Explaining Hyperproperty Violations

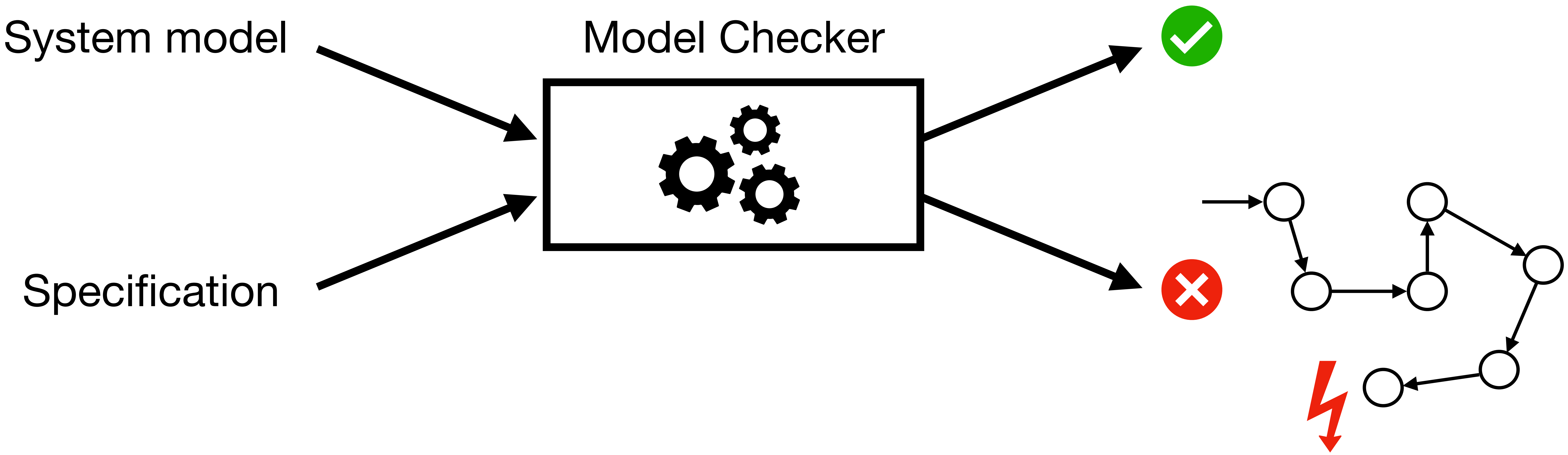
Norine Coenen[1], Raimund Dachselt[2], Bernd Finkbeiner[1], Hadar Frenkel[1],
Christopher Hahn[1], Tom Horak[2], Niklas Metzger[1], *Julian Siber*[1]

[1]CISPA Helmholtz Center for Information Security

[2]Technische Universität Dresden

As presented at the 34th International Conference on Computer-Aided Verification (CAV 2022).

Model Checking

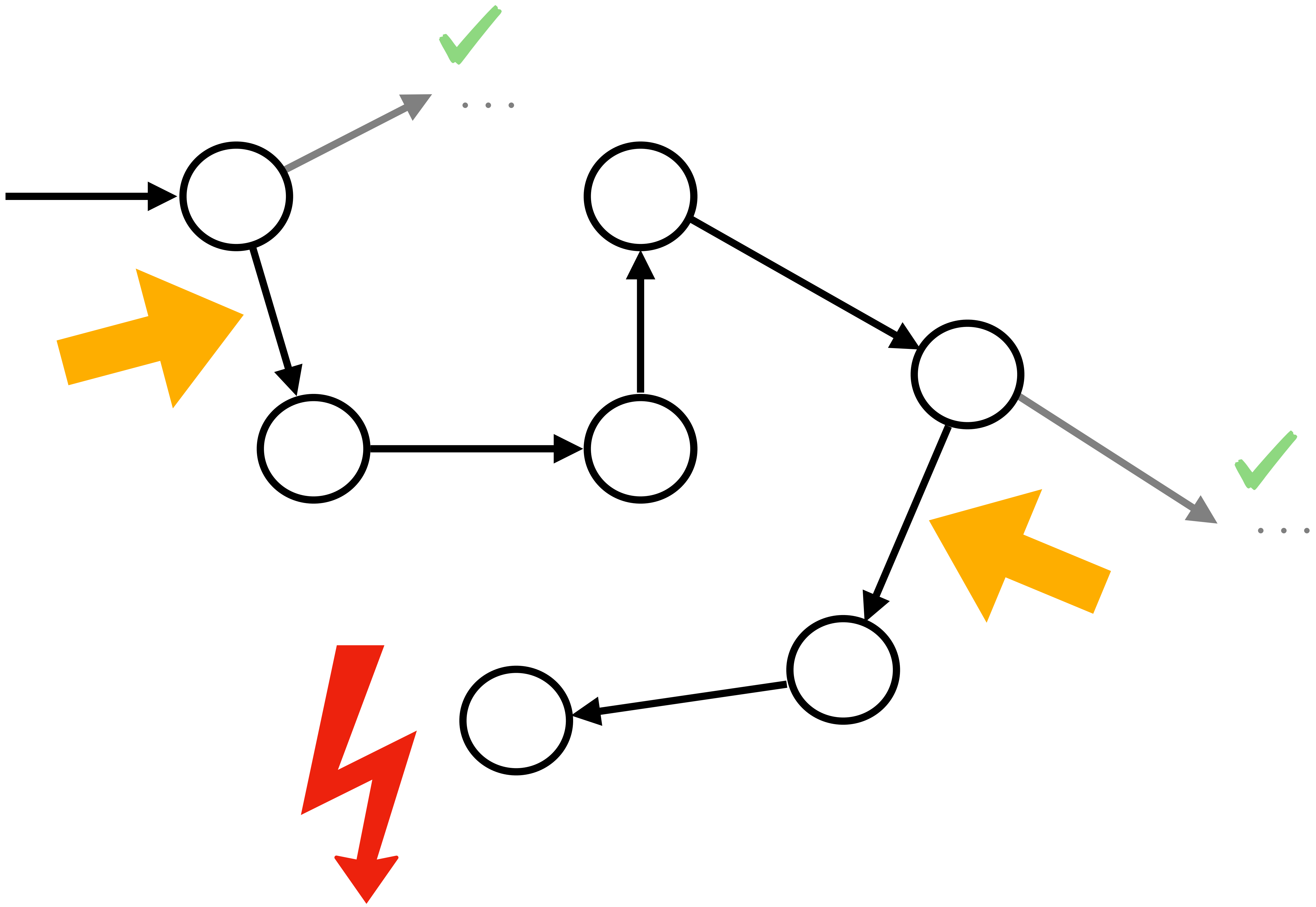


System model

Specification

Model Checker

Explaining Counterexamples



We give explanations by identifying *causes* in the non-deterministic input sequences.

E.g.: *Explaining Counterexamples Using Causality.* Beer, Ben-David, Chockler, Orni, and Trefler. (CAV 2009).
Error explanation with distance metrics. Groce, Chaki, Kroening, Strichman. Int. J. Softw. Tools Technol. Transf. **8** (2006)

Hyperproperties

- Observational determinism: *“A system appears deterministic to low-security users”.*

- (Generalized) Noninterference

- Declassification

- Trace equality

Hyperproperties. Clarkson and Schneider. (CSF 2008).

HyperLTL

- Observational determinism: *“A system appears deterministic to low-security users”.*

$$\forall \pi . \forall \pi' . \square (li_\pi \leftrightarrow li_{\pi'}) \rightarrow \square (lo_\pi \leftrightarrow lo_{\pi'})$$

- (Generalized) Noninterference
- Declassification
- Trace equality

Temporal Logics for Hyperproperties. Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).

HyperLTL Model Checking

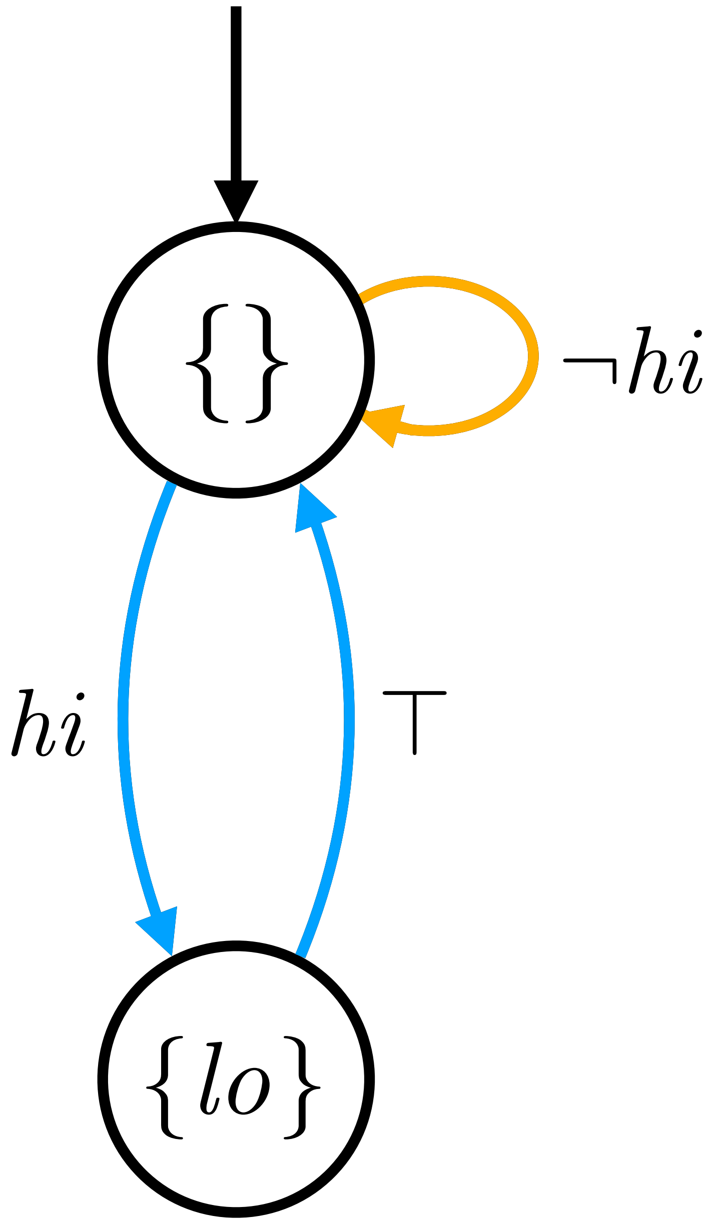
Specification φ :

$$\forall\pi.\forall\pi'.\square(li_\pi \leftrightarrow li_{\pi'}) \rightarrow \square(lo_\pi \leftrightarrow lo_{\pi'})$$

Counterexample Γ :

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

System T :



Violation of φ on Γ is due to interactions between inputs on multiple traces.

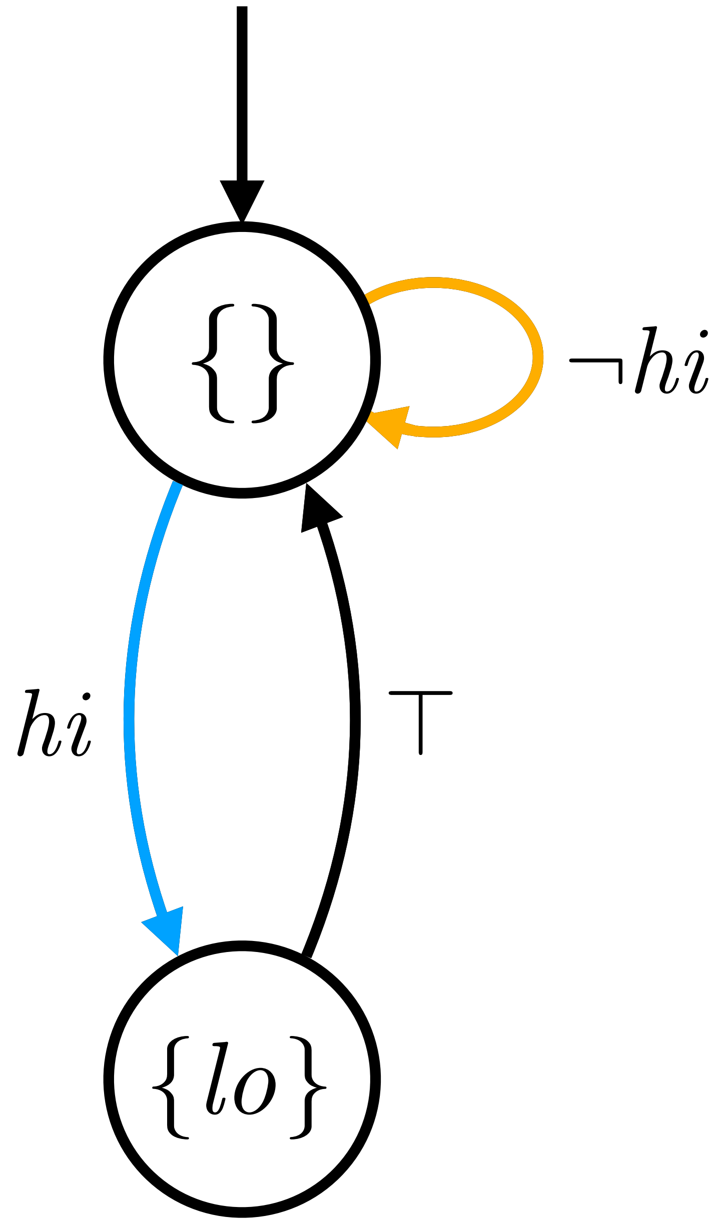
Temporal Logics for Hyperproperties. Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).
Algorithms for Model Checking HyperLTL and HyperCTL.* Finkbeiner, Rabe, and Sánchez. (CAV 2015).

HyperLTL Model Checking

Specification φ :

$$\forall\pi.\forall\pi'.\Box(li_{\pi} \leftrightarrow li_{\pi'}) \rightarrow \Box(lo_{\pi} \leftrightarrow lo_{\pi'})$$

System T :



Counterexample Γ :

$\pi =$	$\{\}$	$\{\}$	$\{\}^{\omega}$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^{\omega}$

Violation of φ on Γ is due to interactions between inputs on multiple traces.

Temporal Logics for Hyperproperties. Clarkson, Finkbeiner, Koleini, Micinski, Rabe, and Sánchez. (POST 2014).
Algorithms for Model Checking HyperLTL and HyperCTL.* Finkbeiner, Rabe, and Sánchez. (CAV 2015).

Causal Analysis

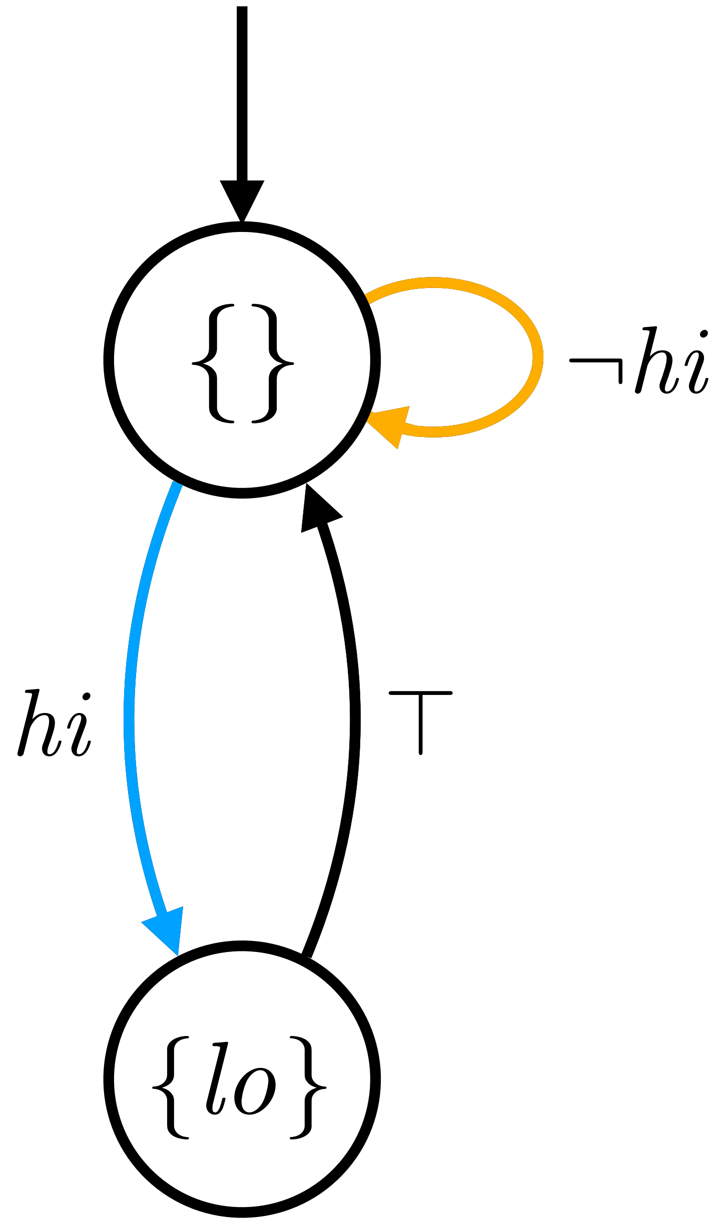
Specification φ :

$$\forall \pi . \forall \pi' . \Box (li_{\pi} \leftrightarrow li_{\pi'}) \rightarrow \Box (lo_{\pi} \leftrightarrow lo_{\pi'})$$

Counterexample Γ :

$\pi =$	$\{\}$	$\{\}$	$\{\}^{\omega}$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^{\omega}$

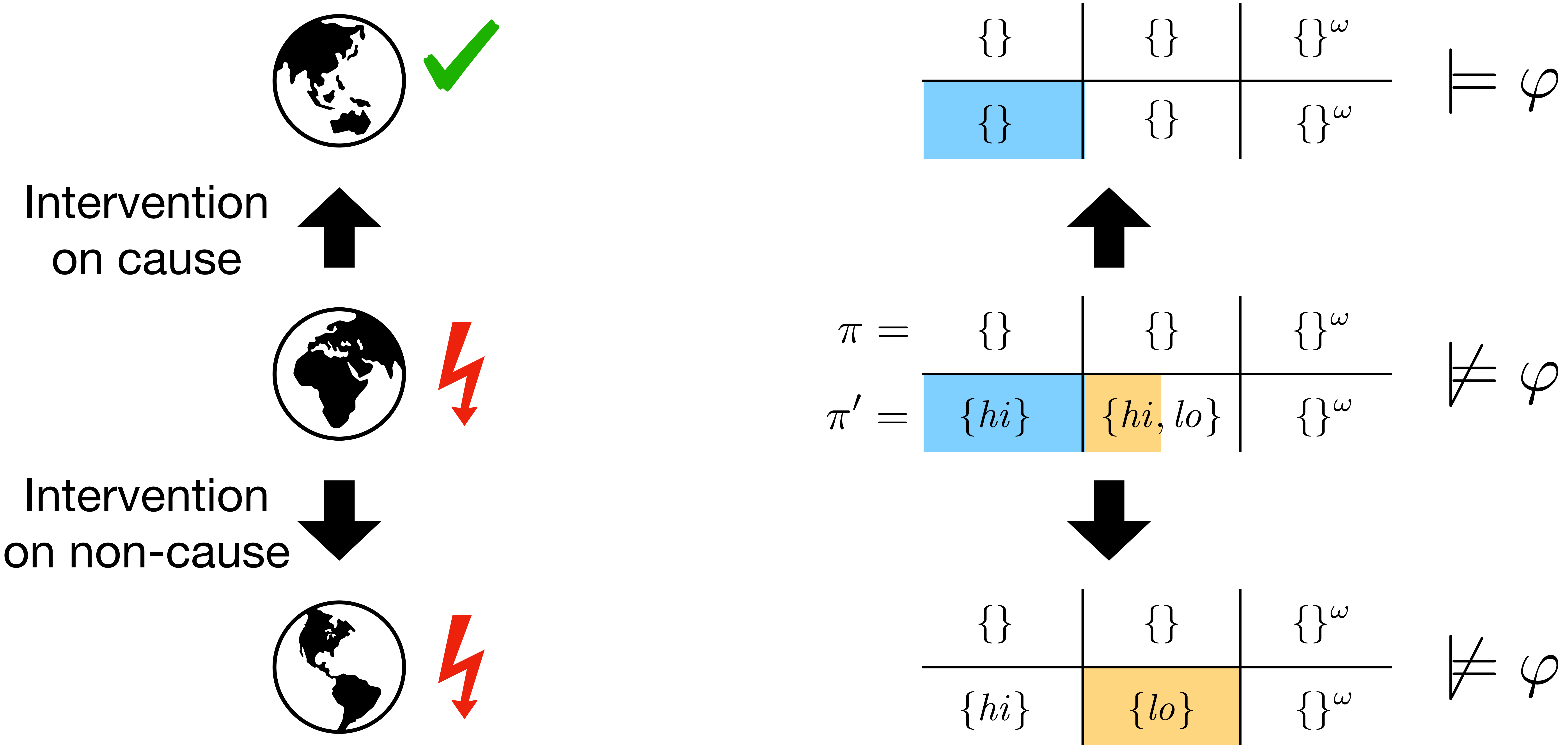
System T :



We highlight the *causes* on the input sequences.

Causes and Explanations: A Structural-Model Approach. Halpern and Pearl. Brit. J. Phil. Sci. **56** (2005).
A Modification of the Halpern-Pearl Definition of Causality. Halpern. (IJCAI 2015).

Counterfactual Reasoning



	$\{\}$	$\{\}$	$\{\}^\omega$
	$\{\}$	$\{\}$	$\{\}^\omega$

$\models \varphi$

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

$\not\models \varphi$

	$\{\}$	$\{\}$	$\{\}^\omega$
	$\{hi\}$	$\{lo\}$	$\{\}^\omega$

$\not\models \varphi$

We extend HP's actual causality to hyperproperty effects and reactive systems.

Events and Causes

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

An *event* $\langle l_a, n, \pi \rangle$ is the value of an atomic proposition a at position n in π .

$$(\pi, \pi') \models \langle hi, 0, \pi' \rangle$$

Events and Causes

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

An *event* $\langle l_a, n, \pi \rangle$ is the value of an atomic proposition a at position n in π .

A *cause* $\mathcal{C}$ is a set of events.

$\mathcal{C}$ is a Cause if…

SAT CF MIN

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

SAT: Γ satisfies all events in $\mathcal{C}$.

Interventions

SAT CF MIN

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

An *intervention* on $\mathcal{C}$ *flips* the values of all events in $\mathcal{C}$.

$intervene(\Gamma, \{\langle hi, 0, \pi' \rangle\}, \emptyset)$ →

$\pi =$	$\{\}$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{\}$	$\{hi\}$	$\{lo\}$	$\{\}^\omega$

Contingencies

SAT CF MIN

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

A *contingency* $\mathcal{W}$ allows to reset states back to Γ.

$$intervene(\Gamma, \{\langle hi, 0, \pi' \rangle\}, \{\langle \neg lo, 2, \pi' \rangle\})$$

➡

$\pi =$	$\{\}$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{\}$	$\{hi\}$	$\{\}$	$\{\}^\omega$

$\mathcal{C}$ is a Cause if…

SAT CF MIN

$\pi =$	$\{\}$	$\{\}$	$\{\}^{\omega}$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^{\omega}$

SAT: Γ satisfies all events in $\mathcal{C}$.

CF: There exists a $\mathcal{W}$ and $\mathcal{C}' \subseteq \mathcal{C}$ s.t.: $intervene(\Gamma, \mathcal{C}', \mathcal{W}) \models \varphi$.

MIN: No $\mathcal{C}' \subset \mathcal{C}$ satisfies **SAT** and **CF**.

$\mathcal{C}$ is a Cause if…

SAT CF MIN

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

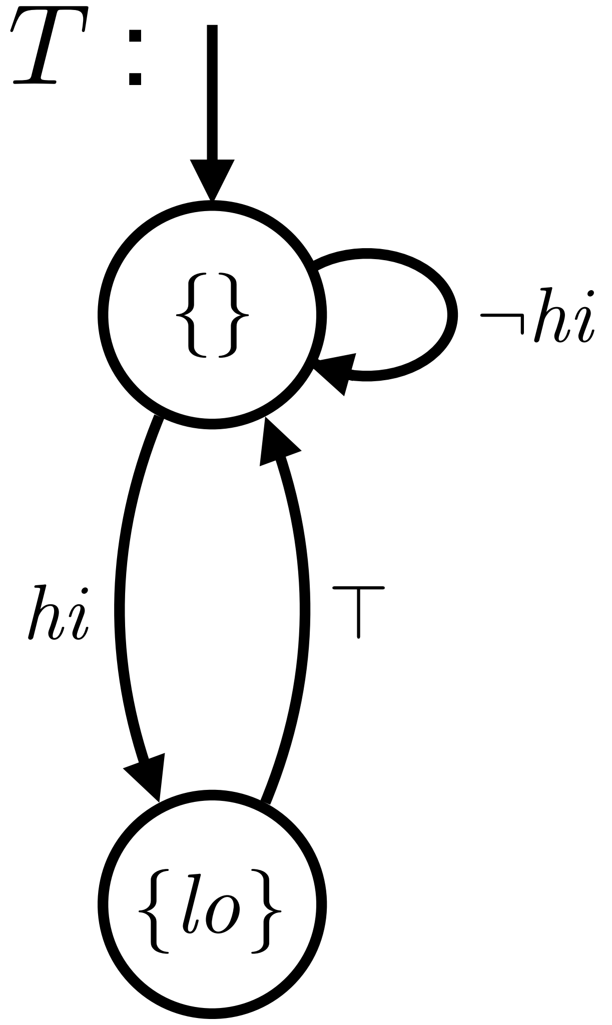
SAT: Γ satisfies all events in $\mathcal{C}$.

CF: There exists a $\mathcal{W}$ and $\mathcal{C}' \subseteq \mathcal{C}$ s.t.: $intervene(\Gamma, \mathcal{C}', \mathcal{W}) \models \varphi$.

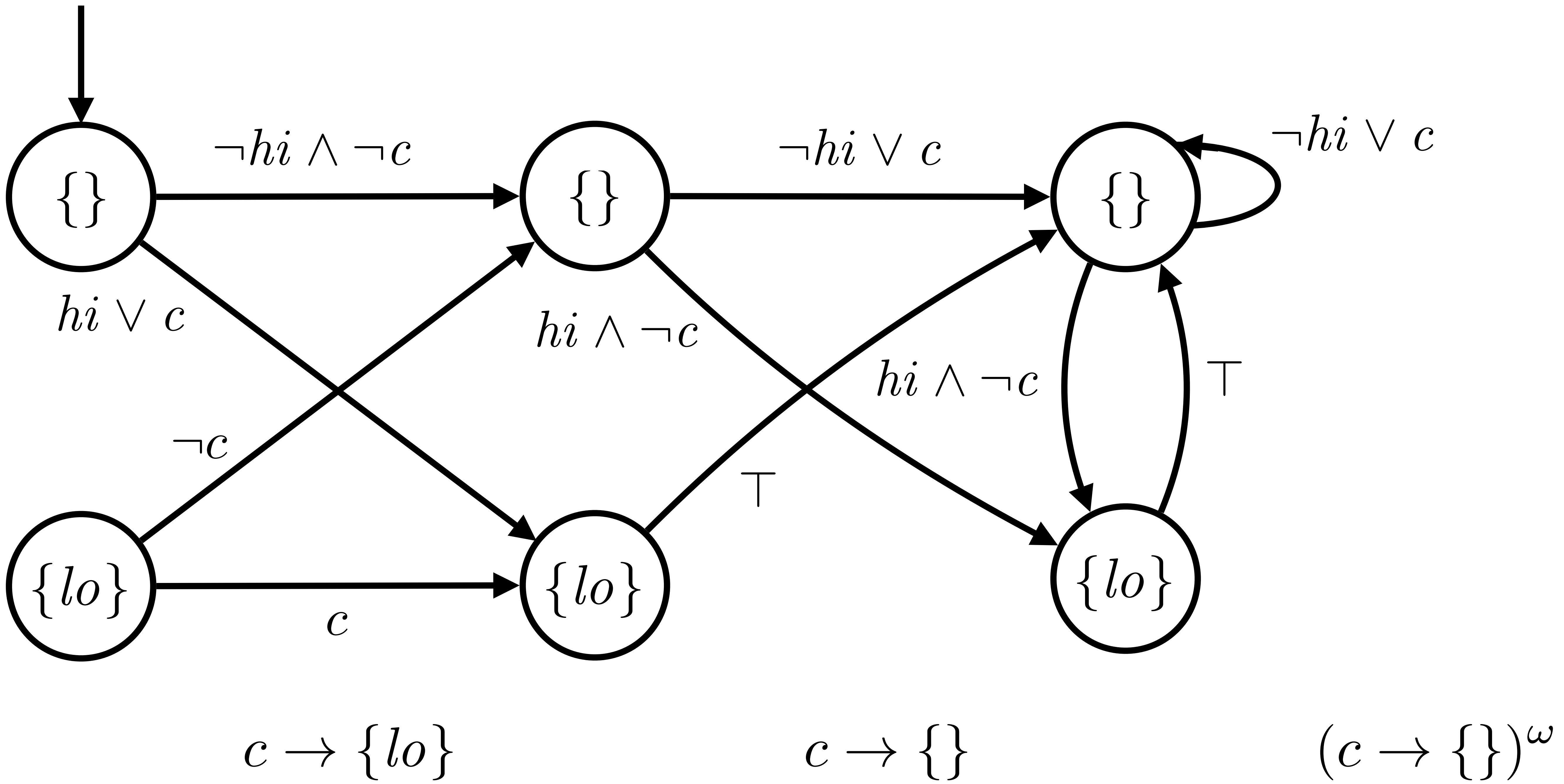
MIN: No $\mathcal{C}' \subset \mathcal{C}$ satisfies **SAT** and **CF**.

Encoding the CF Criterion

System T :



Counterfactual automaton (T, π') :

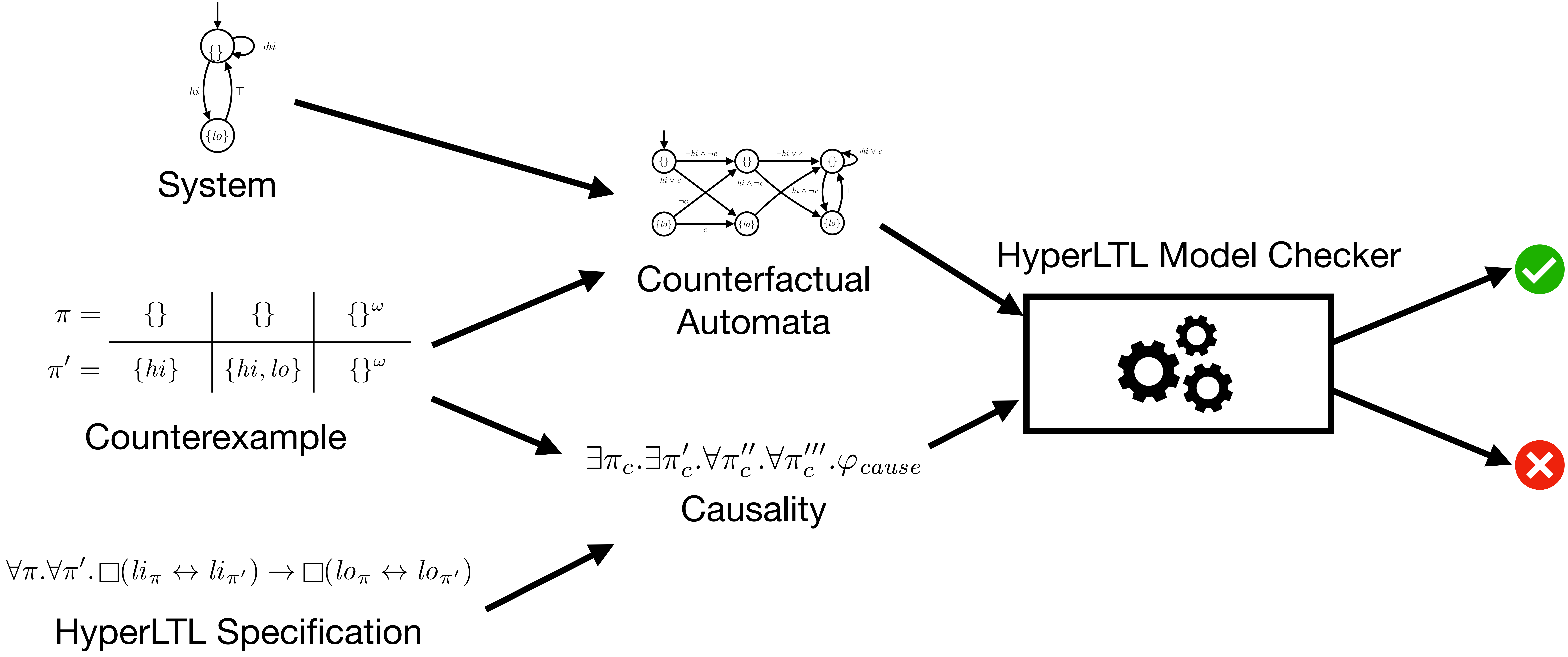


(Partial) counterexample:

$$\pi' = \{hi\}\{hi, lo\}\{\}^\omega$$

Counterfactual automata have additional inputs (here: c) for setting a contingency.

Finding a Cause as a Hyperproperty



System

$\pi =$	$\{\}$	$\{\}$	$\{\}^\omega$
$\pi' =$	$\{hi\}$	$\{hi, lo\}$	$\{\}^\omega$

Counterexample

$\forall \pi . \forall \pi' . \square(li_\pi \leftrightarrow li_{\pi'}) \rightarrow \square(lo_\pi \leftrightarrow lo_{\pi'})$

HyperLTL Specification

Counterfactual Automata

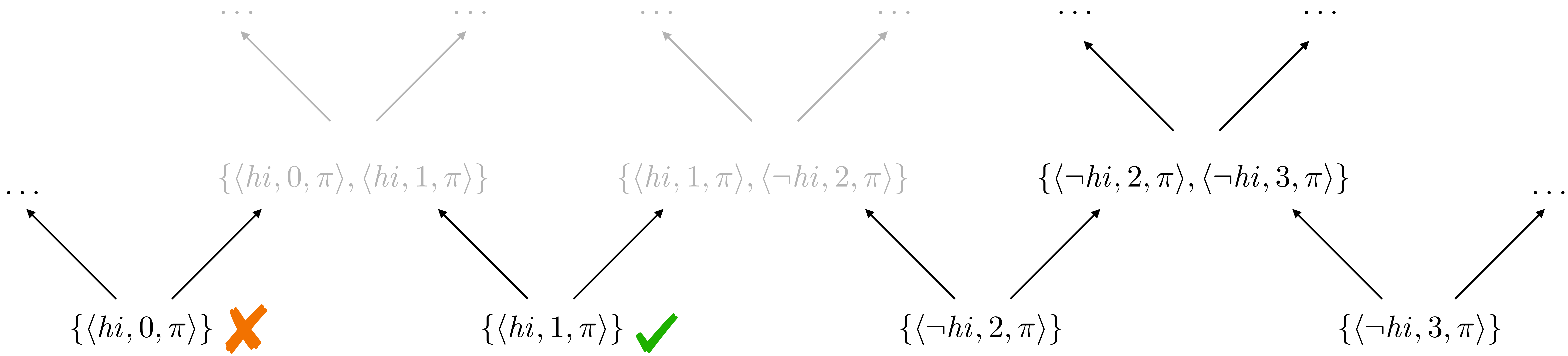
$\exists \pi_c . \exists \pi'_c . \forall \pi''_c . \forall \pi'''_c . \varphi_{cause}$

Causality

HyperLTL Model Checker

Encoding of causality in φ_{cause} : see our paper.

Computing All Causes



$\dots$ $\dots$ $\dots$ $\dots$ $\dots$ $\dots$

$\dots$ $\{\langle hi, 0, \pi\rangle, \langle hi, 1, \pi\rangle\}$ $\{\langle hi, 1, \pi\rangle, \langle \neg hi, 2, \pi\rangle\}$ $\{\langle \neg hi, 2, \pi\rangle, \langle \neg hi, 3, \pi\rangle\}$ $\dots$

$\{\langle hi, 0, \pi\rangle\}$ ✗ $\{\langle hi, 1, \pi\rangle\}$ ✓ $\{\langle \neg hi, 2, \pi\rangle\}$ $\{\langle \neg hi, 3, \pi\rangle\}$

If some $\mathcal{C}$ is a cause, then no strict superset $\mathcal{C}' \supset \mathcal{C}$ is a cause.

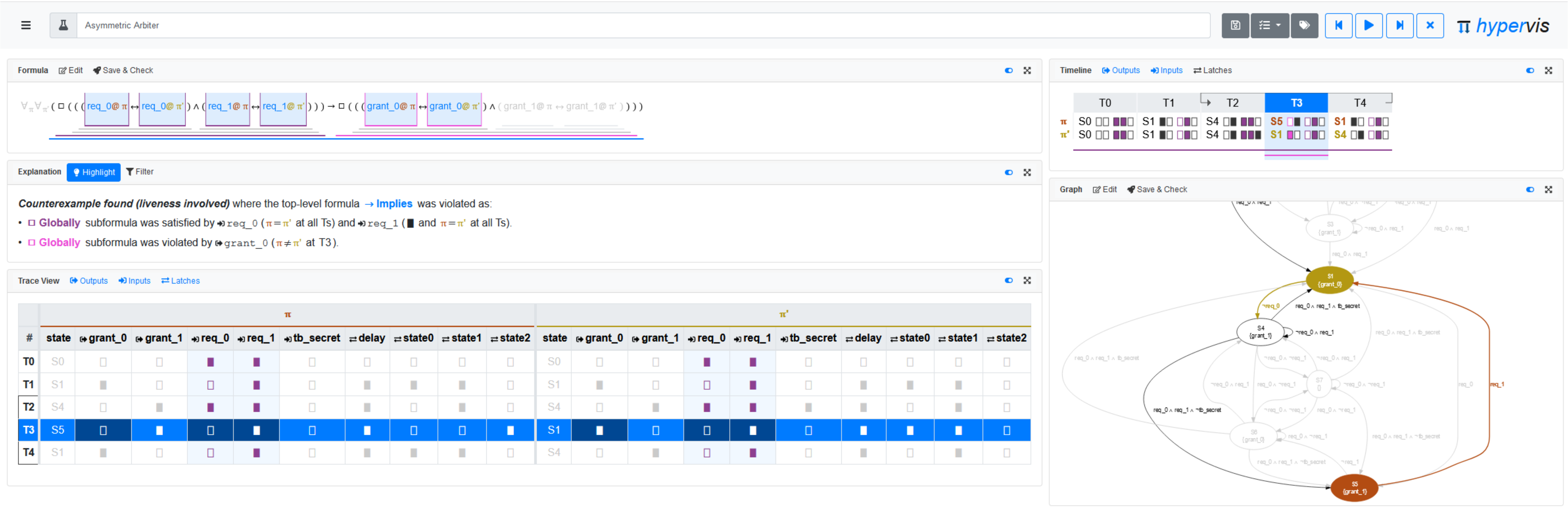
Experiments

Instance	$\|\Gamma\|$	$\|\varphi\|$	$\#(\mathcal{C})$	time(ms)
Running example (paper)	10	9	2	55
Security in & out	35	19	8	798
Drone example 1	24	19	5	367
Drone example 2	18	36	3	256
Asymmetric arbiter '19	28	35	10	490
Asymmetric arbiter	72	35	24	1480

CAV
Artifact
Evaluation
★
Available

CAV
Artifact
Evaluation
★ ★
Functional

HyperVis



$\forall_{\pi} \forall_{\pi'} (\square (((req\_0@\pi \leftrightarrow req\_0@\pi') \wedge (req\_1@\pi \leftrightarrow req\_1@\pi'))) \rightarrow \square (((grant\_0@\pi \leftrightarrow grant\_0@\pi') \wedge (grant\_1@\pi \leftrightarrow grant\_1@\pi'))))$

Counterexample found (liveness involved) where the top-level formula → **Implies** was violated as:

- □ **Globally** subformula was satisfied by `req_0` ($\pi = \pi'$ at all Ts) and `req_1` (■ and $\pi = \pi'$ at all Ts).
- □ **Globally** subformula was violated by `grant_0` ($\pi \neq \pi'$ at T3).

	π									π'									
#	state	grant_0	grant_1	req_0	req_1	tb_secret	delay	state0	state1	state2	state	grant_0	grant_1	req_0	req_1	tb_secret	delay	state0	state1	state2
T0	S0	□	□	■	■	□	□	□	□	□	S0	□	□	■	■	□	□	□	□	□
T1	S1	■	□	□	■	□	■	■	■	□	S1	■	□	□	■	□	■	■	■	□
T2	S4	□	■	■	■	□	■	□	■	□	S4	□	■	■	■	■	■	□	■	□
T3	S5	□	■	□	■	□	■	□	□	■	S1	■	□	□	■	□	■	■	■	□
T4	S1	■	□	□	■	□	■	■	■	□	S4	□	■	□	■	□	■	□	■	□

Visual Analysis of Hyperproperties for Understanding Model Checking Results. Horak, Coenen, Metzger, Hahn, Flemisch, Méndez, Dimov, Finkbeiner, and Dachselt. (VIS 2021). https://hypervis.tools.react.cs.uni-saarland.de

Conclusion

? Counterexamples of hyperproperties are *difficult to understand* and *debug*.

! Extending HP’s actual causality to hyperproperties gives *precise explanations*.

💡 *Causal inference* can itself be stated *as a hyperproperty* model-checking problem.

⚙ Symbolic causes, explicit relations, existential quantifiers