the universal preimage for an arbitrary $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$. To ensure closure, we impose certain restrictions on f.

Informally, to show that reachability is closed under f_U^{-1}, we need to show that if for some π'' every trace $\pi' \in f(\pi'')$ has a good prefix, then π'' must also have a good prefix.

The first property of f, which we need, is that we can build a prefix tree of $f(\pi'')$ while reading π''. It could be the case, for example, that the second level of the tree (the set of prefixes of $f(\pi'')$ of the length 2) cannot be determined unless the whole word π'' is observed. If such a case, the reachability class might not be closed under the universal preimage, since we do not know if every trace from $f(\pi'')$ has a good prefix unless we observe the whole π''.

Definition 4. *A function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is called **prefix-continuous** if for every $n \in \mathbb{N}$ and $\pi \in \Sigma_1^\omega$ there exists $m \in \mathbb{N}$ such that, for every $\pi' \in \Sigma_1^\omega$ if $\pi'_{(m)} = \pi_{(m)}$, then*

$$pref_n(f(\pi)) = pref_n(f(\pi')) \ .$$

Informally, this ensures that the prefix tree of $f(\pi'')$ can be constructed incrementally.

The second property that we need is that every infinite trace in the prefix tree corresponds to some trace in $f(\pi'')$.

Definition 5. *A function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is called **prefix-closed** if for every $\pi \in \Sigma_1^\omega$ and $\pi' \in \Sigma_2^\omega$:*

$$pref(\pi') \subseteq pref(f(\pi)) \Rightarrow \pi' \in f(\pi) \ .$$

We already have everything that we need for the case of a finite alphabet Σ_2, but if it is infinite then we set one additional restriction. The last property that we require from f is the finite branching of the prefix tree. Clearly, this is trivially satisfied for a finite alphabet.

Definition 6. *A function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is called **prefix-compact** if for every $n \in \mathbb{N}$ and $\pi \in \Sigma_1^\omega$ the set $pref_n(f(\pi))$ is finite.*

Remark 1. *If the alphabet Σ_2 is finite, then any function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is prefix-compact.*

Now everything is ready to formulate the main Theorem of this section, which formally states that several classes from the temporal hierarchy are closed under the universal preimage operation, for functions that satisfy the previously introduced requirements.

Theorem 1 (Universal Closure)**.** *Let Σ_1 and Σ_2 be alphabets. Assume a function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is **prefix-continuous, prefix-closed**, and **prefix-compact**, then:*

1) *If L is a safety property, then $f_U^{-1}(L)$ is a safety property.*
2) *If L is a guarantee property, then $f_U^{-1}(L)$ is a guarantee property.*
3) *If L is a recurrence property, then $f_U^{-1}(L)$ is a recurrence property.*

The proof of Theorem 1 requires establishing some auxiliary results using a topological argument, which we formulate in Section IV-B. With the theorem at hand, we can show that the general result in particular covers the previously introduced similarity relation $\leq^{subset}$, since this relation satisfies all the introduced requirements.

Proposition 1. *Let $\mathcal{T}$ be a finite state system, $\pi \in traces(\mathcal{T})$ a trace, $\leq^{subset}$ a subset similarity relation, and $\mathsf{E} \subseteq (2^{AP})^\omega$ an effect property. Suppose C is a cause of E on π. Then C is a universal preimage of E under a prefix-continuous, prefix-closed, and prefix-compact function.*

The immediate corollary from Theorem 1 and Proposition 1 is the closure of temporal hierarchy under causality.

Corollary 1 (Causality Closure)**.** *Let $\mathcal{T}$ be a finite state system, $\pi \in traces(\mathcal{T})$ a trace, $\leq^{subset}$ a subset similarity relation, and $\mathsf{E} \subseteq (2^{AP})^\omega$ an effect property. Suppose C is a cause of E on π in $\mathcal{T}$. Then the following statements hold.*

1) *If E is a safety property, then C is a safety property.*
2) *If E is a guarantee property, then C is a guarantee property.*
3) *If E is a recurrence property, then C is a recurrence property.*

Note that we will discuss the classes that are not closed in this way in Section IV-C, after proving Theorem 1 with results from general topology in the next section.

B. Proof via Topology

This section is dedicated to a topological characterization of the concepts presented before and proves Theorem 1 using results from general topology. We first introduce the basic topological definitions. In parallel, we define the Cantor metric and Cantor topology on the space of words.

1) Metric spaces: A metric space is a set equipped with a notion of distance between its elements. Formally, a *metric space* is a pair (M, d), where M is a set and $d : M \times M \to \mathbb{R}$ is a non-negative, symmetric function, which satisfies the triangle inequality and $d(x, y) = 0 \iff x = y$.

The set of words over an alphabet Σ can be equipped with a distance function d_C called the *Cantor distance*. The Cantor distance between two words π_1 and π_2 is defined as 0 if they are identical, and as $d_C(\pi_1, \pi_2) := 2^{-j}$ otherwise, where j is the length of the longest common prefix of π_1 and π_2. Intuitively, the closer two words are in the Cantor metric, the longer their common prefix. Note that the Cantor distance is bounded by 1.

For a metric space M, a set $O \subseteq M$ is *open* if, for every $x \in O$ there exists $\epsilon > 0$, such that the ball $B_\epsilon(x)$ of radius ϵ with the center in x lies entirely in O.

$$B_\epsilon(x) := \{x' \in M \mid d(x, x') \leq \epsilon\} \subseteq O \ .$$

The complement of an open set is called *closed*. The collection of open sets forms a *topology* on M, induced by the metric d.

For Σ^ω with the Cantor metric, open sets are of the form $X\Sigma^\omega$, where $X \subseteq \Sigma^*$ is a finite word language according to Proposition 3.1 in [37]. Intuitively, a set $L \subseteq \Sigma^\omega$ is open if, for every $\pi \in L$, there exists $n \in \mathbb{N}$, such that any π' coinciding with π in the first n letters also belongs to L. This is equivalent to saying that the ball of radius 2^{-n} centered at π is contained in L. The topology induced by the Cantor distance is called the *Cantor topology*.

According to Proposition 3.5 in [37], a set $L \subseteq \Sigma^\omega$ is closed if for every $\pi \in \Sigma^\omega$: $pref(\pi) \subseteq pref(L) \Rightarrow \pi \in L$.

The function $f : M_1 \to M_2$ between two metric spaces (M_1, d_1) and (M_2, d_2) is *continuous* if for every $x \in M_1$ and every $\epsilon > 0$, there exists $\delta > 0$ such that:

$$\forall x' \in M_1 : \text{ if } d_1(x, x') < \delta, \text{ then } d_2(f(x), f(x')) < \epsilon \ .$$

Cantor topology derives its name from the fact that the space of words with this topology can be continuously injected into the Cantor space, as shown by Plotkin [38].

2) Borel hierarchy: In a metric space M the union of any collection of open sets is also open and the intersection of a finite number of open sets remains open. However, the intersection of a countable collection of open sets may no longer be open. The set of countable intersections of open sets is noted as $\mathbf{\Pi_2}$.

Symmetrically, the set of closed sets is closed under arbitrary intersections but only finite unions. The set of countable unions of closed sets is denoted as $\mathbf{\Sigma_2}$. Together, $\mathbf{\Sigma_2}$ and $\mathbf{\Pi_2}$ form the second level of the Borel hierarchy.

The Borel hierarchy organizes subsets of the metric space M into classes. A set is called Borel if it belongs to some level of the Borel hierarchy. The first level of the Borel hierarchy consists of the set of open sets $\mathbf{\Sigma_1}$ and the set of closed sets $\mathbf{\Pi_1}$. The higher levels are defined recursively as follows:

$$\mathbf{\Sigma_n} := \{\cup_{i\in\mathbb{N}} X_i \mid X_i \in \mathbf{\Pi_{n-1}}\},$$

$$\mathbf{\Pi_n} := \{X \mid M \setminus X \in \mathbf{\Sigma_n}\},$$

$$\mathbf{\Delta_n} := \mathbf{\Sigma_n} \cap \mathbf{\Pi_n}.$$

The first two and a half levels of the Borel hierarchy for the set of infinite words Σ^ω with the Cantor topology correspond to the temporal hierarchy, as established by Mana and Pnueli [31].

Theorem 2 (Mana and Pnueli [31])**.** *Let $L \subseteq \Sigma^\omega$ be an infinite word language.*

1) *L is a safety property iff L is a closed set.*
2) *L is a guarantee property iff L is an open set.*
3) *L is a obligation property iff $L \in \mathbf{\Delta_2}$.*
4) *L is a recurrence property iff $L \in \mathbf{\Pi_2}$.*
5) *L is a persistence property iff $L \in \mathbf{\Sigma_2}$.*
6) *L is a reactivity property iff $L \in \mathbf{\Delta_3}$.*

Thus, the problem of the closure of temporal classes is equivalent to the problem of the closure of Borel classes.

A fundamental fact about Borel classes is that they are closed under the continuous preimage of a function [37].

Proposition 2 ([37])**.** *The preimage of a Borel set under a continuous function is a Borel set of the same Borel class.*

3) Hausdorff metric: A set $X \subseteq M$ is *compact* if for every collection of open sets $\{Y_i\}_{i\in\alpha}$ that covers X: $X \subseteq \bigcup_{i\in\alpha} Y_i$ there exists a finite subset of $\{Y_i\}_{i\in\alpha}$ that also covers X. Every compact subset of a metric space is closed.

According to Proposition 3.12 in [37], a set $L \subseteq \Sigma^\omega$ is compact in the Cantor topology if it is closed and for every $n \in \mathbb{N}$ the set $pref_n(L)$ is finite.

Since we are addressing the problem of closure of Borel classes under the preimages of a function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$, which maps traces to sets of traces, it is essential to define a metric on the set of sets of traces. If f is prefix-closed and prefix-compact it maps traces to compact sets of traces. Therefore, it suffices to define a metric on the set of compact sets. For a set $X \subseteq M$, we denote the set of all nonempty compact subsets of X as $\mathcal{K}(X)$. The Hausdorff distance between two nonempty sets $X, Y \subseteq M$ is defined as follows:

$$d_H(X, Y) = max(\sup_{x\in X} d(x, Y), \sup_{y\in Y} d(X, y)) \ ,$$

$$\text{where } d(a, B) = \inf_{b\in B} d(a, b) \text{ for } B \subseteq M \text{ and } a \in M.$$

On the set $\mathcal{K}(M)$, the Hausdorff distance d_H is a metric, which induces a topology known as the Hausdorff topology [18]. The next proposition shows that prefix-continuous functions are essentially functions that are continuous in Hausdorff topology.

Proposition 3. *A function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is prefix-continuous, prefix-closed and prefix-compact if and only if it is a continuous function from the metric space Σ_1^ω to the metric space $\mathcal{K}(\Sigma_2^\omega)$ with the Hausdorff metric d_H.*

4) Borel Classes in Hausdorff topology: Combining Propositions 2 and 3 we find that prefix-closed, prefix-compact, and prefix-continuous functions preserve Borel classes of the Hausdorff topology. Therefore, it remains to determine how Borel classes behave when lifted from Σ^ω to $\mathcal{K}(\Sigma^\omega)$.

Lemma 2. *Let M be a metric space with a subset $X \subseteq M$. Then the following statements are true.*

1) *If X is open in M, then $\mathcal{K}(X)$ is open in $\mathcal{K}(M)$.*
2) *If X is closed in M then $\mathcal{K}(X)$ is closed in $\mathcal{K}(M)$.*
3) *If $X \in \mathbf{\Pi_2}$ in M then $\mathcal{K}(X) \in \mathbf{\Pi_2}$ in $\mathcal{K}(M)$.*

Proof. The first two statements can be demonstrated through the equivalence of the Hausdorff topology and the Vietoris topology on the space of compact subsets [33]. However, for the sake of clarity, we present an explicit proof.

1): Let $X \subseteq M$ be an open set and let $K \in \mathcal{K}(X)$ be a compact subset of X. Define a continuous function f on K for every $x \in K$ as follows: $f(x) := d(x, M \setminus X)$. This is a continuous function from the compact set to $\mathbb{R}$. By the Extreme Value Theorem [41], f is bounded, and there exists $q \in K$ such that $f(q) = inf_{x\in K} f(x)$. Since $q \notin M \setminus X$ and $M \setminus X$ is closed, we conclude that $f(q) = d(q, M \setminus X) > 0$.

By the definition of Hausdorff distance, for every set $Z \in \mathcal{K}(M)$ with $d_H(K, Z) < f(q)$, it follows that for every $x \in Z$, $d(x, K) < f(q)$. Consequently, $x \notin M \setminus X$, thus, $Z \subseteq X$. Therefore the ball $B_{f(q)}(K)$, of radius $f(q)$ centered at K lies entirely within $\mathcal{K}(X)$. Since this holds for any $K \in \mathcal{K}(X)$, we have shown that $\mathcal{K}(X)$ is open.

2): Let $X \subseteq M$ be a closed set and let $K \in \mathcal{K}(M)$ a compact set not in $\mathcal{K}(X)$. Then there exists $x \in K$, such that $x \notin X$, and since X is closed, $d(x, X) > 0$.

For any $Z \in \mathcal{K}(M)$ with $d_H(Z,K) < d(x,X)/2$, there exists $z \in Z$, such that $d(x,z) < d(x,X)$. This implies $z \notin X$, thus, $Z \notin \mathcal{K}(X)$. Therefore, the ball $B_{d(x,X)/2}(K)$, of radius $d(x,X)/2$ centered at K, lies entirely within $\mathcal{K}(M) \setminus \mathcal{K}(X)$. Since this holds for any $K \in \mathcal{K}(M) \setminus \mathcal{K}(X)$, we have shown that $\mathcal{K}(M) \setminus \mathcal{K}(X)$ is open. Thus, $\mathcal{K}(X)$ is closed.

3): Let X be a set from the Borel class $\mathbf{\Pi_2}$ in M. Then by definition of $\mathbf{\Pi_2}$:

$$X = \bigcap_{n \in \mathbb{N}} Y_n \ ,$$

where each Y_n is an open set. Thus,

$$\mathcal{K}(X) = \bigcap_{n \in \mathbb{N}} \mathcal{K}(Y_n) \ .$$

Each $\mathcal{K}(Y_n)$ is open in $\mathcal{K}(M)$ by the first statement, completing the proof. □

Finally, we combine all the results to prove Theorem 1.

Proof of Theorem 1. By Proposition 3, f can be viewed as a continuous function between two metric spaces $f : \Sigma_1^\omega \to \mathcal{K}(\Sigma_2^\omega)$. By the definition of the universal preimage, we get

$$f_U^{-1}(L) = f^{-1}(\mathcal{K}(L)) \ .$$

By Lemma 2, we know that $\mathcal{K}$ preserves open, closed, and $\mathbf{\Pi_2}$ subsets. By Proposition 2 we know that f^{-1} preserves all Borel classes, as f is continuous. Thus, we prove all three statements of the theorem, since by Theorem 2 safety, guarantee, and recurrence properties are exactly open, closed, and $\mathbf{\Pi_2}$ sets. □

C. Existential preimage and non-closure of persistence

Since the universal preimage formalizes the universal quantification inherent in temporal causes, it is natural to seek a similar formalization for existential quantification. To this end, we introduce the concept of the existential preimage.

Definition 7. *Let X and Y be sets, and let $f : X \to \mathcal{P}(Y)$ be a function. The **existential preimage** of the set S under f, denoted $f_E^{-1}(S)$, is the set of elements whose images under f intersect S: $f_E^{-1}(S) := \{x \in X \mid f(x) \cap S \neq \emptyset\}$.*

The existential preimage is dual to the universal preimage in the sense that, for a set $S \subseteq Y$, $f_E^{-1}(S) = X \setminus (f_U^{-1} Y \setminus S))$. Using this duality we immediately get the following corollary from Theorem 1.

Corollary 2. *Assume a function $f : \Sigma_1^\omega \to \mathcal{P}(\Sigma_2^\omega)$ is prefix-continuous, prefix-closed, and prefix-compact. Then the safety, guarantee, and persistence classes are closed under f_E^{-1}.*

Hence, the persistence class is closed under the existential preimage, but as we show in the following, it is not closed under the dual universal preimage which encodes causal inference. The rest of the subsection discusses its behavior under the universal preimage, along with the obligation class which also is not closed under the universal preimage.

First, we consider causal inference with the subset similarity relation $\leq^{subset}$, a special case of the universal preimage by Proposition 1. We prove that an obligation effect can have a non-obligation cause.

Theorem 3. *There exists a system $\mathcal{T}$, a trace π, and an obligation ω-regular effect E, such that the cause of E on π in $\mathcal{T}$ with similarity relation $\leq^{subset}$ is not an obligation property.*

Proof. Define the input alphabet $I := \{a\}$. The output alphabet is empty, i.e., $O := \emptyset$. The system is the trivial set of all possible traces: $traces(\mathcal{T}) := (2^{I \cup O})^\omega$. The observed trace enables a continuously: $\pi := a^\omega$. Consider the effect $E := (\Box a) \vee (\Diamond(\neg a \wedge (\bigcirc a)))$. E is an obligation property since it is a disjunction of safety and guarantee properties. In essence, E does allow any trace but $a^+(\neg a)^\omega$. The cause of E on π in $\mathcal{T}$ is $\Box\Diamond a$, since if π'' satisfies $\Box\Diamond a$, then clearly any $\pi' \leq_\pi^{subset} \pi''$ satisfies $\Box\Diamond a$, and hence E. If π'' does not satisfy $\Box\Diamond a$ that $\pi'' \in (2^I)^n(\neg a)^\omega$ for some n, hence $a^n(\neg a)^\omega \leq_\pi^{subset} \pi''$, thus π'' is not in the cause. □

Together with Theorem 1 that states that recurrence is closed, this immediately implies that persistence also is not closed under causality. This is because from Theorem 1 it follows that cause for the obligation property is a recurrence property, so the only way it cannot be an obligation property is by not being a persistence property.

Corollary 3. *There exists a system $\mathcal{T}$, a trace π, and a persistence ω-regular effect E, such that the cause of E on π in $\mathcal{T}$ with similarity relation $\leq^{subset}$ is not a persistence property.*

Besides the fact that persistence is not closed under causality and consequently under the universal preimage, we want to investigate for how many Borel classes we can find similar counterexamples. Clearly, we must go beyond the ω-regular setting, as ω-regular properties are contained in $\mathbf{\Delta_3}$ and are themselves closed under causal inference.

Existential and universal preimages are dual to each other, as recurrence ($\mathbf{\Pi_2}$) and persistence ($\mathbf{\Sigma_2}$) classes. Thus, it suffices to examine the behavior of recurrence properties under existential preimage.

Lemma 3. *Let $A \subseteq \Sigma^*$ be a finite word language. Then the infinite word language A^ω is an existential preimage of a recurrence property under a prefix-continuous, prefix-closed, and prefix-compact function.*

Proof. Define the alphabet Σ' as consisting of primed copies of symbols from Σ.

The language $A' \subseteq (\Sigma \cup \Sigma')^\omega$ consists of words from A in which the last letter is replaced with its primed version. Formally:

$$A' := \{\pi_{(|\pi|-1)}\pi'_{|\pi|} \mid \pi \in A\} \ .$$

Clearly, $(A')^\omega$ is a recurrence property, as a word π is in $(A')^\omega$ if and only if it has infinitely many prefixes from $(A')^*$.

Define a prefix-continuous, prefix-closed, and prefix-compact function $f : \Sigma^\omega \to (\Sigma \cup \Sigma')^\omega$ for $\pi \in \Sigma^\omega$ as follows:

$$f(\pi) := \{\pi' \mid \forall i : \ \pi(i) = \pi'(i) \text{ or } \pi(i)' = \pi'(i)\} \ .$$

Essentially, f randomly replaces each symbol in π with its primed version. Intuitively it tries to split π into words from A, with primed letters representing the endings of words from A. If f can split π into the words from A, this splitting is in $f(\pi) \cap (A')^\omega$. Hence, $A^\omega = f_E^{-1}((A')^\omega)$. □

Theorem 4. *There exists a recurrence property L and a prefix-continuous, prefix-closed, and prefix-compact function f, such that $f_E^{-1}(L)$ is not a Borel set.*

Proof. There exists a finite language A, such that A^ω is not Borel, as shown in [17]. Therefore, the theorem follows directly from Lemma 3. □

V. Complexity

In this section, we take a closer look at the complexity of synthesizing causes as automata and study bounds for the size of these automata. We show lower bounds for all property classes, which are the first lower bounds for the problem and witness that the exponential scaling in the algorithm of Finkbeiner et al. [15] cannot be avoided. However, we also show that the upper bounds can still be improved for several property classes as logarithmic factors can be avoided. We focus our attention to the case of subset similarity relation $\leq^{subset}$. The presented characterization is precise with respect to the system size, but with respect to the effect size, there remains a minor gap between the lower and upper bounds on the higher levels of the hierarchy.

We heavily use the characterization of causes as downward closed sets of traces satisfying the effect (cf. Lemma 1). For the upper bounds, we present the algorithms that output the set from Lemma 1 matching a cause when it exists.

A. Lower bounds

First, we prove that the NBW and NCW complementation problems can be reduced to the cause synthesis problem linearly in the size of the system with persistence or recurrence effects, respectively.

The complement of an NBW can be expressed as a UCW (universal co-Büchi automaton) with the same structure. Hence it can be viewed as a universal quantification over a DCW automaton. Similarly, the complement of an NCW can be viewed as a universal quantification over a DBW automaton.

The trick in the proof is now to interpret the automaton to be complemented as the system in a causal inference tasks. The acceptance condition of the complement automaton can be encoded as a recurrence/persistence property for Büchi and Co-Büchi acceptance, respectively. It remains to ensure that only identical words are related in the similarity relation, such that the universal quantification as described in Lemma 1 only ranges over the same word. In the end, a word is then in the cause automaton only if all runs satisfy the complementary acceptance condition. Hence, the language of the cause automaton is exactly the complement of the original automaton language. Technical details of this construction are provided in the proof of the following lemma.

Lemma 4. *For every NBW (NCW) A there exists a system $\mathcal{T}$ with $|A|+1$ states, an effect E represented by a DCW (DBW) of constant size and a trace π of constant size, such that if the cause of E on π in $\mathcal{T}$ is expressed as a NBW C then there exists a NBW for the complement of $L(A)$ of the size $\mathcal{O}(|C|)$.*

Proof. Denote $A = (Q, \Sigma, q_0, F, \Delta)$, with $\Sigma = \{1, \dots, m\}$ and $|Q| = n$. Let us define a system $\mathcal{T} = (S, s_0, AP, \delta, l)$, where $AP = I \bigcup O$.

$$I := \{i_1, \dots, i_k, j_1, \dots, j_{\lceil \log(n) \rceil}\}.$$

Here k is a minimal integer such that $\binom{k}{k/2} \geq m$. Outputs and states are defined as follows.

$$O := \{o\}, \;\; S := Q \cup \{s_\top\}, s_0 := q_0.$$

To define δ at first we need to encode pairs $q, \sigma \in Q \times \Sigma$ as elements from 2^I. We fix an injective function enc_Q from Q to the set of subsets of $\{j_1, \dots, j_{\lceil \log(n) \rceil}\}$.

$$enc_Q : Q \to \mathcal{P}(j_1, \dots, j_{\lceil \log(n) \rceil}).$$

Additionally, we fix an injective function enc_Σ from Σ to the set of subsets of $\{i_1, \dots, i_k\}$ of the size $k/2$. The number of such subsets is $\binom{k}{k/2} \geq m$ by choice of k, hence such injective function exists.

$$enc_\Sigma : \Sigma \to \{V \subseteq \{i_1, \dots, i_k\} \mid \; |V| = k/2\}$$

Please note that for every two different $\sigma, \sigma' \in \Sigma$ subsets $enc_\Sigma(\sigma)$ and $enc_\sigma(\sigma')$ are incomparable.

$$\forall \sigma, \sigma' \in \Sigma : \sigma \neq \sigma' \Rightarrow enc_\Sigma(\sigma) \not\subseteq enc_\Sigma(\sigma'). \qquad (*)$$

We define the transition function δ as follows.

$$\delta(q, enc_Q(q') \cup enc_\Sigma(\sigma)) := \begin{cases} q', & \text{if } q' \in \Delta(q, \sigma), \\ s_\top, & \text{otherwise.} \end{cases}$$

For every $q \in Q$ and $W \subseteq I$ which cannot be presented as $enc_Q(q') \cup enc_\Sigma(\sigma)$ for $q' \in Q$ and $\sigma \in \Sigma$:

$$\delta(q, W) := s_\top \quad \text{and} \quad \delta(s_\top, W) := s_\top \;\; .$$

If A is an NBW labeling l_{NBW} is defined as follows:

$$l_{NBW}(s) := \begin{cases} \{o\} & \text{if } s \in F \;\; , \\ \emptyset & \text{if } s \in Q \setminus F \cup \{s_\top\} \;\; . \end{cases}$$

If A is an NCW the labeling l_{NCW} is defined differently:

$$l_{NCW}(s) := \begin{cases} \{o\} & \text{if } s \in Q \setminus F \;\; , \\ \emptyset & \text{if } s \in F \cup \{s_\top\} \;\; . \end{cases}$$

Trace $\pi = \emptyset^\omega$ does not depend on the automaton. We define the effect depending on whether A is an NBW or an NCW:

$$E_{NBW} := \Diamond \Box \neg o \quad \text{or} \quad E_{NCW} := \Box \Diamond \neg o \;\; .$$

For a word $\sigma \in \Sigma^\omega$ we denote the encoding of σ with 2^I as $Enc_I(\sigma) \in (2^I)^\omega$. For every k the k-th letter of $Enc_I(\sigma)$ is defined as follows:

$$Enc_I(\sigma)_k = enc_\Sigma(\sigma_k) \cup \{j_1, \dots, j_{\lceil log(n) \rceil}\} \;\; .$$

Claim: For every word $\sigma \in \Sigma^\omega$: $\sigma \in \overline{A} \iff Enc_I(\sigma) \in C$.

The first direction: $\sigma \in \overline{A} \implies Enc_I(\sigma) \in C$. Assume a word $\sigma \in \overline{A}$. All runs of A on σ must be rejecting, hence they must visit F finitely many times if A is an NBW or infinitely many times if A is an NCW.

Assume a trace $\pi' \in (2^{I \cup O})^\omega$, such that $\pi' \leq_\pi^{subset} Enc_I(\sigma)$. If while producing π' the system $\mathcal{T}$ visits state $s_\top$, then π' satisfies E. Suppose while producing π' system $\mathcal{T}$ does not visit $s_\top$. Thus, by the definition of the transition system for every k: $\pi'_k(I) = enc_Q(q_k) \cup enc_\Sigma(\sigma_k)$, where $\{q_j\}_{j \in \mathbb{N}}$ is a run of A on σ. Please note that the Σ part of π' in this case is the encoding of σ and not an encoding of any other trace from Σ^ω, by the fact that $\pi' \leq_\pi^{subset} Enc_I(\sigma)$ and the property of encoding (*).

As we noted before, $\{q_j\}_{j \in \mathbb{N}}$ must be rejecting, hence it cannot visit F infinitely many times in the case of NBW or it must visit F infinitely many times in the case of NCW. Thus π' satisfies E. We proved that $Enc_I(\sigma) \in C$, since for any $\pi' \leq_\pi^{subset} Enc_I(\sigma)$: $\pi' \in E$.

The second direction: $\sigma \in \overline{A} \impliedby Enc_I(\sigma) \in C$. Assume a word $\sigma \in \Sigma^\omega$ such that $Enc_I(\sigma) \in C$. Let us take a run $\{q_k\}_{k \in \mathbb{N}}$ of A on σ. Let us define the word $\pi' \in (2^I)^\omega$ as follows. For every k:

$$\pi'_k := enc_Q(q_k) \cup enc_\Sigma(\sigma_k) \ .$$

By the definition of $\leq^{subset}$ we get $\pi' \leq_\pi^{subset} Enc_I(\sigma)$. Hence, $\mathcal{T}(\pi')$ satisfies E by the definition of the cause, since $Enc_I(\sigma) \in C$. Hence, by the definition of E $\{q_k\}_{k \in \mathbb{N}}$ does not visit F infinitely often in the case of NBW or visits F infinitely often in the case of NCW, thus it is a rejecting run of A. Therefore, $\sigma \in \overline{A}$ since every run of A on σ is rejecting. □

Similarly, we provide the linear encoding of the NFW complementation problem into the cause synthesis problem with safety or reachability effect.

Lemma 5. *For every NFW A there exists a system $\mathcal{T}$ with $|A| + 1$ states, an effect E in the form of a safety (or reachability) DBW of constant size and a trace π of constant size, such that if the cause of E on π in $\mathcal{T}$ expressed as a NBW C then there exists a NFW for the complement of $L(A)$ of the size $\mathcal{O}(|C|)$.*

Proof. Similar to Lemma 4. We add to the alphabet of A one additional symbol $\#$, which represents the end of the word. Then we construct $\mathcal{T}$,E, and π in the same way as we did in the Lemma 4 to track all possible executions of A. The only difference is that now E detects if the last state that appeared before the first occurrence of $\#$ was accepting in A or not.

Please note that E in this case can be a safety or reachability DBW since we can either accept or reject words without occurrences of $\#$. If C is the NBW for the cause, we can easily turn it into NFW for $\overline{A}$ just by calling a state accepting if C accepts $\#^\omega$ from this state. □

Using the provided reductions of complementation problems we derive the lower bounds of the cause size with respect to the system $\mathcal{T}$ for different temporal classes from the well-known lower bounds on the automata complementation problems.

Theorem 5. *The following states lower bounds for the cause automaton with respect to the size of the system.*

1) *An NBW for the cause of a persistence effect in a system $\mathcal{T}$ requires at least $2^{\Omega(|\mathcal{T}| log |\mathcal{T}|)}$ states in the worst case.*
2) *An NBW for the cause of a recurrence effect in a system $\mathcal{T}$ requires at least $\Omega(3^{|\mathcal{T}|})$ states in the worst case.*
3) *An NBW for the cause of a safety or guarantee effect in a system $\mathcal{T}$ requires at least $\Omega(2^{|\mathcal{T}|})$ states in the worst case.*

Proof. 1) For an NBW with n states, an NBW for the complement requires at least $2^{\Omega(nlogn)}$ states in the worst case [45]. By Lemma 4 NBW complementation can be reduced to cause synthesis with persistence effect linearly in the size of $\mathcal{T}$.

2) For an NCW with n states, an NBW for the complement requires at least $\Omega(3^n)$ states in the worst case [4]. By Lemma 4, NCW complementation can be reduced to cause synthesis with a recurrence effect linearly in the size of the system $\mathcal{T}$.

3) For an NFW with n states, an NFW for the complement requires at least $\Omega(2^n)$ states in the worst case [23]. By Lemma 5, NFW complementation can be reduced to cause synthesis with a safety or guarantee effect linearly in the size of the system. □

We turn our focus to the complexity with respect to the effect E. We prove a doubly exponential lower bound. For that purpose for every $n \in \mathbb{N}$, we define a language that is the cause of an effect of size $\mathcal{O}(n)$, while any NBW recognizing this language requires at least $2^{2^{\Omega(n)}}$ states.

For $n \in \mathbb{N}$ and $\pi \in \Sigma^*$ let us denote as $subword_n(\pi)$ the set of words of length n that appear in π from a position divisible by n.

$$subword_n(\pi) := \{\pi_{kn}, \pi_{kn+1}, \ldots, \pi_{(k+1)n-1} \mid k \in \mathbb{N}\}$$

Let us define a finite word language $L_n \subseteq \{0, 1, \#\}^*$ for $n \in \mathbb{N}$ as follows.

$$L_n := \{\pi \in \{0, 1, \#\}^* \mid \forall w \in \{0, 1\}^n,\ w \in subword_n(\pi)\}$$

In other words, L_n consists of words π such that $\{0, 1\}^n \subseteq subword_n(\pi)$. We prove the doubly exponential lower bound on the NBW which recognizes L_n.

Lemma 6. *An NFW for L_n requires at least $2^{2^{\Omega(n)}}$ states.*

Now we show that L_n is the cause of an effect of size $\mathcal{O}(n)$. The idea uses that a word π belongs to L_n if $subword_n(\pi)$ contains every word $w \in \{0, 1\}^n$.

For any single w consider the infinite word w^ω. For such a word there exists a position kn such that π and w^ω coincide for n consecutive symbols starting at kn. This position marks where w appears in π.

To verify this, an automaton with $\mathcal{O}(n)$ states suffices. It tracks the position modulo n and nondeterministically selects kn as the starting point where π and w^ω coincide. Hence, universally quantifying this check over all w we get a construction that recognizes L_n.

Lemma 7. *For every n there is a system $\mathcal{T}$ and a trace $\pi_\emptyset$ of constant sizes and a safety (or reachability) effect of size $\mathcal{O}(n)$, such that if the cause of E on $\pi_\emptyset$ in $\mathcal{T}$ expressed as a NBW C then there exists a NFW for L_n with $\mathcal{O}(|C|)$ states.*

Proof. Let us define an input alphabet consisting of four letters $I = \{i_0, i_1, i_\#, i_*\}$ and an output alphabet consisting of one letter $O = \{o\}$. The system is trivial and models every trace over the alphabet. The trace is the trivial trace without enabled atomic propositions: $traces(\mathcal{T}) := (2^{I\cup O})^\omega,\ \pi_\emptyset := \emptyset^\omega$. The effect E is defined as the union $E := E_1 \cup E_2 \cup E_3$. The first part E_1 requires that at some position all input variables $i_0, i_1, i_\#, i_*$ become false.

$$E_1 := \{\pi \in (2^{I\cup O})^\omega \mid \exists k : \neg(\pi_k(i_0) \vee \pi_k(i_1) \vee \pi_k(i_\#) \vee \pi_k(i_*)))\}$$

E_2 requires that the output part of the trace does not repeat every n states, in other words, there is a position such that after n positions the output variable takes a different value.

$$E_2 := \{\pi \in (2^{I\cup O})^\omega | \exists k : \pi_k(o) \neq \pi_{k+n}(o)\}$$

E_3 requires that from some position divisible by n before i_* becomes true for the first time variables i_1 and o take the same value n steps in the row.

$$E_3 := \{\pi \in (2^{I\cup O})^\omega | \exists k\ \forall l < kn\ \forall j < n : \neg\pi_l(i_*) \wedge (\pi_{kn+j}(i_1) = \pi_{kn+j}(o))\}$$

Since E must remember only the number of the steps modulo n it can be modeled by an NBW with n states which nondeterministically decides on which step E_2 or E_3 is satisfied.

For a word $\sigma \in \{0, 1, \#\}^*$ we denote the encoding of σ with 2^I as $Enc_I(\sigma) \in (2^I)^\omega$. For every k the k-th letter of $Enc_I(\sigma)$ is defined as follows:

$$Enc_I(\sigma)_k = \begin{cases} \{i_c\} & \text{if } k < |\sigma| \text{ and } \sigma_k = c \\ \{i_*\} & \text{if } k \geq |\sigma| \end{cases},$$

Claim: $\forall \sigma \in \{0, 1, \#\}^*$: $\sigma \in L_n \iff Enc_I(\sigma) \in C$.

Assuming the Claim, given a NBW for the cause C an NFW of size $\mathcal{O}(|C|)$ for L_n can be easily constructed by d combining C with encoding Enc_I and defining accepting states as the states from which C accepts $\{i_*\}^\omega$. The rest of the proof is devoted to proving the claim.

The first direction: $\sigma \in L_n \implies Enc_I(\sigma) \in C$. Let σ be a word from L_n. Let us prove that $Enc_I(\sigma) \in C$. For that, we need to prove that for every $\pi' \leq^{subset}_{\pi_\emptyset} Enc_I(\sigma)$: $\pi' \in E$. If $\pi'_I \neq Enc_I(\sigma)$, then by the definition of $\leq^{subset}$ and the encoding for some k:

$\pi'_k(i_0) = \pi'_k(i_1) = \pi'_k(i_\#) = \pi'_k(i_*) = \bot$, hence $\pi' \in E_1$.

If $\pi'(o)$ does not repeat every n states $\pi' \in E_2$: Assume that $\pi' \notin E_1 \cup E_2$. Then the prefix $\pi'(o)_{(n)}$ repeats in $\pi'(o)$ every n states. Since $\sigma \in L_n$ the word $\pi'(o)_{(n)}$ considered as a word over $\{0, 1\}$ must appear in σ from the position kn for some k. Hence, $\pi'(o)$ and $Enc_I(\sigma)(i_1)$ take the same value n steps in the row from position kn. Since $\pi' \notin E_1$, we can conclude that $Enc_I(\sigma(i_1)) = \pi'(i_1)$. Thus, $\pi' \in E_3$.

The second direction: $\sigma \in L_n \impliedby Enc_I(\sigma) \in C$. Assume that $Enc_I(\sigma) \in C$. Let us prove that $\sigma \in L_n$. For that, we need to prove that for every $w \in \{0, 1\}^n$: $w \in subword_n(\pi)$.

Consider $\pi' \in (2^{I\cup O})^\omega$ such that $\pi'_I = Enc_I(\sigma)$ and for every k: $\pi'_k(o) = w_{k\%n}$, where $k\%n$ is k modulo n. Obviously $\pi' \leq^{subset}_{\pi_\emptyset} Enc_I(\sigma)$, hence $\pi' \in E$. Moreover, $\pi' \in E_3$, since $\pi' \notin E_1 \cup E_2$.

Hence, there exists k, such that $\pi'(o)$ and $\pi'(i_1)$ take the same value n steps in a row from the step kn. Thus, w appears in π from the position kn. □

The lower bound in the size of the effect then immediately follows from the last two lemmas.

Theorem 6. *An NBW for the cause of a safety or reachability effect, given as an NBW E requires at least $2^{2^{\Omega(|E|)}}$ states in the worst case.*

B. Upper bounds

This subsection establishes upper bounds that match the lower bounds presented in the previous subsection. With respect to the system size $|\mathcal{T}|$ all the bounds are tight.

The first upper bounds presented in Theorem 7 is $\mathcal{O}(|\pi| \cdot 3^{(|\mathcal{T}|\cdot|E|)})$ for the recurrence effect. Note that this assumes that E is provided as a DBW. Converting a recurrence NBW to a DBW requires a blow-up of $2^{\Omega(n \log n)}$ [11]. Hence, the combined upper bound for the cause synthesis for the recurrence effect presented as NBW becomes $2^{2^{\mathcal{O}(|E| \log |E|)}}$.

In contrast, the lower bound established in Theorem 6 is $2^{2^{\Omega(|E|)}}$, leaving a question about the tight bound unresolved. The same applies to the persistence class, as the upper bound on the effect derived from [15] is also $2^{2^{\mathcal{O}(|E| \log |E|)}}$.

Theorem 7. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and an effect E given as a DBW, there exists a DBW for the cause of the size $\mathcal{O}(|\pi| \cdot 3^{(|\mathcal{T}|\cdot|E|)})$.*

Sketch of Proof. We construct a universal Büchi automaton of the size $|\mathcal{T}| \cdot |E|$ for the cause. Using the fact that a universal Büchi automaton of the size n can be translated to a nondeterministic one of the size 3^n [34], we derive the stated upper bound. For the construction details of the UBW and the handling of π please see the full version of the proof. □

Next, we present upper bounds for safety (Theorem 8) and guarantee (Theorem 9) properties. Both results assume the effect is given as a DFW for the good or bad prefixes respectively. Since translating safety (or guarantee) NBW to the DFW for good (or bad) prefixes requires an exponential blowup [24], these upper bounds match the lower bound

established in Theorem 6, with respect to the size of the effect given as an NBW.

Theorem 8. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and a safety effect E given as a DFW E_{bad_pref} for the bad prefixes of E, there exists a DFW for the bad prefixes of the cause of the size $\mathcal{O}(|\pi| \cdot 2^{(|\mathcal{T}|\cdot|E_{bad\_pref}|)})$.*

Sketch of Proof. We construct a NFW which recognizes pairs of finite traces between which a bad prefix exists. The non-deterministic choice represents the selection of such traces. Subsequently, we convert it to a DFW and combine it with a trace π. For details, please refer to the full proof. □

Theorem 9. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and a guarantee effect E given as a DFW E_{good_pref} for the good prefixes of E, there exists a DFW for the good prefixes of the cause of the size $\mathcal{O}(|\pi| \cdot 2^{(|\mathcal{T}|\cdot|E_{good\_pref}|)})$.*

Sketch of Proof. By the definition a finite word w is a good prefix of the cause C if every continuation of it is in C. And for that we need that all finite words between w and π have a good prefix of E. We construct a UFW of size $|\mathcal{T}|\cdot|E_{good\_pref}|$ that recognizes the good prefixes of the cause and the convert it into a DFW. For the details on this construction please refer to the full version of the proof. □

VI. Related Work

Methodology: The topological concepts used in Section IV have long been established in mathematics. The Hausdorff distance was introduced by Felix Hausdorff in [22]. The Vietoris topology [46] is another topology on the space of subsets. It coincides with the Hausdorff topology on compact subsets but differs when generalized to arbitrary closed subsets [33]. These concepts are employed in the Powerdomain theory, studied by Plotkin [38], [39] and Smyth [42]. The relation between Powerdomains and the Vietoris topology is discussed in [43]. The Vietoris topology can be split into the upper and lower Vietoris topologies. The lower Vietoris topology was applied by Clarkson and Schneider in [8] to characterize different classes of hyperproperties. The established connection between these topological concepts, universal preimages, and causality enable us to use the existing mathematical theory to study the framework.

Causality: The complexity of checking and computing actual causes in finite, so-called *structural equation models* [21] has been studied extensively [13], [14], [19]. In that framework, causes are essentially finite sets of explicit events (comparable to $\mathbf{\Delta_0}$ in Figure 1) and not arbitrary properties as considered in this paper. Moreover, the structural equation approach does not model time explicitly and can only be used to model system executions up to a fixed bound [9]. An extension to a state-based attribution method for transitions systems has been studied recently [32]. For approaches that combine counterfactual reasoning with temporal properties, there are a number of related complexity results not pertaining to temporal causality as considered in this paper. *Event Order Logic* [28] is an approach that expresses what order of events is causal for some property violation. Upper bounds for the time needed to compute causes in this logic are known to be exponential in the system size [27]. Parreuax et al. [36] define causes for reachability and safety effects in transition systems and two-player games and show that causes can be checked in polynomial time, but they do not consider the problem of cause synthesis or more complex effects. To the best of our knowledge, our paper is the first to establish explicit results on the closure under causal inference. We believe this is because temporal causality is the first formalism in this domain that uses the same language for both cause and effect, in the spirit of earlier work on counterfactual modal logic [29], [44].

VII. Summary & Conclusion

We have conducted a detailed investigation of closure under causal inference and complexity of cause synthesis for properties belonging to all classes of the hierarchy of temporal properties [31]. Our discoveries can be summarized as follows:

1) Reachability, guarantee, and recurrence properties are closed under causal inference: An effect from these classes always has a cause from the same class. This complements previous results on ω-regular properties [15] and the intersection of safety and guarantee properties [3].
2) Obligation and recurrence properties are not closed in the same way, which completes the picture regarding the hierarchy of ω-regular properties.
3) Based on 1), we provide improved upper bounds for the size of causes synthesized from reachability and guarantee properties.
4) We show lower bounds on the size of causes for all classes from the hierarchy, which confirm that the known algorithms are optimal in the number of exponents and with respect to the system size. For some classes, a gap in the logarithmic factors with respect to the effect remains.

Contribution 3) is of high practical relevance for explaining model-checking results of safety and guarantee properties, which are common in verification tasks. Contributions 1 and 2 were proven via results for a more abstract mathematical operation that promises to generalize beyond cause synthesis to other problems that involve trace-quantifier alternations. This includes synthesis with incomplete information [25], [26] and automata-based algorithms for hyperproperties [2], [16]. We plan on investigating these connections in future work.

Acknowledgements

This work was partially supported by the DFG in project 389792660 (TRR 248 – CPEC) and by the ERC Grant HYPER (No. 101055412). Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] I. Beer, S. Ben-David, H. Chockler, A. Orni, and R. J. Trefler, "Explaining counterexamples using causality," in *Computer Aided Verification, 21st International Conference, CAV 2009, Grenoble, France, June 26 - July 2, 2009. Proceedings*, ser. Lecture Notes in Computer Science, A. Bouajjani and O. Maler, Eds., vol. 5643. Springer, 2009, pp. 94–108. [Online]. Available: https://doi.org/10.1007/978-3-642-02658-4_11

[2] R. Beutner, D. Carral, B. Finkbeiner, J. Hofmann, and M. Krötzsch, "Deciding hyperproperties combined with functional specifications," in *LICS '22: 37th Annual ACM/IEEE Symposium on Logic in Computer Science, Haifa, Israel, August 2 - 5, 2022*, C. Baier and D. Fisman, Eds. ACM, 2022, pp. 56:1–56:13. [Online]. Available: https://doi.org/10.1145/3531130.3533369

[3] R. Beutner, B. Finkbeiner, H. Frenkel, and J. Siber, "Checking and sketching causes on temporal sequences," in *Automated Technology for Verification and Analysis - 21st International Symposium, ATVA 2023, Singapore, October 24-27, 2023, Proceedings, Part II*, ser. Lecture Notes in Computer Science, É. André and J. Sun, Eds., vol. 14216. Springer, 2023, pp. 314–327. [Online]. Available: https://doi.org/10.1007/978-3-031-45332-8_18

[4] U. Boker, O. Kupferman, and A. Rosenberg, "Alternation removal in büchi automata," in *International Colloquium on Automata, Languages and Programming*, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:8054966

[5] H. Chockler and J. Y. Halpern, "Explaining image classifiers," in *Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning, KR 2024, Hanoi, Vietnam. November 2-8, 2024*, P. Marquis, M. Ortiz, and M. Pagnucco, Eds., 2024. [Online]. Available: https://doi.org/10.24963/kr.2024/25

[6] H. Chockler, J. Y. Halpern, and O. Kupferman, "What causes a system to satisfy a specification?" *ACM Trans. Comput. Log.*, vol. 9, no. 3, pp. 20:1–20:26, 2008. [Online]. Available: https://doi.org/10.1145/1352582.1352588

[7] M. R. Clarkson, B. Finkbeiner, M. Koleini, K. K. Micinski, M. N. Rabe, and C. Sánchez, "Temporal logics for hyperproperties," in *Principles of Security and Trust - Third International Conference, POST 2014, Grenoble, France, April 5-13, 2014, Proceedings*, ser. Lecture Notes in Computer Science, M. Abadi and S. Kremer, Eds., vol. 8414. Springer, 2014, pp. 265–284. [Online]. Available: https://doi.org/10.1007/978-3-642-54792-8_15

[8] M. R. Clarkson and F. B. Schneider, "Hyperproperties," *J. Comput. Secur.*, vol. 18, no. 6, pp. 1157–1210, 2010. [Online]. Available: https://doi.org/10.3233/JCS-2009-0393

[9] N. Coenen, R. Dachselt, B. Finkbeiner, H. Frenkel, C. Hahn, T. Horak, N. Metzger, and J. Siber, "Explaining hyperproperty violations," in *Computer Aided Verification - 34th International Conference, CAV 2022, Haifa, Israel, August 7-10, 2022, Proceedings, Part I*, ser. LNCS, S. Shoham and Y. Vizel, Eds., vol. 13371. Springer, 2022, pp. 407–429. [Online]. Available: https://doi.org/10.1007/978-3-031-13185-1_20

[10] N. Coenen, B. Finkbeiner, H. Frenkel, C. Hahn, N. Metzger, and J. Siber, "Temporal causality in reactive systems," in *Automated Technology for Verification and Analysis - 20th International Symposium, ATVA 2022, Virtual Event, October 25-28, 2022, Proceedings*, ser. LNCS, A. Bouajjani, L. Holík, and Z. Wu, Eds., vol. 13505. Springer, 2022, pp. 208–224. [Online]. Available: https://doi.org/10.1007/978-3-031-19992-9_13

[11] T. Colcombet and K. Zdanowski, "A tight lower bound for determinization of transition labeled büchi automata," in *International Colloquium on Automata, Languages and Programming*, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:1073464

[12] A. Datta, D. Garg, D. K. Kaynar, D. Sharma, and A. Sinha, "Program actions as actual causes: A building block for accountability," in *IEEE 28th Computer Security Foundations Symposium, CSF 2015, Verona, Italy, 13-17 July, 2015*, C. Fournet, M. W. Hicks, and L. Viganò, Eds. IEEE Computer Society, 2015, pp. 261–275. [Online]. Available: https://doi.org/10.1109/CSF.2015.25

[13] T. Eiter and T. Lukasiewicz, "Complexity results for structure-based causality," *Artif. Intell.*, vol. 142, no. 1, pp. 53–89, 2002. [Online]. Available: https://doi.org/10.1016/S0004-3702(02)00271-0

[14] ——, "Causes and explanations in the structural-model approach: Tractable cases," *Artif. Intell.*, vol. 170, no. 6-7, pp. 542–580, 2006. [Online]. Available: https://doi.org/10.1016/j.artint.2005.12.003

[15] B. Finkbeiner, H. Frenkel, N. Metzger, and J. Siber, "Synthesis of temporal causality," in *Computer Aided Verification - 36th International Conference, CAV 2024, Montreal, QC, Canada, July 24-27, 2024, Proceedings, Part III*, ser. Lecture Notes in Computer Science, A. Gurfinkel and V. Ganesh, Eds., vol. 14683. Springer, 2024, pp. 87–111. [Online]. Available: https://doi.org/10.1007/978-3-031-65633-0_5

[16] B. Finkbeiner, M. N. Rabe, and C. Sánchez, "Algorithms for model checking hyperltl and hyperctl ^*," in *Computer Aided Verification - 27th International Conference, CAV 2015, San Francisco, CA, USA, July 18-24, 2015, Proceedings, Part I*, ser. LNCS, D. Kroening and C. S. Pasareanu, Eds., vol. 9206. Springer, 2015, pp. 30–48. [Online]. Available: https://doi.org/10.1007/978-3-319-21690-4_3

[17] O. Finkel, "Borel hierarchy and omega context free languages," *Theor. Comput. Sci.*, vol. 290, pp. 1385–1405, 2003. [Online]. Available: https://api.semanticscholar.org/CorpusID:11381251

[18] B. Goldlücke, "Variational analysis," in *Computer Vision, A Reference Guide*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:6223674

[19] J. Y. Halpern, "A modification of the halpern-pearl definition of causality," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Q. Yang and M. J. Wooldridge, Eds. AAAI Press, 2015, pp. 3022–3033. [Online]. Available: http://ijcai.org/Abstract/15/427

[20] ——, *Actual Causality*. MIT Press, 2016.

[21] J. Y. Halpern and J. Pearl, "Causes and explanations: A structural-model approach: Part 1: Causes," in *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, University of Washington, Seattle, Washington, USA, August 2-5, 2001*, J. S. Breese and D. Koller, Eds. Morgan Kaufmann, 2001, pp. 194–202.

[22] F. Hausdorff, "Grundzüge der mengenlehre," 1914. [Online]. Available: https://api.semanticscholar.org/CorpusID:170951128

[23] J. E. Hopcroft and J. D. Ullman, "Introduction to automata theory, languages and computation," 1979. [Online]. Available: https://api.semanticscholar.org/CorpusID:31901407

[24] O. Kupferman and M. Y. Vardi, "Model checking of safety properties," *Formal Methods in System Design*, vol. 19, pp. 291–314, 1999. [Online]. Available: https://api.semanticscholar.org/CorpusID:909779

[25] ——, *Synthesis with Incomplete Informatio*. Dordrecht: Springer Netherlands, 2000, pp. 109–127. [Online]. Available: https://doi.org/10.1007/978-94-015-9586-5_6

[26] ——, "Synthesizing distributed systems," in *16th Annual IEEE Symposium on Logic in Computer Science, LICS 2001, Boston, Massachusetts, USA, June 16-19, 2001, Proceedings*. IEEE Computer Society, 2001, pp. 389–398. [Online]. Available: https://doi.org/10.1109/LICS.2001.932514

[27] F. Leitner-Fischer, "Causality checking of safety-critical software and systems," Ph.D. dissertation, University of Konstanz, Germany, 2015. [Online]. Available: http://kops.uni-konstanz.de/handle/123456789/30778

[28] F. Leitner-Fischer and S. Leue, "Causality checking for complex system models," in *Verification, Model Checking, and Abstract Interpretation, 14th International Conference, VMCAI 2013, Rome, Italy, January 20-22, 2013. Proceedings*, ser. LNCS, R. Giacobazzi, J. Berdine, and I. Mastroeni, Eds., vol. 7737. Springer, 2013, pp. 248–267. [Online]. Available: https://doi.org/10.1007/978-3-642-35873-9_16

[29] D. K. Lewis, *Counterfactuals*. Cambridge, MA, USA: Blackwell, 1973.

[30] O. Lichtenstein, A. Pnueli, and L. D. Zuck, "The glory of the past," in *Logics of Programs, Conference, Brooklyn College, New York, NY, USA, June 17-19, 1985, Proceedings*, ser. Lecture Notes in Computer Science, R. Parikh, Ed., vol. 193. Springer, 1985, pp. 196–218. [Online]. Available: https://doi.org/10.1007/3-540-15648-8_16

[31] Z. Manna and A. Pnueli, "A hierarchy of temporal properties," in *Proceedings of the Ninth Annual ACM Symposium on Principles of Distributed Computing, Quebec City, Quebec, Canada, August 22-24, 1990*, C. Dwork, Ed. ACM, 1990, pp. 377–410. [Online]. Available: https://doi.org/10.1145/93385.93442

[32] C. Mascle, C. Baier, F. Funke, S. Jantsch, and S. Kiefer, "Responsibility and verification: Importance value in temporal logics," in *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 2021, pp. 1–14. [Online]. Available: https://doi.org/10.1109/LICS52264.2021.9470597

[33] E. Michael, "Topologies on spaces of subsets," *Transactions of the American Mathematical Society*, vol. 71, pp. 152–182, 1951. [Online]. Available: https://api.semanticscholar.org/CorpusID:11182986
[34] S. Miyano and T. Hayashi, "Alternating finite automata on omega-words," *Theor. Comput. Sci.*, vol. 32, pp. 321–330, 1984. [Online]. Available: https://api.semanticscholar.org/CorpusID:6926486
[35] R. K. Mothilal, D. Mahajan, C. Tan, and A. Sharma, "Towards unifying feature attribution and counterfactual explanations: Different means to the same end," in *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, M. Fourcade, B. Kuipers, S. Lazar, and D. K. Mulligan, Eds. ACM, 2021, pp. 652–663. [Online]. Available: https://doi.org/10.1145/3461702.3462597
[36] J. Parreaux, J. Piribauer, and C. Baier, "Counterfactual causality for reachability and safety based on distance functions," in *Proceedings of the Fourteenth International Symposium on Games, Automata, Logics, and Formal Verification, GandALF 2023, Udine, Italy, 18-20th September 2023*, ser. EPTCS, A. Achilleos and D. D. Monica, Eds., vol. 390, 2023, pp. 132–149. [Online]. Available: https://doi.org/10.4204/EPTCS.390.9
[37] D. Perrin and J. Pin, *Infinite words - automata, semigroups, logic and games*, ser. Pure and applied mathematics series. Elsevier Morgan Kaufmann, 2004, vol. 141.
[38] G. D. Plotkin, "A powerdomain construction," *SIAM J. Comput.*, vol. 5, pp. 452–487, 1976. [Online]. Available: https://api.semanticscholar.org/CorpusID:18371115
[39] ——, "A powerdomain for countable non-determinism (extended abstract)," in *International Colloquium on Automata, Languages and Programming*, 1982. [Online]. Available: https://api.semanticscholar.org/CorpusID:16699700
[40] A. Pnueli, "The temporal logic of programs," in *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*. IEEE Computer Society, 1977, pp. 46–57. [Online]. Available: https://doi.org/10.1109/SFCS.1977.32
[41] W. Rudin, "Principles of mathematical analysis," 1964. [Online]. Available: https://api.semanticscholar.org/CorpusID:50742905
[42] M. B. Smyth, "Power domains," *J. Comput. Syst. Sci.*, vol. 16, pp. 23–36, 1978. [Online]. Available: https://api.semanticscholar.org/CorpusID:27946887
[43] ——, "Power domains and predicate transformers: A topological view," in *International Colloquium on Automata, Languages and Programming*, 1983. [Online]. Available: https://api.semanticscholar.org/CorpusID:262041251
[44] R. Stalnaker, *A Theory of Conditionals*. Dordrecht: Springer Netherlands, 1981, pp. 41–55. [Online]. Available: https://doi.org/10.1007/978-94-009-9117-0_2
[45] M. Y. Vardi, "The büchi complementation saga," in *Symposium on Theoretical Aspects of Computer Science*, 2007. [Online]. Available: https://api.semanticscholar.org/CorpusID:763052
[46] L. Vietoris, "Monatsh. f. math. u. phys. 31," 1921.

APPENDIX

A. Detailed Proofs

Lemma 6. *An NFW for L_n requires at least $2^{2^{\Omega(n)}}$ states.*

Proof. Assume A is an NFW that recognizes language L_n. Let Y be a set of sets of binary words of length n of size 2^{n-1}:

$$Y = \{S \subseteq \{0,1\}^n \mid |S| = 2^{n-1}\} \ .$$

For each $y \in Y$ after reading a finite word π with $subword_n(\pi) = y$ the NFW A must reach at least one such state from which it accepts a word $\overline{\pi}$ with $subword_n(\overline{\pi}) = \{0,1\}^n \setminus y$. Such states must be different for different y, hence $|A| \geq |Y| = \binom{2^n}{2^{n-1}} = 2^{2^{\Omega(n)}}$. □

Theorem 7. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and an effect E given as a DBW, there exists a DBW for the cause of the size $\mathcal{O}(|\pi| \cdot 3^{(|\mathcal{T}|\cdot|E|)})$.*

Proof. We construct a UBW U of the size $|\mathcal{T}| \cdot |E|$, such that U accepts pairs of traces $\pi_1 \in (2^I)^\omega$ and $\pi_2 \in (2^{I \cup O})^\omega$, where π_1 is in the cause of E on π_2 in $\mathcal{T}$. Afterward, U can be translated to the DBW of the size $\mathcal{O}(3^{|U|})$ by [34]. Combining it with π we get an automaton for the cause. We denote $\mathcal{T} = (S, s_0, AP, \delta, l)$ and $E = (Q, 2^{AP}, q_0, F, \Delta)$. Let us define $U := (S \times Q, \Sigma, s_0 \times q_0, \Delta_U, S \times F)$ over the alphabet $\Sigma := 2^I \times 2^{AP}$. Here transition relation Δ_U is defined for $s \in S$, $q \in Q$, $I'_1, I'_2 \subseteq I$ and $O' \subseteq O$ as follows:

$$\Delta_U((s,q),(I'_1,I'_2,O')) := \{(s',q') \mid \exists I' \subseteq I : \ I'_1 \cap I'_2 \subseteq I' \subseteq I'_1 \cup I'_2, \ s' \in \delta(s,I'), \ \Delta(q,I',l(s')) = q'\} \ .$$

It is easy to see that runs of U on π_1, π_2 correspond to runs of E on traces π_3 such that $\pi_3 \leq_{\pi_2} \pi_1$. Hence, U accepts π_1, π_2 iff E accepts all such π_3, which means that π_2 is in the cause of E on π_1. □

Theorem 8. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and a safety effect E given as a DFW E_{bad_pref} for the bad prefixes of E, there exists a DFW for the bad prefixes of the cause of the size $\mathcal{O}(|\pi| \cdot 2^{(|\mathcal{T}|\cdot|E_{bad\_pref}|)})$.*

Proof. The proof is similar to the proof of Theorem 7, but now we work with finite word languages.

First, we construct the NFW A over the alphabet $(2^I)^* \times (2^{I \cup O})^*$, which recognizes such pairs of words $w_1 \in (2^I)^*$ and $w_2 \in (2^{I \cup O})^*$, that there exists a trace $w \leq^{subset}_{w_2} w_1$ and $w \in E_{bad\_pref}$.

Denote the system $\mathcal{T} = (S, s_0, AP, \delta, l)$ and the effect $E_{bad\_pref} = (Q, 2^{AP}, q_0, F, \Delta)$ and denote the cause as C.

$$A := (S \times Q, (2^I)^* \times (2^{I \cup O})^*, s_0 \times q_0, \Delta_U, S \times F)$$

The transition function Δ_U is defined as follows:

$$\Delta_U((s,q),(I'_1,I'_2,O')) := \{(s',q') \mid \exists I' \subseteq I : \ I'_1 \cap I'_2 \subseteq I' \subseteq I'_1 \cup I'_2, \ s' \in \delta(s,I'), \ \Delta(q,I',l(s')) = q'\} \ .$$

We build the DFW of the size $2^{|A|} = 2^{|\mathcal{T}|\cdot|E_{bad\_pref}|}$ which recognizes the same language as A and combine it with π getting the automaton C'_{bad_pref}. Obviously, C'_{bad_pref} accepts only bad prefixes of C. But unfortunately, there may be some bad prefixes of C which C'_{bad_pref} does not accept.

For every $\pi'' \notin C$ there exists a trace $\pi' \leq^{subset}_{\pi} \pi''$, such that $\pi' \notin E$. Hence, π' has a prefix from E_{bad_pref}. Thus, π'' has a prefix from C'_{bad_pref}.

The last step is to denote accepting all states of C'_{bad_pref} from which every infinite trace eventually visits an accepting state getting the automaton C_{bad_pref}. □

Theorem 9. *For a system $\mathcal{T}$, a similarity relation $\leq^{subset}$, a trace π and a guarantee effect E given as a DFW E_{good_pref} for the good prefixes of E, there exists a DFW for the good prefixes of the cause of the size $\mathcal{O}(|\pi| \cdot 2^{(|\mathcal{T}|\cdot|E_{good\_pref}|)})$.*

Proof. By Corollary 1 the cause must also be a guarantee property. Moreover, by Theorem 1 the set C_{pairs} of pairs of traces $\pi_1 \in (2^I)$ and $\pi_2 \in (2^{I \cup O})^\omega$, such that π_1 is in the cause of π_2, is also a guarantee property, since it can be universally projected on the set E in the similar way as in Proposition 1.

First, we construct the UFW U of the size $|\mathcal{T}| \cdot |E_{good\_pref}|$ that recognizes the good prefixes of C_{pairs}.

Denote $\mathcal{T} = (S, s_0, AP, \delta, l)$ and $E_{good\_pref} = (Q, 2^{AP}, q_0, F, \Delta)$ and denote the cause as C.

Let us define $U = (S \times Q, \Sigma, s_0 \times q_0, \Delta_U, S \times F)$ over the alphabet $\Sigma := 2^I \times 2^{AP}$. Here transition relation Δ_U is defined for $s \in S$, $q \in Q$, $I'_1, I'_2 \subseteq I$ and $O' \subseteq O$ as follows:

$$\Delta_U((s,q),(I'_1,I'_2,O')) := \{(s',q') \mid \exists I' \subseteq I : \; I'_1 \cap I'_2 \subseteq I' \subseteq I'_1 \cup I'_2, \; s' \in \delta(s,I'), \; \Delta(q,I',l(s')) = q'\} \; .$$

Turning this U to DFW of the size $2^{|U|} = 2^{|\mathcal{T}| \cdot |E_{good\_pref}|}$ and combining it with π we get the DFW for the good prefixes of the cause. $\square$