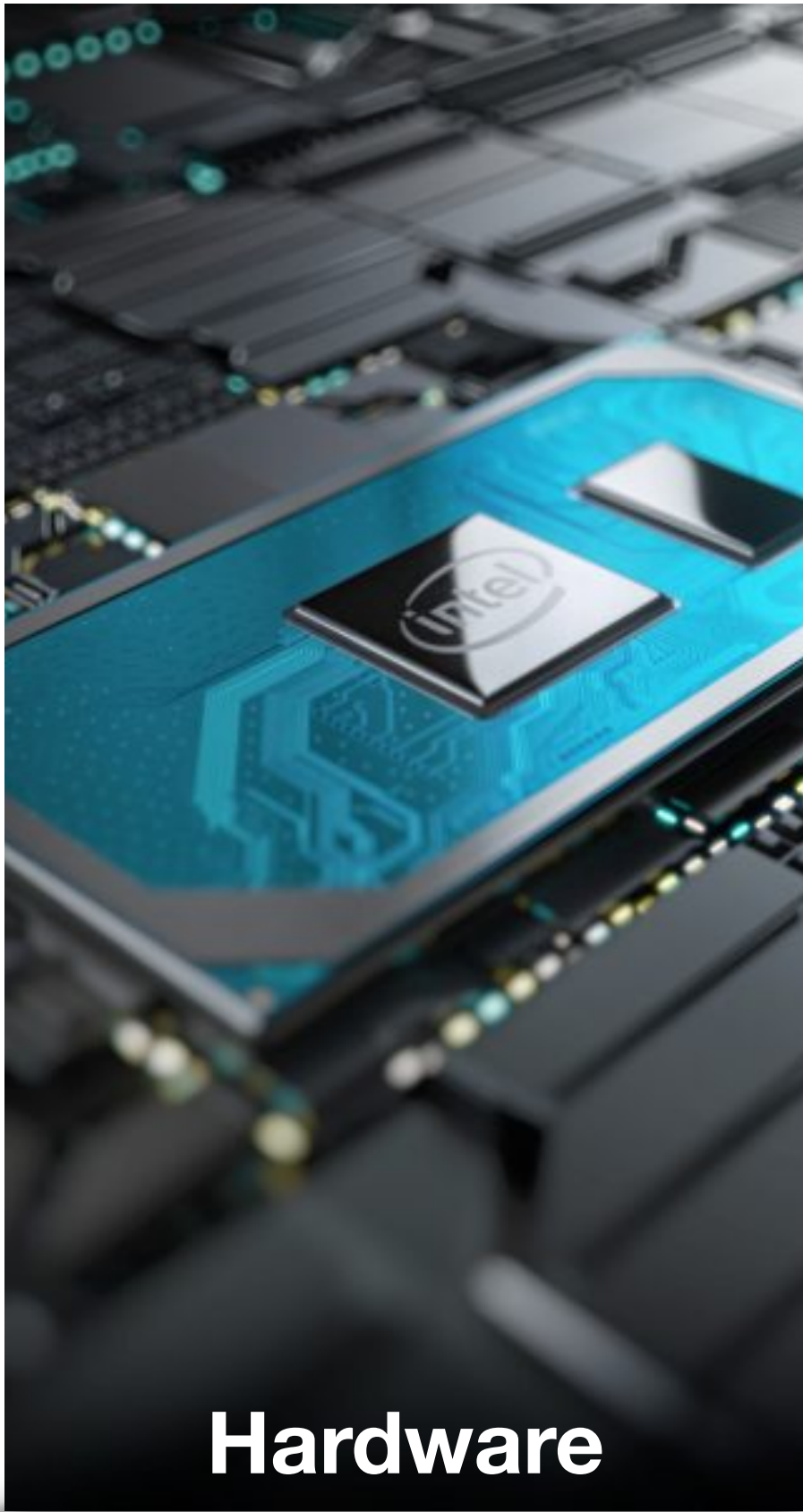




Hyperproperties for Privacy, Fairness, and Legal Requirements

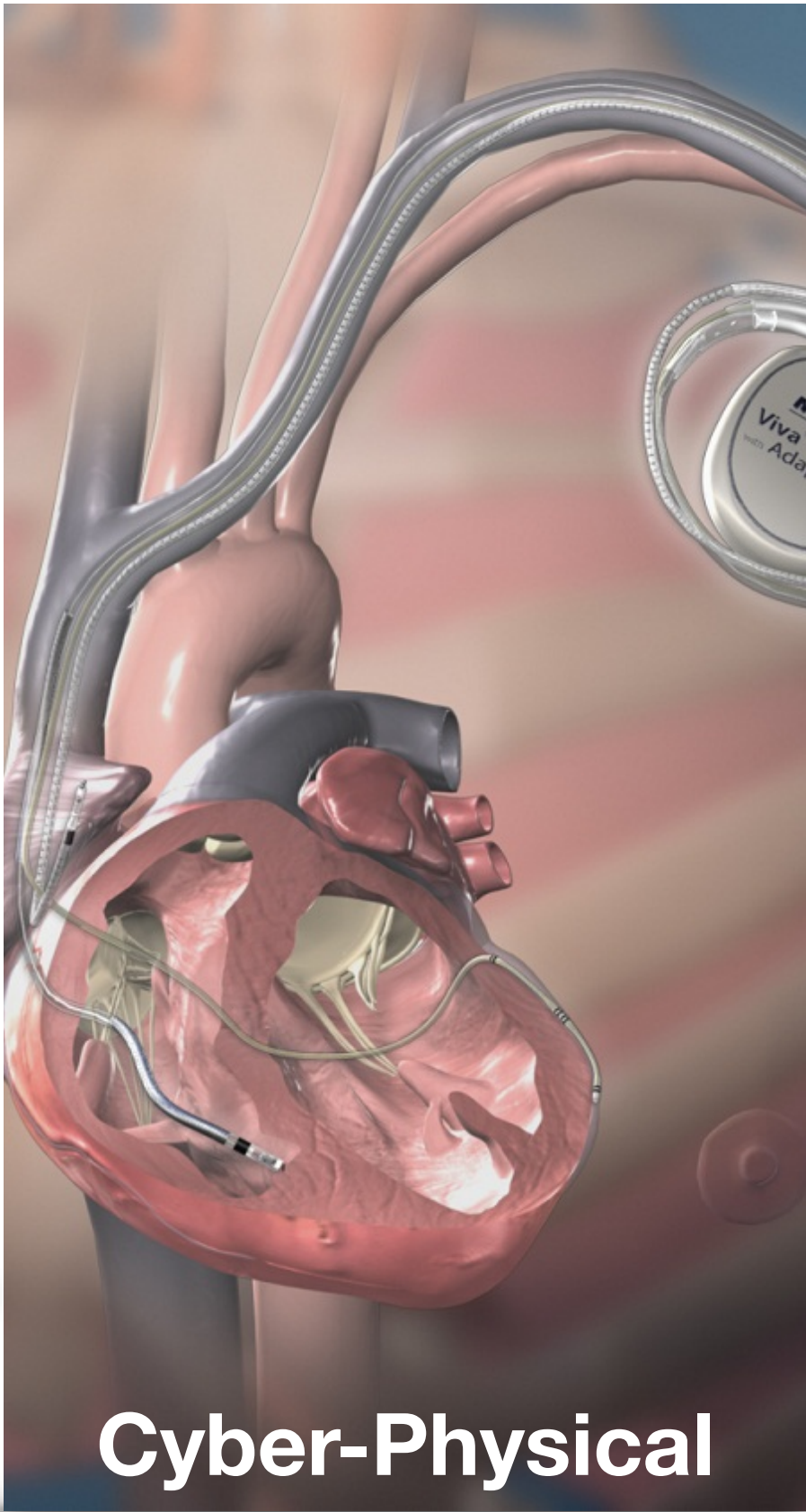
Ashutosh **Trivedi** (University of Colorado Boulder)



Hardware

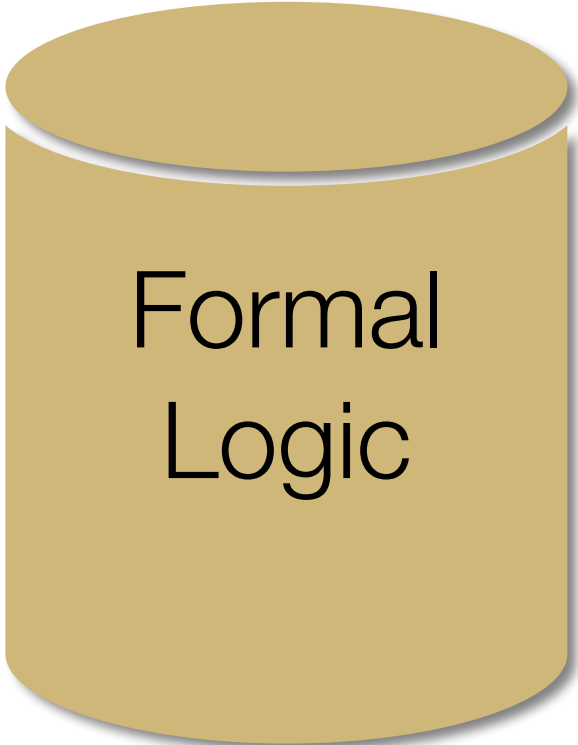


Software

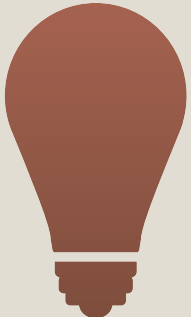


Cyber-Physical

Safety-Critical Systems

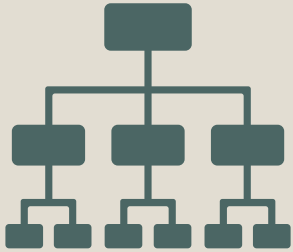


Formal
Logic



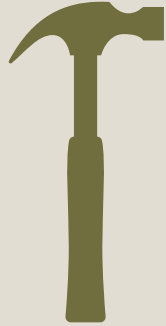
Techniques

- Automata-theoretic
- Abstraction-refinement
- Inductive invariants
- Barrier certificates



Algorithms

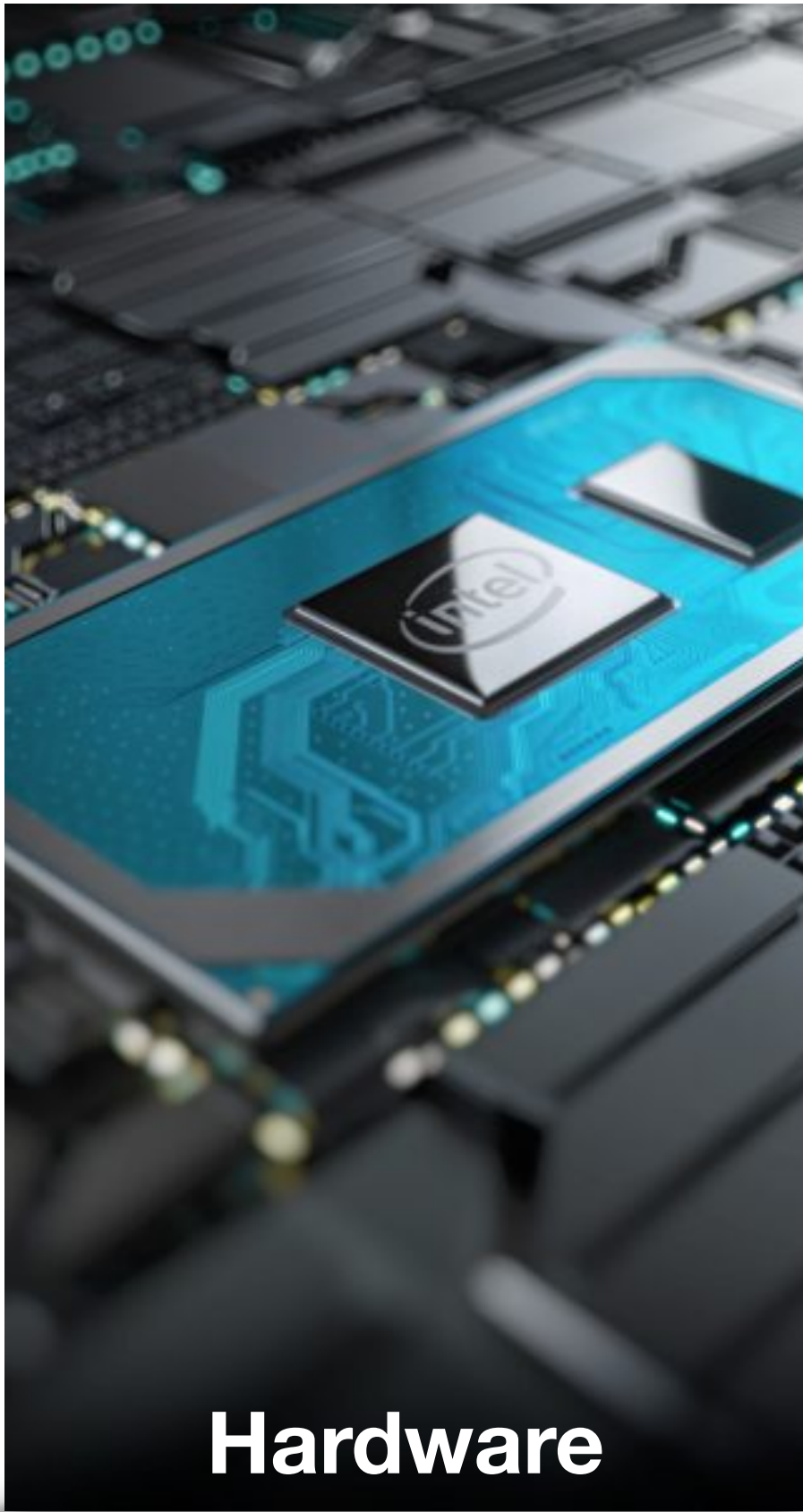
- Symbolic model checking
- Bounded model checking
- SAT/SMT/SOS



Tools

- NuSMV/PRISM/UPPAAL/..
- Phaver/DReal/Flow*
- PRISM/Storm/Owl/Mungojerrie

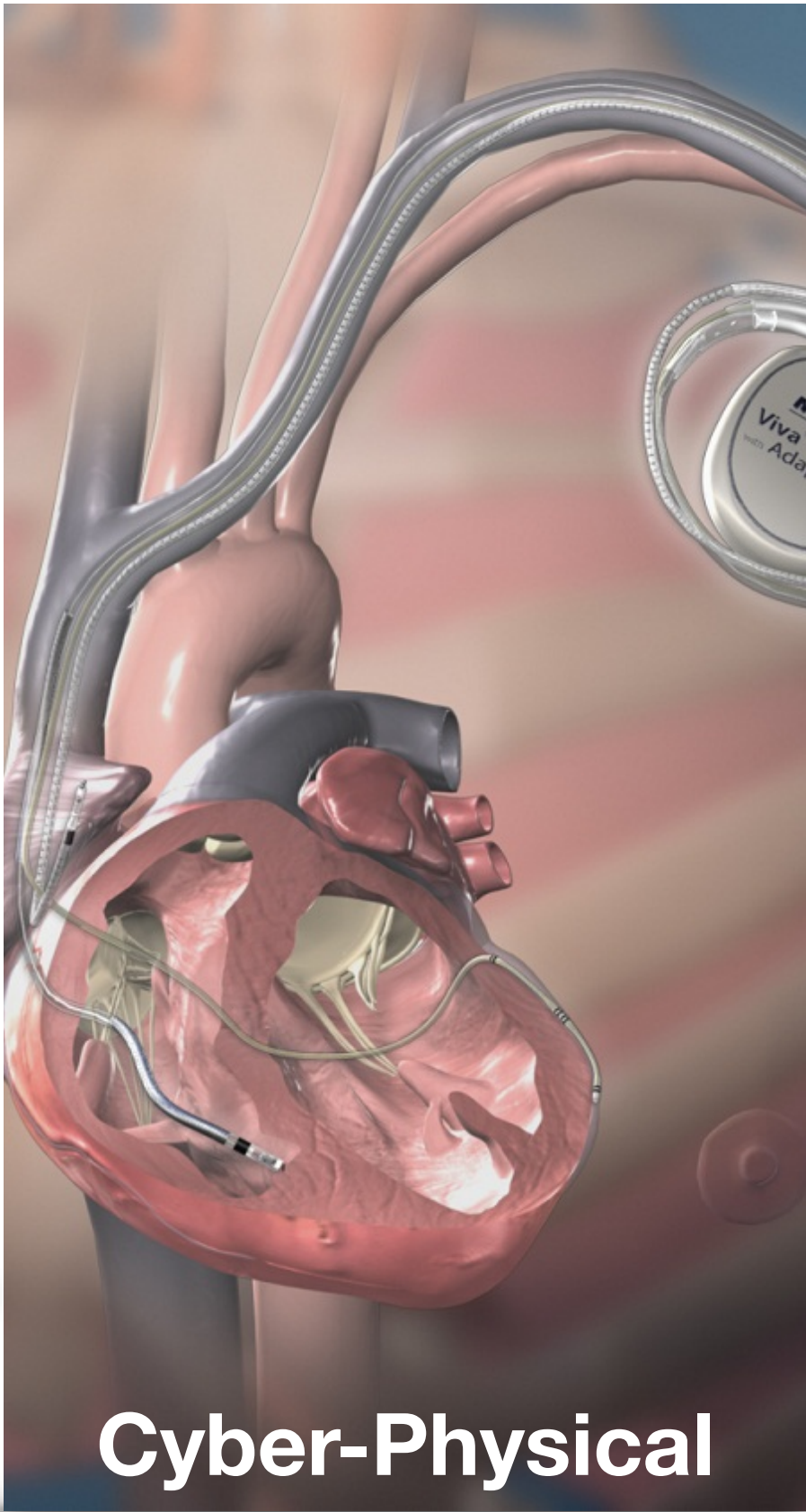
Formal Methods



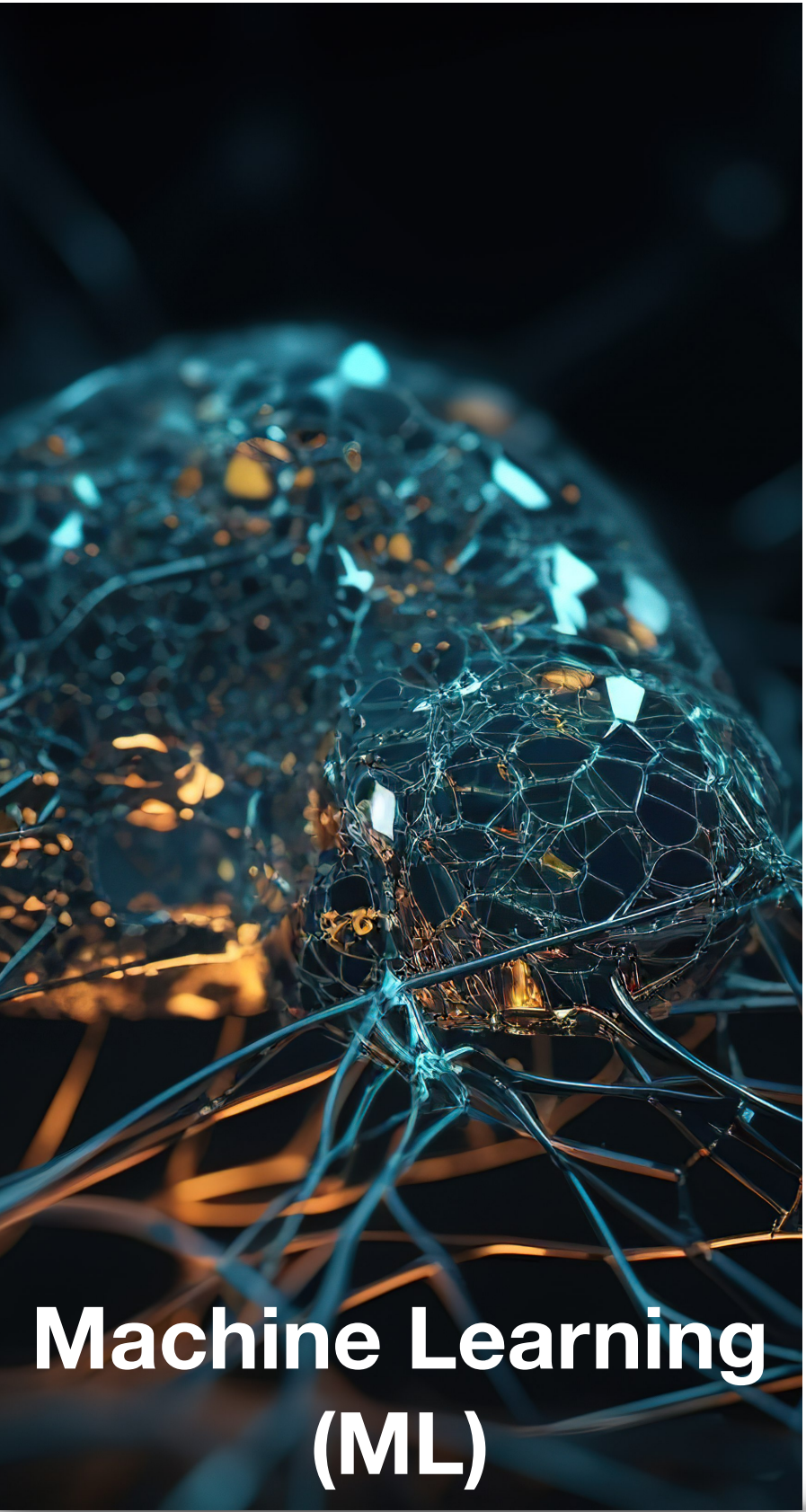
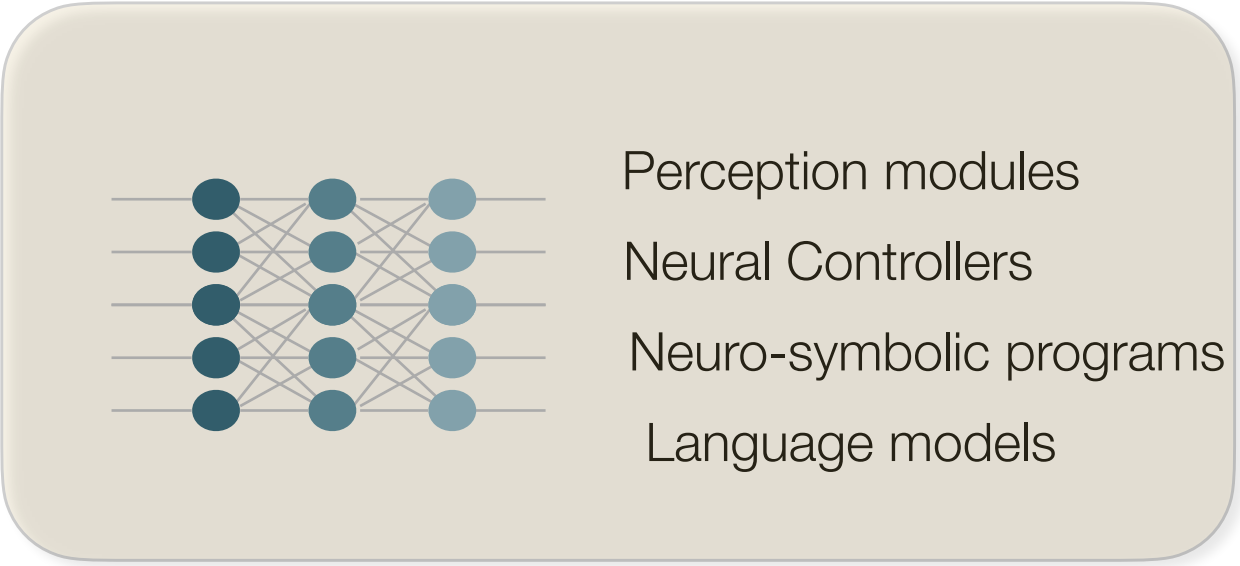
Hardware



Software



Cyber-Physical



Machine Learning
(ML)

Safety-Critical Systems



Safety



Privacy



Fairness



Law



Privacy, Fairness, and Legal Requirements



Privacy

Concerns beyond differential privacy: systemic leaks

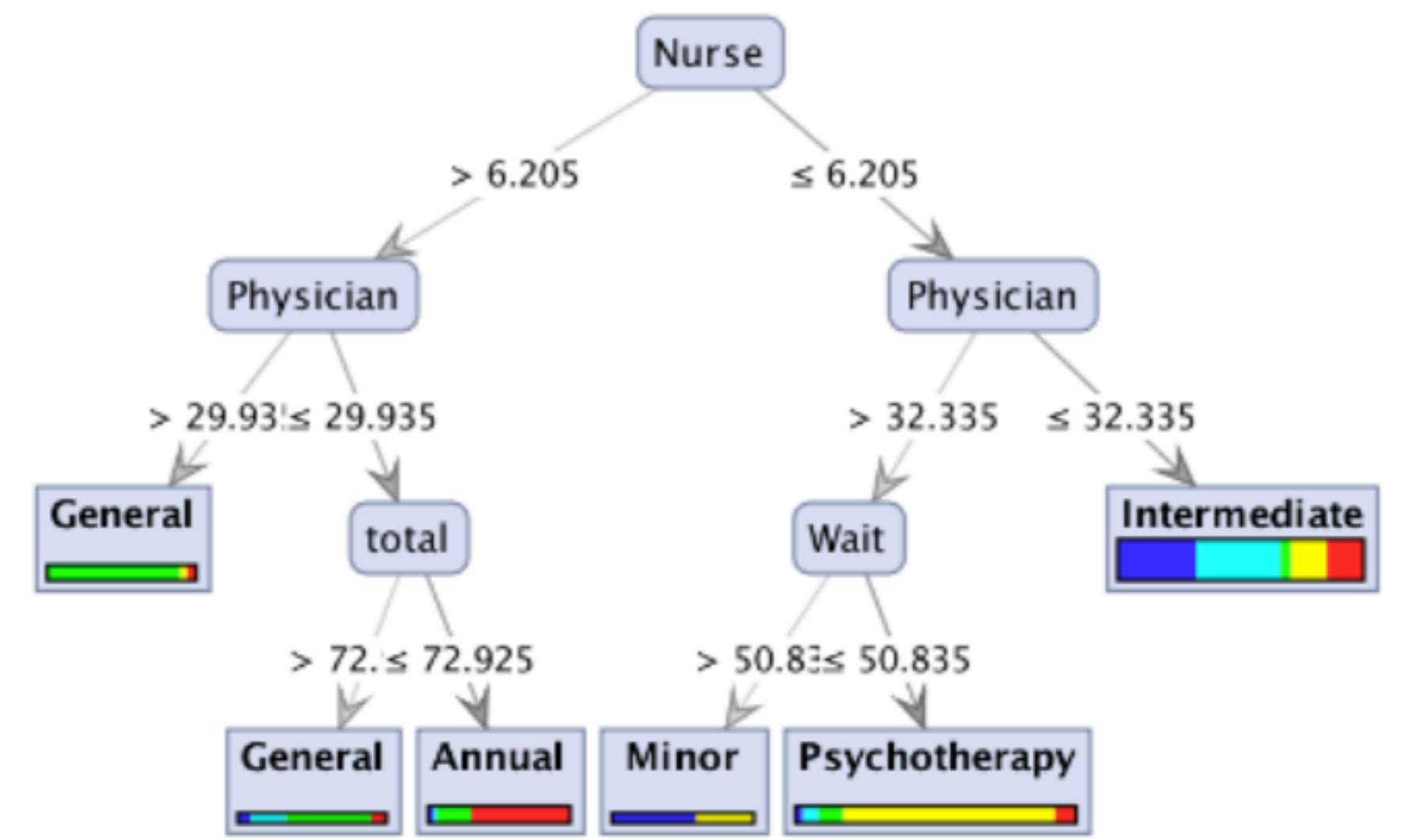


[Can Fam Physician](#). 1988 Jan; 34: 67–70.

Satisfaction and Wait Time of Patients Visiting a Family Practice Clinic

[David Bestvater](#), [Earl V. Dunn](#), [Connie Townsend](#), and [Wendy Nelson](#)

Services provided by average consultation, examination, and wait times						
Service	Avg. Total Time (min)	Avg. Total Wait (min)	Avg. Time with Nurse (min)	Avg. Time with Physician (min)	No. Cases (n)	Incomplete Cases (n)
Minor assessment (std.)	50 (30)	33 (22)	1 (3)	16 (13)	67	11
Intermediate assessment (std.)	55 (24)	37 (21)	2 (3)	16 (12)	400	29
General assessment (std.)	77 (27)	31 (17)	10 (5)	36 (19)	43	1
Psychotherapy (std.)	71 (22)	35 (16)	2 (3)	34 (14)	11	0
Annual exam (after 16 th birthday) (std.)	51 (30)	26 (12)	7 (4)	18 (8)	5	2
Other service					13	N/A
No service code given					74	N/A





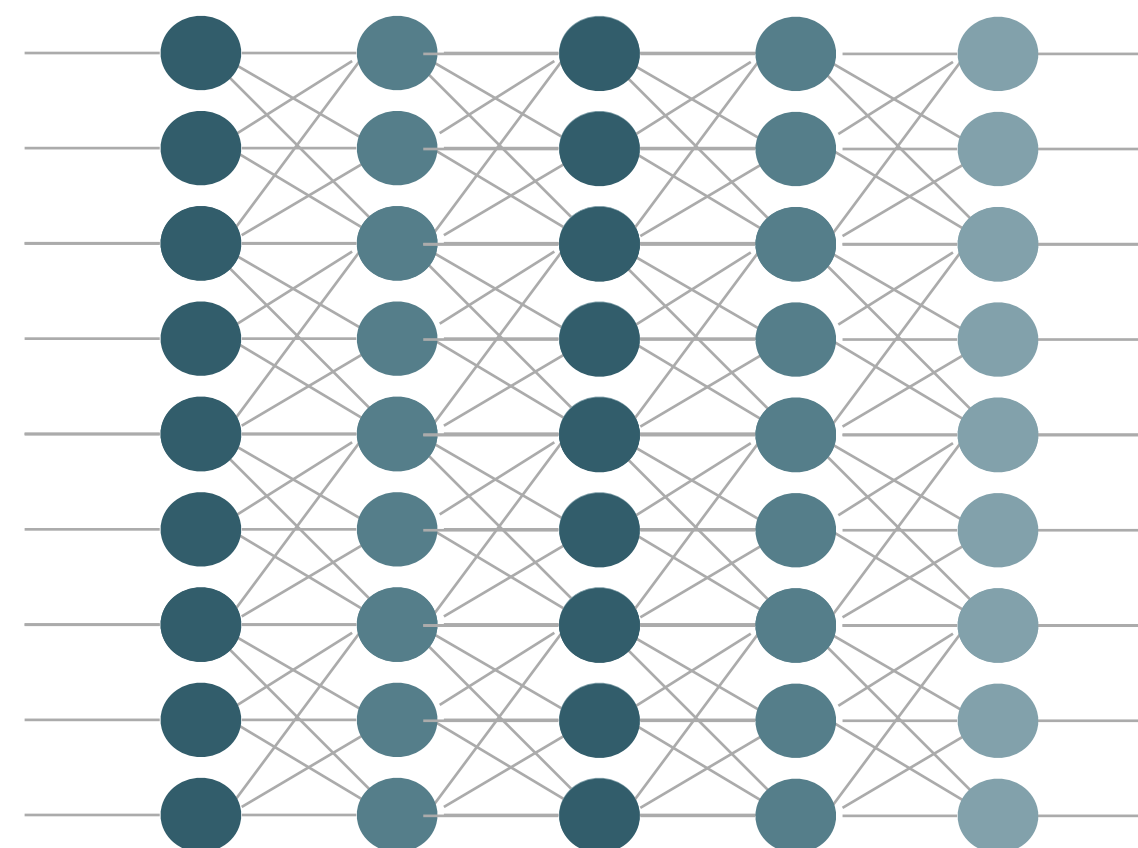
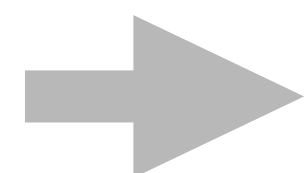
Non-interference

Inputs differing only in secret values produce equivalent observations.

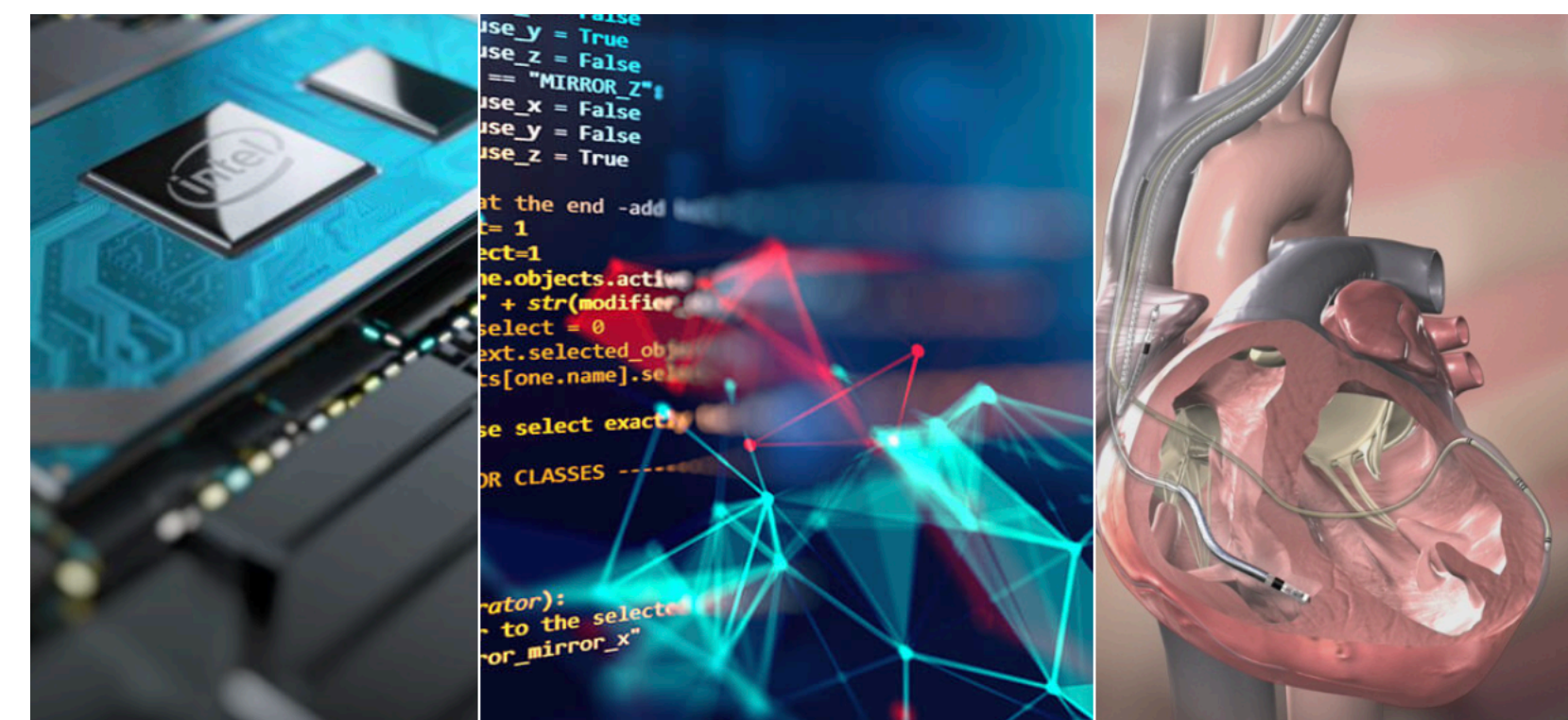
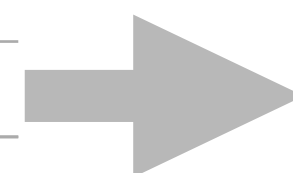
$\forall s_1, s_2$ and $\forall p$ we have that $\mathcal{O}(s_1, p) = \mathcal{O}(s_2, p)$



Historical Data



ML models



Data-driven hardware, software, or CPS

ML trained over historical data may **encode and amplify** discrimination and biases present in the society.

Do we need to worry about such hypothetical scenarios?





State of Wisconsin Department of Corrections

COMPAS

DOC uses the Correctional Offender Management Profiling for Alternative Sanctions tool, commonly known as COMPAS, for criminogenic assessments and unified case planning. This actuarial risk assessment system contains offender information specifically designed to detect and inform dynamic case plans that will guide the offender throughout his or her lifecycle in the criminal justice system.





Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Two Petty Theft Arrests

VERNON PRATER	BRISHA BORDEN
Prior Offenses 2 armed robberies, 1 attempted armed robbery	Prior Offenses 4 juvenile misdemeanors
Subsequent Offenses 1 grand theft	Subsequent Offenses None
LOW RISK 3	HIGH RISK 8

Two Drug Possession Arrests

DYLAN FUGETT	BERNARD PARKER
Prior Offense 1 attempted burglary	Prior Offense 1 resisting arrest without violence
Subsequent Offenses 3 drug possessions	Subsequent Offenses None
LOW RISK 3	HIGH RISK 10

ProPublica's main charge:
Black defendant face higher false positive rates



Individual Fairness

Individuals differing only in protected features, receive equivalent outcomes.

$\forall p_1, p_2$ and $\forall n$ we have that $\mathcal{O}(p_1, n) = \mathcal{O}(p_2, n)$

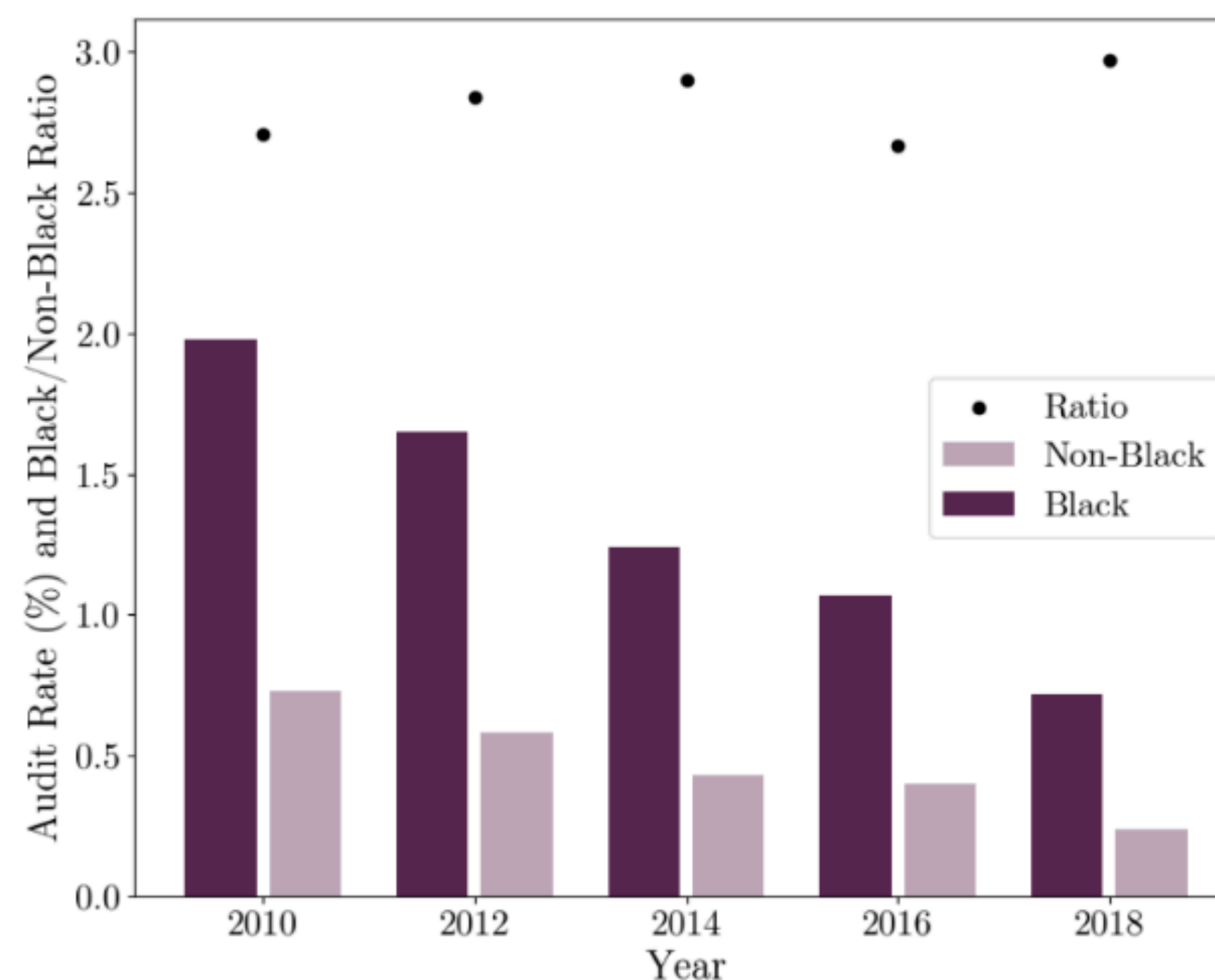
Black Americans Are Much More Likely to Face Tax Audits, Study Finds

A new report documents systemic discrimination in how the I.R.S. selects taxpayers to be audited, with implications for a debate on the agency's funding.

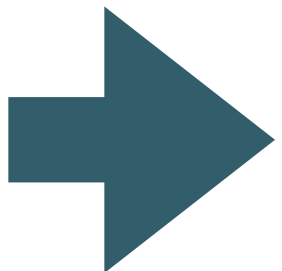
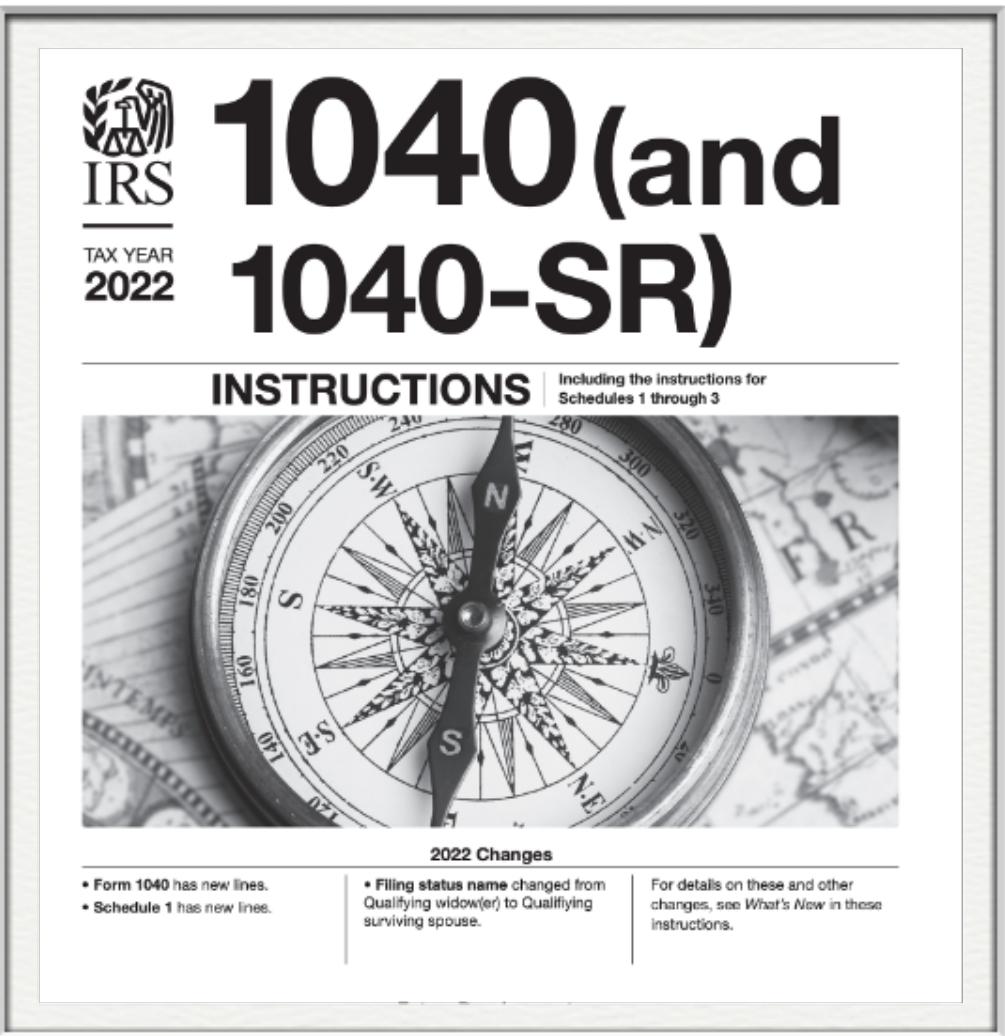
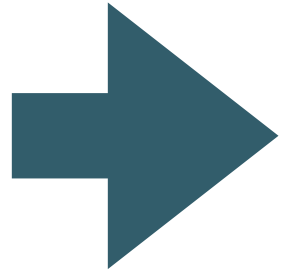
WASHINGTON — Black taxpayers are at least three times as likely to be audited by the Internal Revenue Service as other taxpayers, even after accounting for the differences in the types of returns each group is most likely to file, a team of economists has concluded in one of the most detailed studies yet on race and the nation's tax system.

The findings do not suggest bias from individual tax enforcement agents, who do not know the race of the people they are auditing. They also do not suggest any valid reason for the I.R.S. to target Black Americans at such high rates; there is no evidence that group engages in more tax evasion than others.

Instead, the findings document discrimination in the computer algorithms the agency uses to determine who is selected for an audit, according to the study by economists from Stanford University, the University of Michigan, the University of Chicago and the Treasury Department.



US Tax 101: Manual Filing



Caution: Figure A is an overview of the tests to claim a qualifying child. For details, see the rest of this chapter.

Relationship	A qualifying child is a child who is your . . .
	Son, daughter, stepchild, foster child, or a descendant of any of them (for example, your grandchild)
OR	
	Brother, sister, half brother, half sister, stepbrother, stepsister, or a descendant of any of them (for example, your niece or nephew)
AND	
Age	was . . .
	Under age 19 at the end of 2022 and younger than you (or your spouse, if filing jointly)
OR	
	Under age 24 at the end of 2022, a student, and younger than you (or your spouse, if filing jointly)
OR	
	Permanently and totally disabled at any time during the year, regardless of age

Before you begin:

- ✓ Complete the Earned Income Worksheet, later, in these instructions.
- ✓ **1040 and 1040-SR filers.** Complete line 27; Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.
- ✓ **1040-NR filers.** Complete Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.

CAUTION Use this worksheet only if you meet each of the items discussed under line 3 of Credit Limit Worksheet A, including that you are not filing Form 2555.

Schedule 8812

- Enter the amount from Schedule 8812, line 12 1
- Number of qualifying children under 17 with the required social security number: $\times \$1,500$. Enter the result. 2
TIP: The number of children you use for this line is the same as the number of children you used for line 4 of Schedule 8812.
- Enter your earned income from line 7 of the Earned Income Worksheet. 3
- Is the amount on line 3 more than \$2,500?
☐ No. Leave line 4 blank, enter -0- on line 5, and go to line 6. 4
☐ Yes. Subtract \$2,500 from the amount on line 3. Enter the result.
- Multiply the amount on line 4 by 15% (0.15) and enter the result. 5
- On line 2 of this worksheet, is the amount \$4,500 or more?
☐ No.
If you are a bona fide resident of Puerto Rico and line 5 above is less than line 1 above, go to line 7. Otherwise, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12.
☐ Yes. If line 5 above is equal to or more than line 1 above, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12. Otherwise, go to line 7.
- If your employer withheld or you paid Additional Medicare Tax or Tier 1 RRTA taxes, use the Additional Medicare Tax and RRTA Tax Worksheet to figure the amount to enter; otherwise enter the following amounts.
• Social security tax withheld from Form(s) W-2, box 4, and Puerto Rico Form(s) 499R-2/W-2PR, box 21, and
• Medicare tax withheld from Form(s) W-2, box 6, and Puerto Rico Form(s) 499R-2/W-2PR, box 23. 7
- Enter the total of any amounts from—
• Schedule 1, line 15;
• Schedule 2, line 5;
• Schedule 2, line 6; and
• Schedule 2, line 13. 8
- Add lines 7 and 8. Enter the total. 9

If married filing jointly, include your spouse's amounts with yours when completing lines 7 and 8.

US Tax 101: Tax-Preparation Software



\$11.2 billion industry!



Free (open-source) options for low-income

Who is responsible if your tax-preparation software is buggy?



The misuse of tax preparation software, even if unintentional or accidental, is no defense to accuracy-related penalties under [section 6662](#).

— **Langley v. Comm'r, T.C. Memo. 2013-22**



How to detect bugs in a **legal-critical system** such as tax-preparation software?

Absence of **oracle**

Lack of reliable datasets

Computationally infeasible



Common Law

['kɑːmən 'lɔː]

A body of unwritten laws based on legal precedents established by the courts.





Stare decisis (let the decision stand): courts and judges should honor “precedent” —or the decisions, rulings, and opinions from prior cases.

Corollary. Correctness properties for legal software can be defined in terms of relating similar cases!

$\forall c_1, c_2$ we have that $(c_1 \equiv c_2) \implies \mathcal{O}(c_1) \equiv \mathcal{O}(c_2)$

Tax-Preparation Software: A Legal-Critical System



Metamorphic Testing and Debugging of Tax Preparation Software

Saeid Tizpaz-Niari
Computer Science Department
University of Texas at El Paso
saeid@utep.edu

Verya Monjezi
Computer Science Department
University of Texas at El Paso
vmonjezi@miners.utep.edu

Morgan Wagner
Psychology Department
University of Texas at El Paso
mrwagner@miners.utep.edu

Shiva Darian
Information Science Department
University of Colorado Boulder
shiva.darian@colorado.edu

Krystia Reed
Psychology Department
University of Texas at El Paso
kmreed2@utep.edu

Ashutosh Trivedi
Computer Science Department
University of Colorado Boulder
ashutosh.trivedi@colorado.edu

TENFORTY revealed three types of failures in *OpenTaxSolver*: missing some eligibility conditions (e.g., married people filing separately status is not eligible to take earned income credits); software fails to satisfy the correctness requirements when

the computed tax returns get very close to zero (small non-zero values); and the updated software that allows users to explicitly opt for an option does not satisfy some correctness requirements in the corner cases.

Privacy

Inputs differing only in secret values produce equivalent observations.

$\forall s_1, s_2$ and $\forall p$ we have that $\mathcal{O}(s_1, p) = \mathcal{O}(s_2, p)$

Fairness

Individuals differing only in protected features, receive equivalent outcomes.

Hyperproperties

$\forall p_1, p_2$ and $\forall n$ we have that $\mathcal{O}(p_1, n) = \mathcal{O}(p_2, n)$
Sets of trace properties, i.e. sets of sets of executions traces of a system.

Law

Similar legal cases must have similar rulings.

$\forall c_1, c_2$ we have that $(c_1 \equiv c_2) \implies \mathcal{O}(c_1) \equiv \mathcal{O}(c_2)$



Safe and Secure CPS



Fairness Specifications



Legal Specifications



Safe and Secure CPS



Mahathi Anand
IST, Uni Stuttgart



Vishnu Murali
CU Boulder



Majid Zamani
CU Boulder

A functional certificate based approach for hyperproperty verification

Siyuan Liu , Ashutosh Trivedi, Xiang Yin, Majid Zamani:

Secure-by-construction synthesis of cyber-physical systems. Annu. Rev. Control. 53: 30-50 (2022)

Mahathi Anand, Vishnu Murali, Ashutosh Trivedi, Majid Zamani:

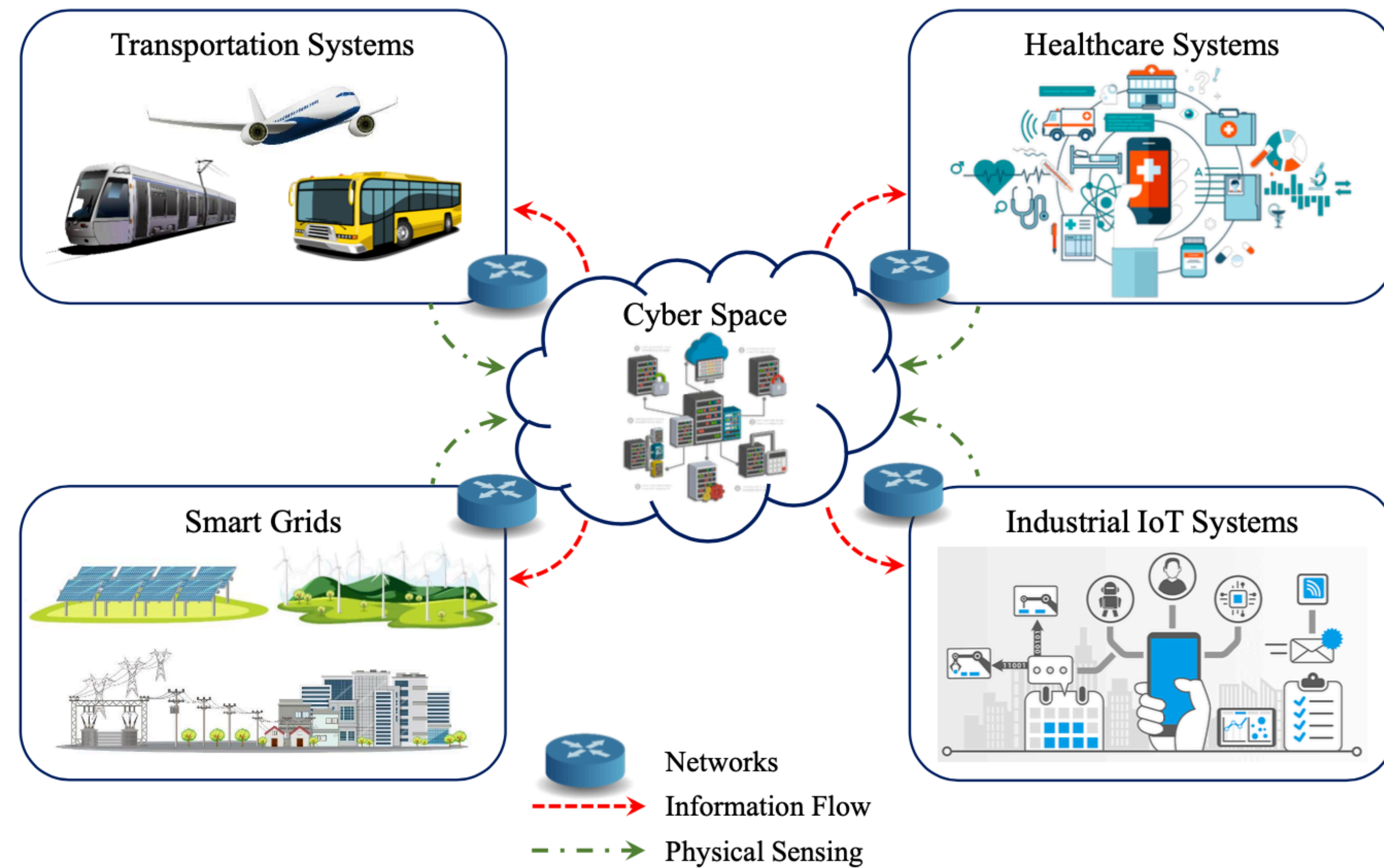
Formal Verification of Control Systems against Hyperproperties via Barrier Certificates. CoRR abs/2105.05493 (2021) IEEE, TAC (2024).

Mahathi Anand, Vishnu Murali, Ashutosh Trivedi, Majid Zamani:

Formal verification of hyperproperties for control systems. CAADCPS@CPSIoTWeek 2021: 29-30

Vishnu Murali , Ashutosh Trivedi , Majid Zamani

Closure Certificates. HSCC 2024: 10:1-10:11





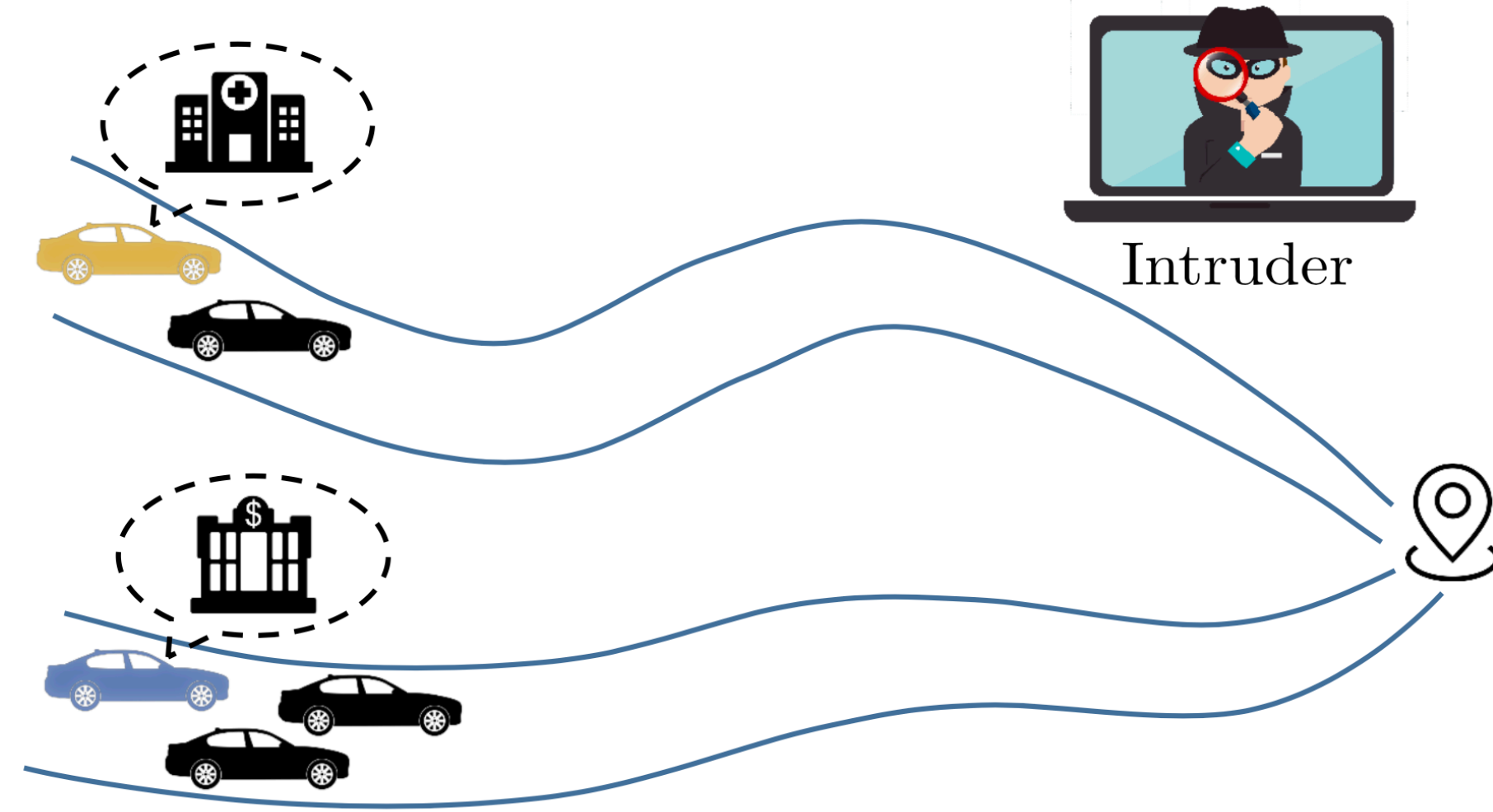
Safety-Critical Cyber-Physical Systems

- Large-scale, interconnected, heterogeneous components
- Complex requirements
- Undecidability
- Powerful deduction and abstraction based approaches

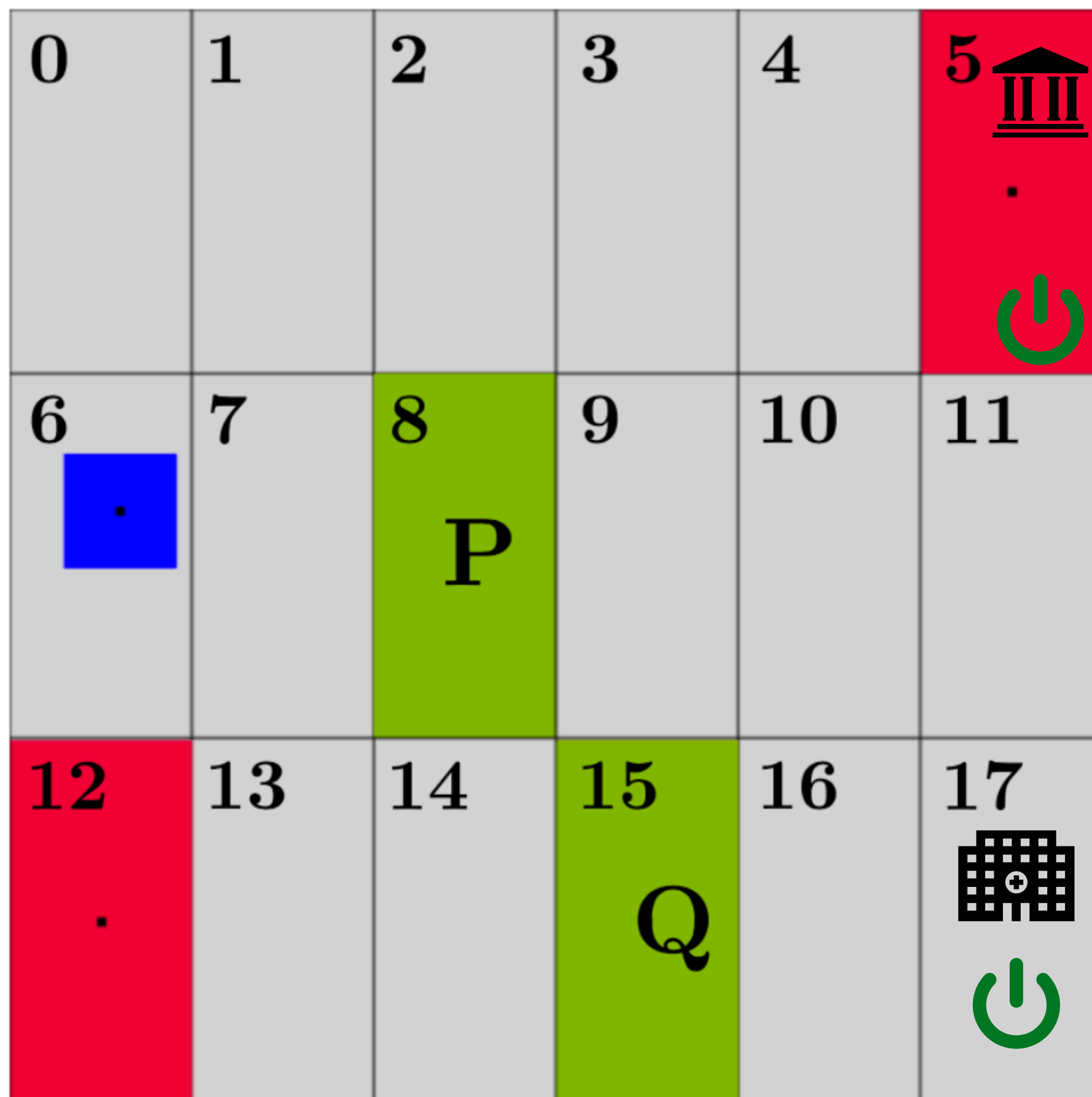


Security-Critical Cyber-Physical Systems

- Novel attack surfaces
- Information leakage in physical variable
- Undecidability
- Rich history of security analysis
(Entropy, opacity, hyperproperties)



Attack Model



$$\Sigma : \begin{cases} \dot{x} = v \cos(\theta) \\ \dot{y} = v \sin(\theta) \\ \dot{\theta} = \frac{v}{L} \tan(u_1) \\ \dot{v} = av + bu_2 \end{cases}$$

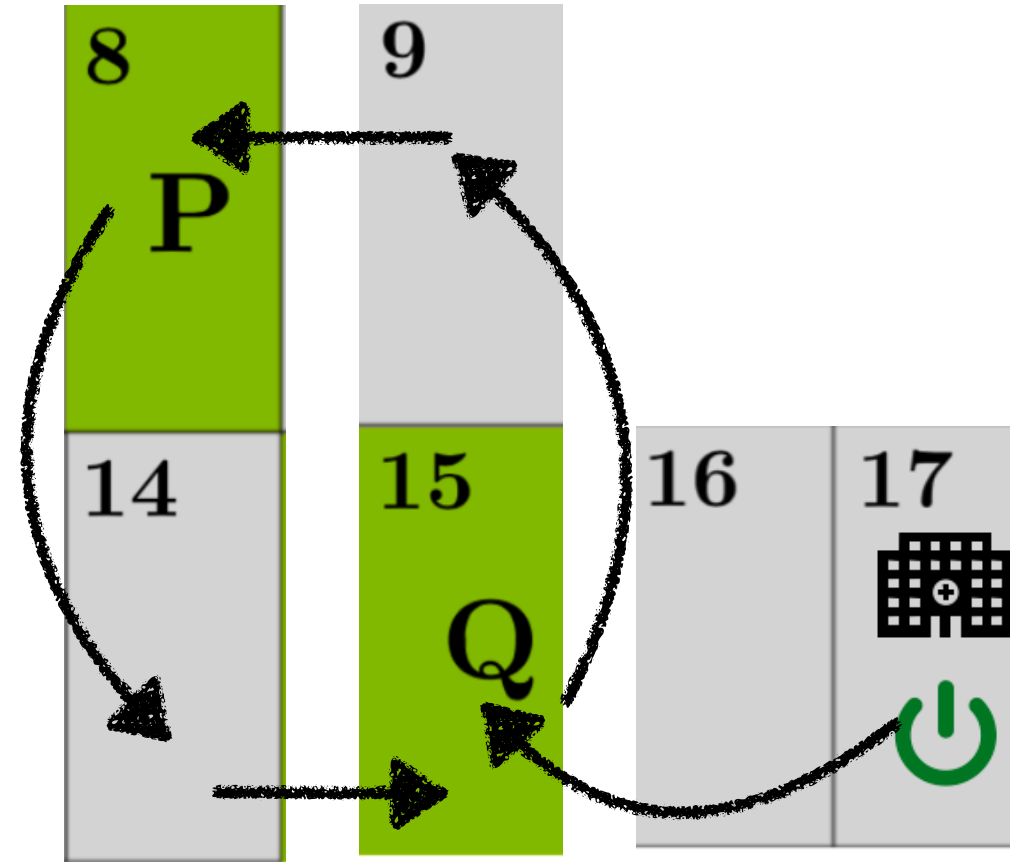
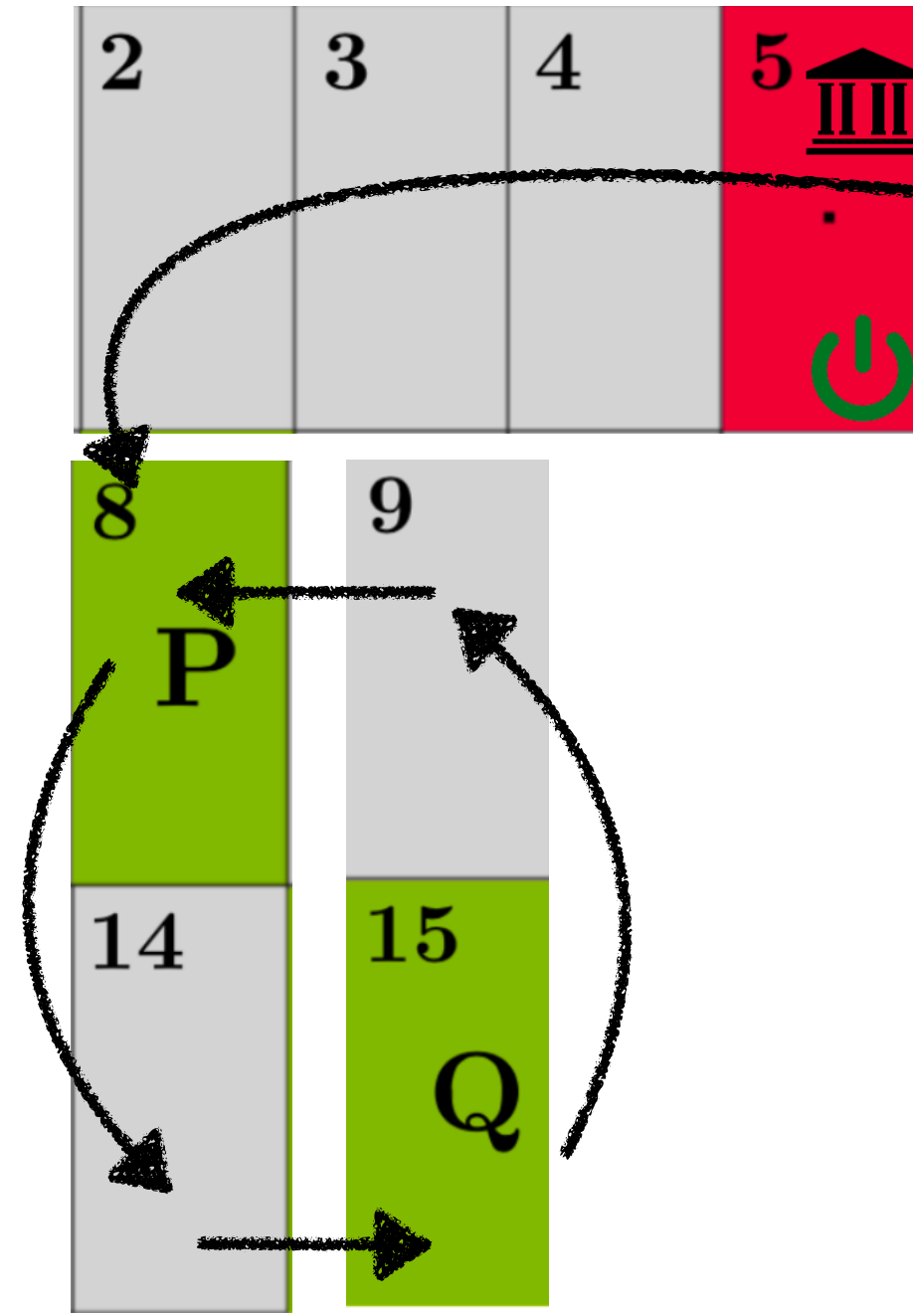
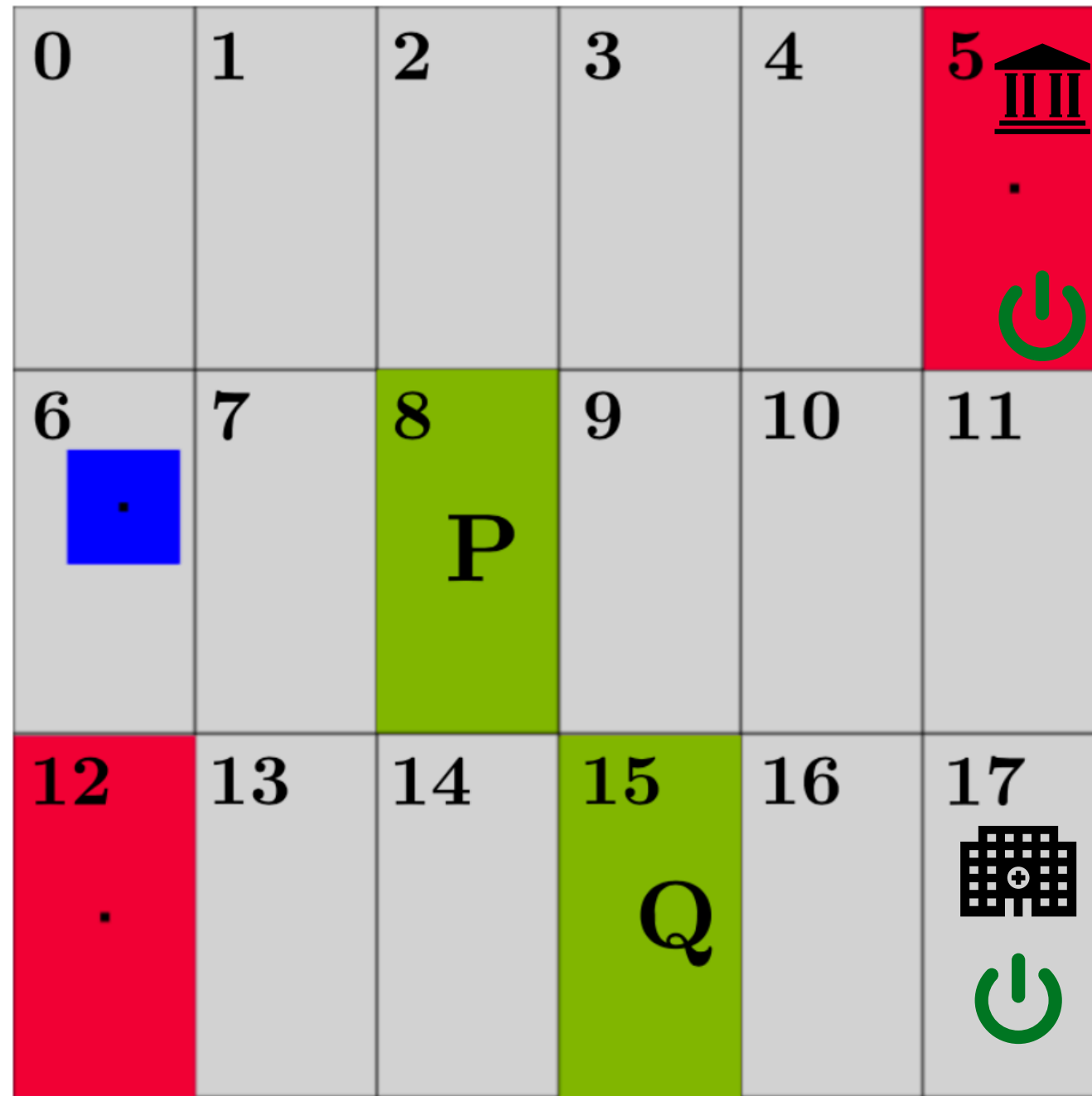


Mission-Objective: The vehicle visits Parking (P) and Quay (Q) locations infinitely often.

$$\forall \pi (\Box \Diamond P_{\pi}) \wedge (\Box \Diamond Q_{\pi})$$

Security-Objective (Opacity): The intruder must not infer whether the vehicle started from a secret location.

$$\forall \pi \exists \pi' ((\text{Init}_{\pi} \wedge \text{Init}_{\pi'} \wedge \text{Secret}_{\pi}) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_{\pi} \iff \mathcal{O}_{\pi'}))$$

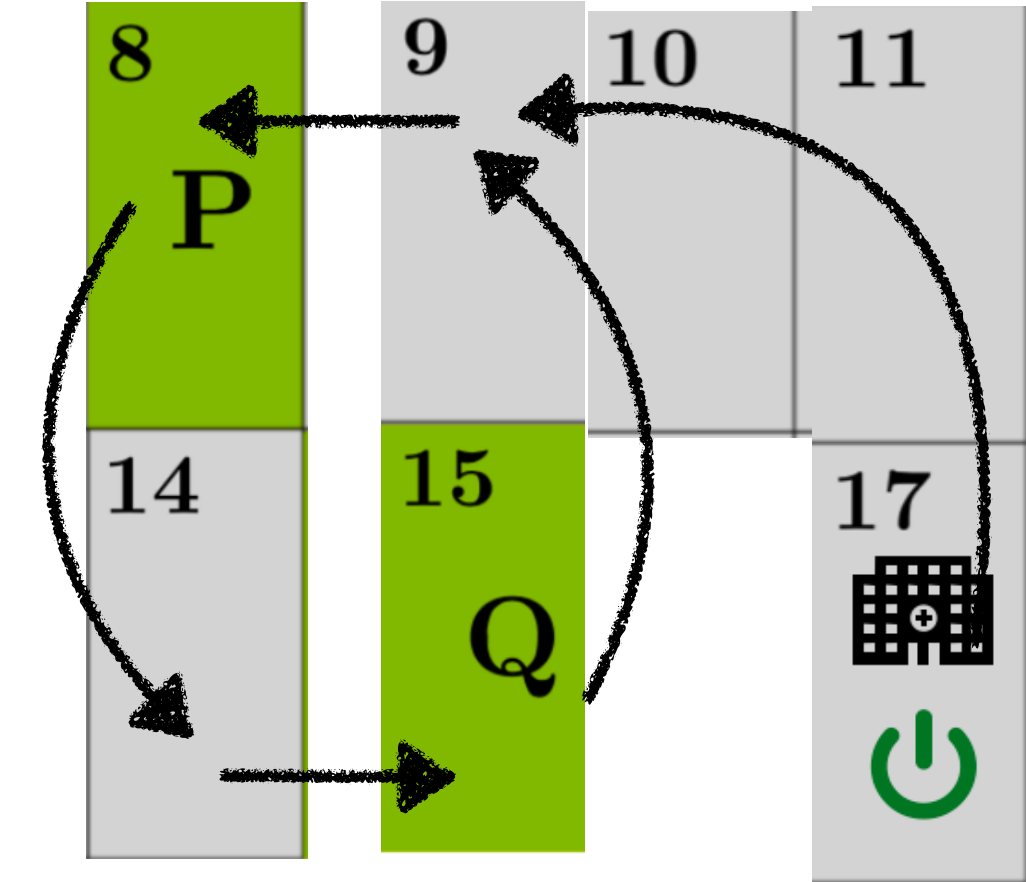
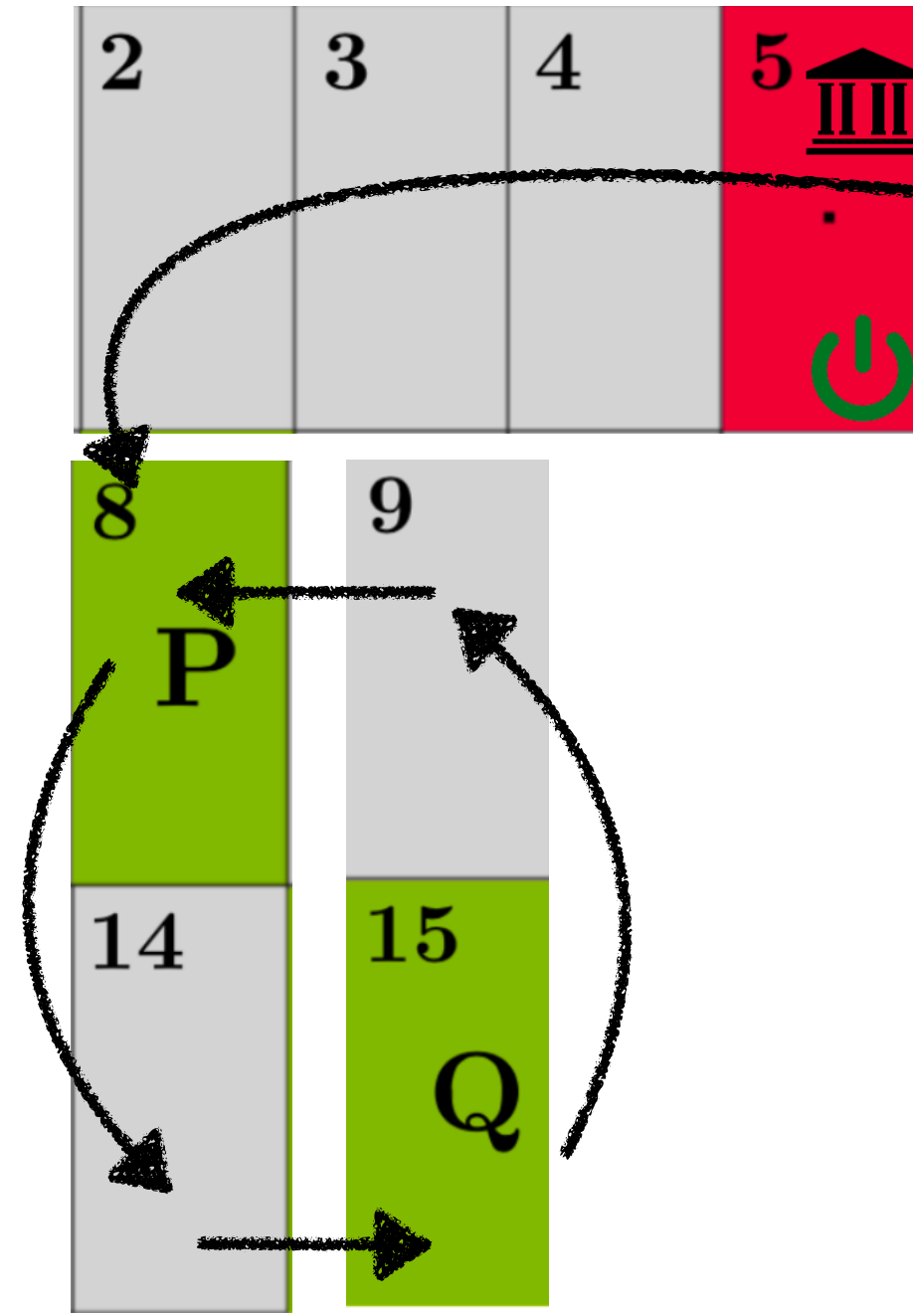
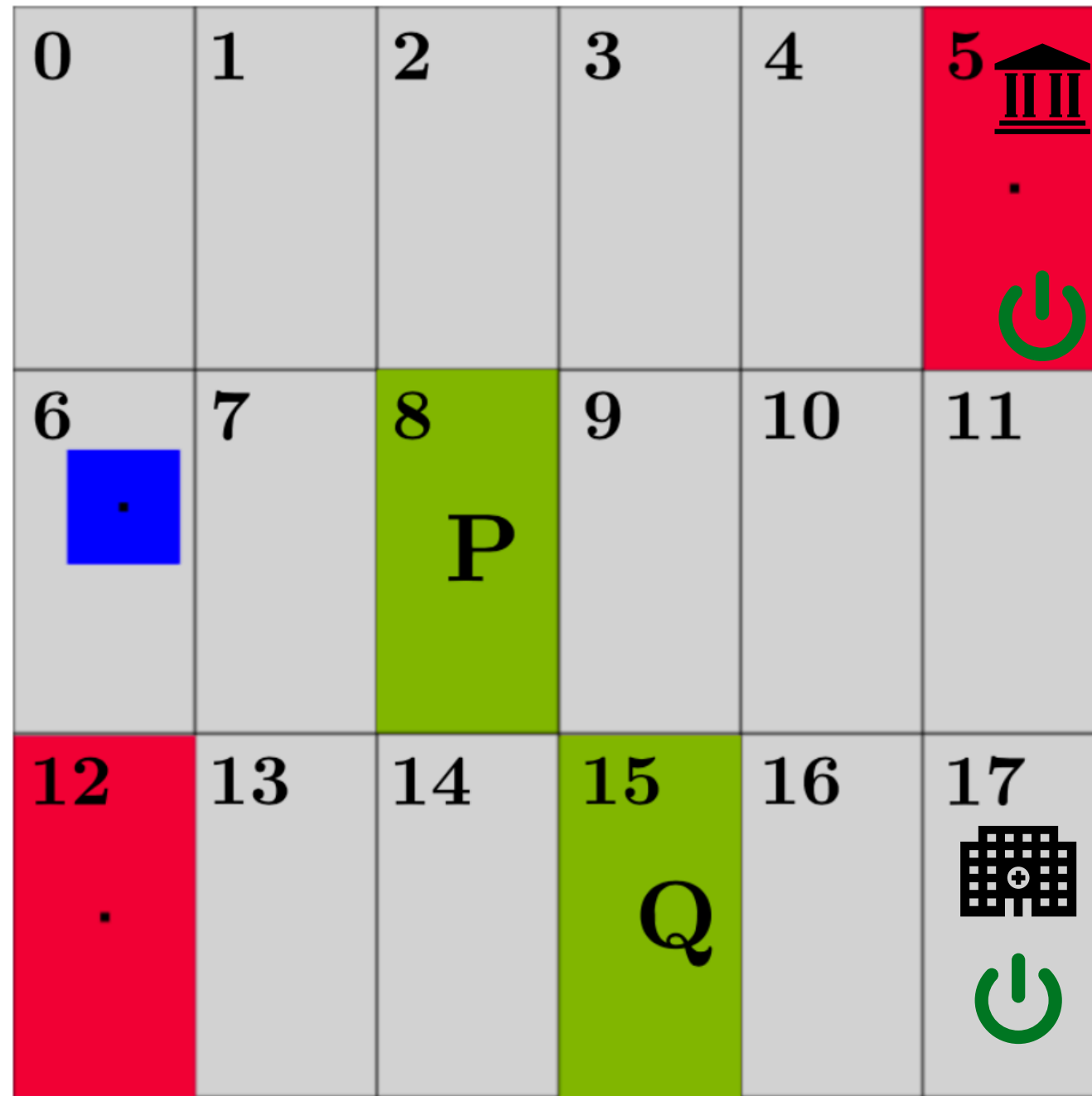


Mission-Objective: The vehicle visits Parking (P) and Quay (Q) locations infinitely often.

$$\forall \pi (\Box \Diamond P_{\pi}) \wedge (\Box \Diamond Q_{\pi})$$

Security-Objective (Opacity): The intruder must not infer whether the vehicle started from a secret location.

$$\forall \pi \exists \pi' ((\text{Init}_{\pi} \wedge \text{Init}_{\pi'} \wedge \text{Secret}_{\pi}) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_{\pi} \iff \mathcal{O}_{\pi'}))$$



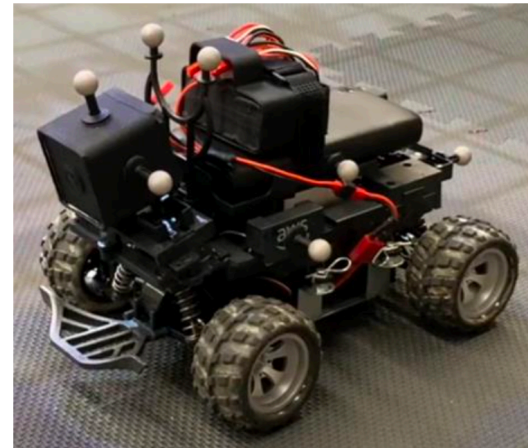
Mission-Objective: The vehicle visits Parking (P) and Quay (Q) locations infinitely often.

$$\forall \pi (\Box \Diamond P_{\pi}) \wedge (\Box \Diamond Q_{\pi})$$

Security-Objective (Opacity): The intruder must not infer whether the vehicle started from a secret location.

$$\forall \pi \exists \pi' ((\text{Init}_{\pi} \wedge \text{Init}_{\pi'} \wedge \text{Secret}_{\pi}) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_{\pi} \iff \mathcal{O}_{\pi'}))$$

$$\Sigma : \begin{cases} \dot{x} = v \cos(\theta) \\ \dot{y} = v \sin(\theta) \\ \dot{\theta} = \frac{v}{L} \tan(u_1) \\ \dot{v} = av + bu_2 \end{cases}$$



0	1	2	3	4	5 
6 	7	8 	9	10	11
12 	13	14	15 	16	17 

$$\forall \pi (\Box \Diamond P_\pi) \wedge (\Box \Diamond Q_\pi)$$

$$\forall \pi \exists \pi' ((\text{Init}_\pi \wedge \text{Init}_{\pi'} \wedge \text{Secret}_\pi) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_\pi \iff \mathcal{O}_{\pi'}))$$

Problem: HyperLTL Verification

Given a discrete-time dynamical system S and a hyperLTL φ specification, decide whether the system satisfies the specification, i.e. $S \models \varphi$.

Undecidable

Classic abstractions don't preserve opacity

Opacity preserving abstractions

Deductive verification (functional certificates)

Rajeev Alur, Pavol Cerný, Steve Zdancewic:
Preserving Secrecy Under Refinement. ICALP (2) 2006: 107-118

Xiang Yin , Majid Zamani , Siyuan Liu 
On Approximate Opacity of Cyber-Physical Systems. IEEE Trans. Autom. Control. 66(4): 1630-1645 (2021)

Verifying CPS using functional certificates

Barrier certificates (inductive state-invariants)

Lyapunov certificates (ranking functions)

Closure certificates (inductive transition invariants)

Stephen Prajna, Ali Jadbabaie:

Safety Verification of Hybrid Systems Using Barrier Certificates. HSCC 2004: 477-492

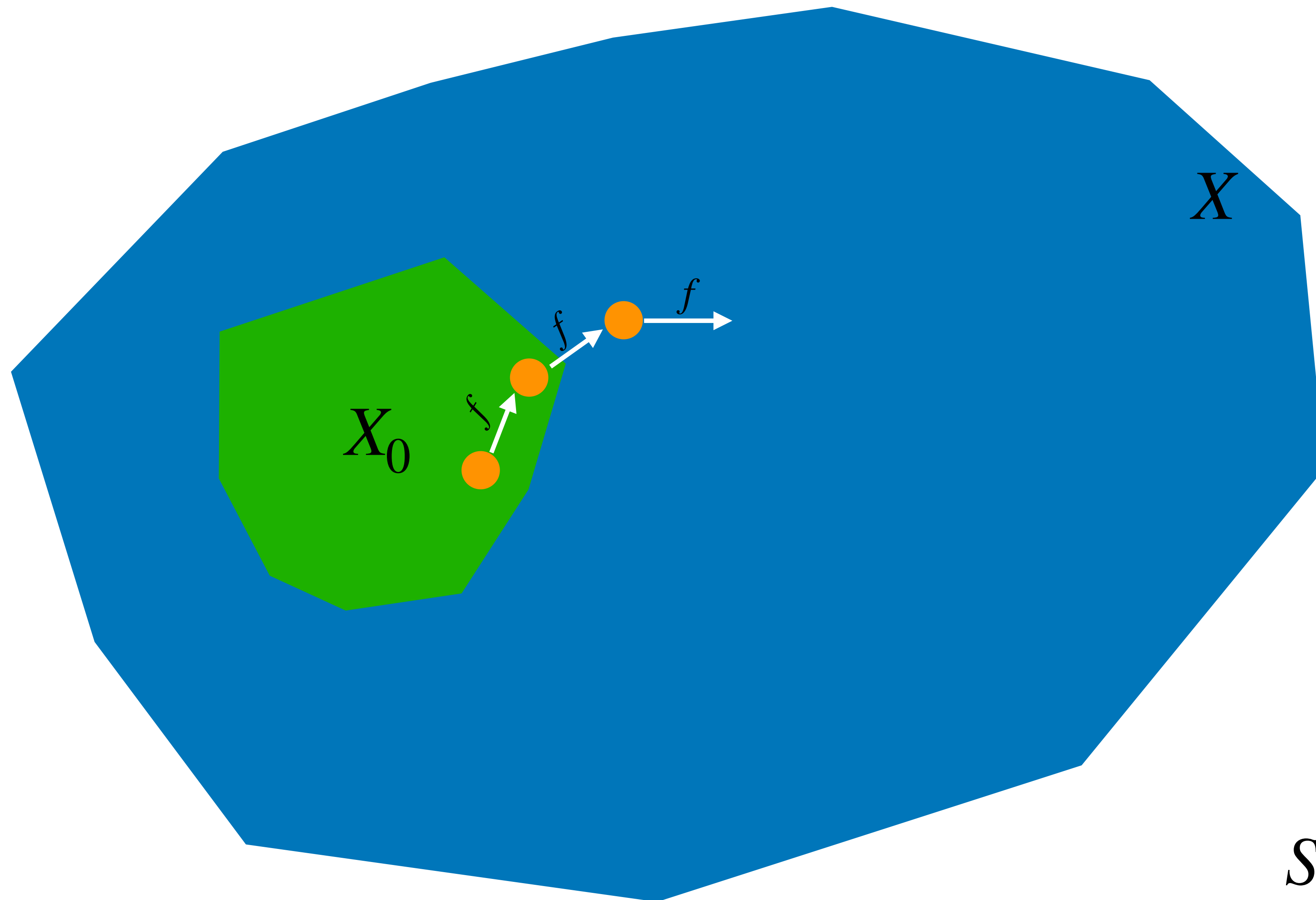
Antonis Papachristodoulou, Stephen Prajna:

On the construction of Lyapunov functions using the sum of squares decomposition. CDC 2002: 3482-3487

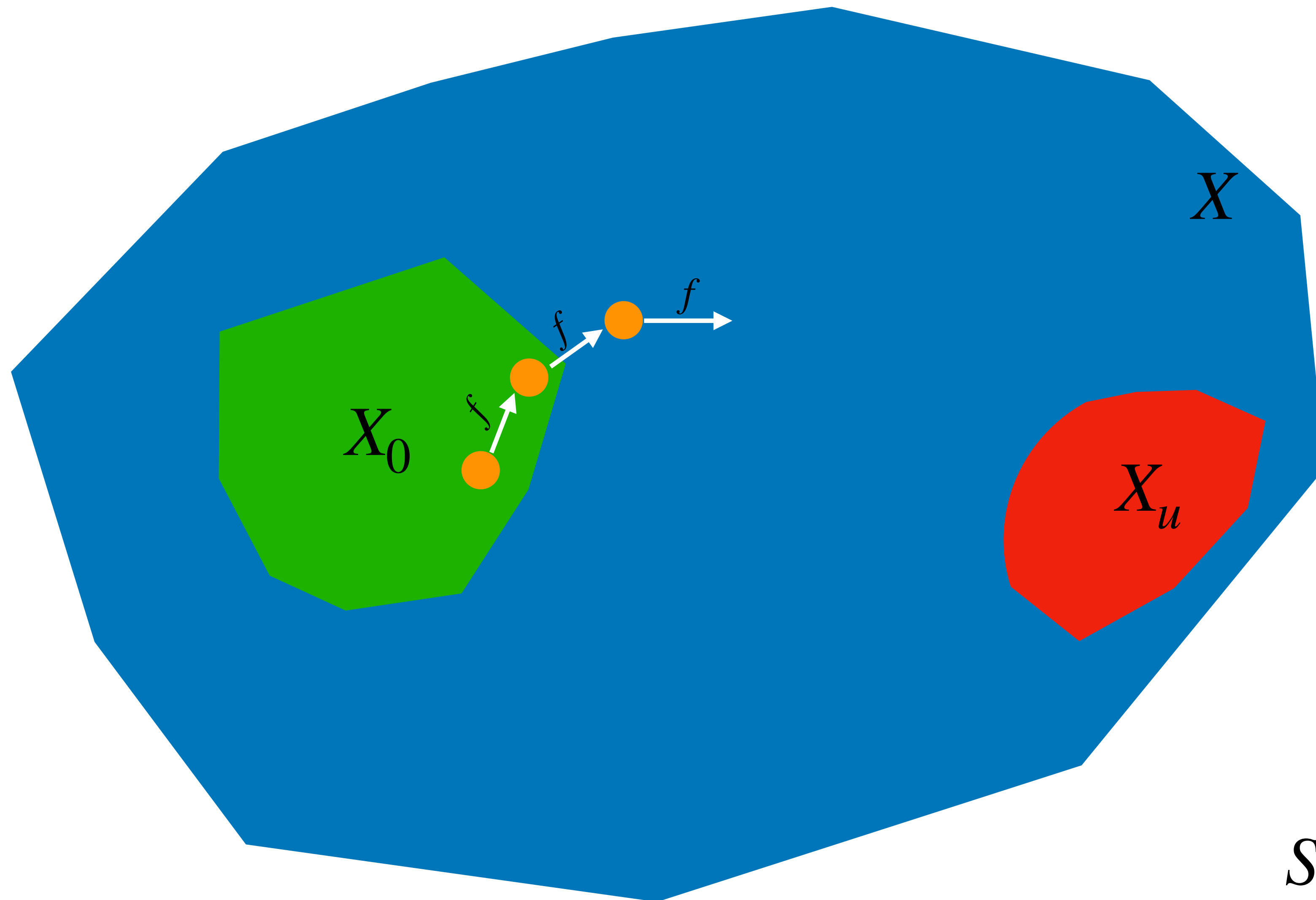
Vishnu Murali , Ashutosh Trivedi , Majid Zamani :

Closure Certificates. HSCC 2024: 10:1-10:11

Verifying Safety using Barrier Certificates



$$S = (X, X_0, f)$$



Unsafe set: X_u

$$S = (X, X_0, f)$$

Barrier Certificates

The function $B : X \rightarrow \mathbb{R}$ is a **barrier certificate** for (X, X_0, f, X_u) , if

► B is **non-positive** in the initial states, i.e.,

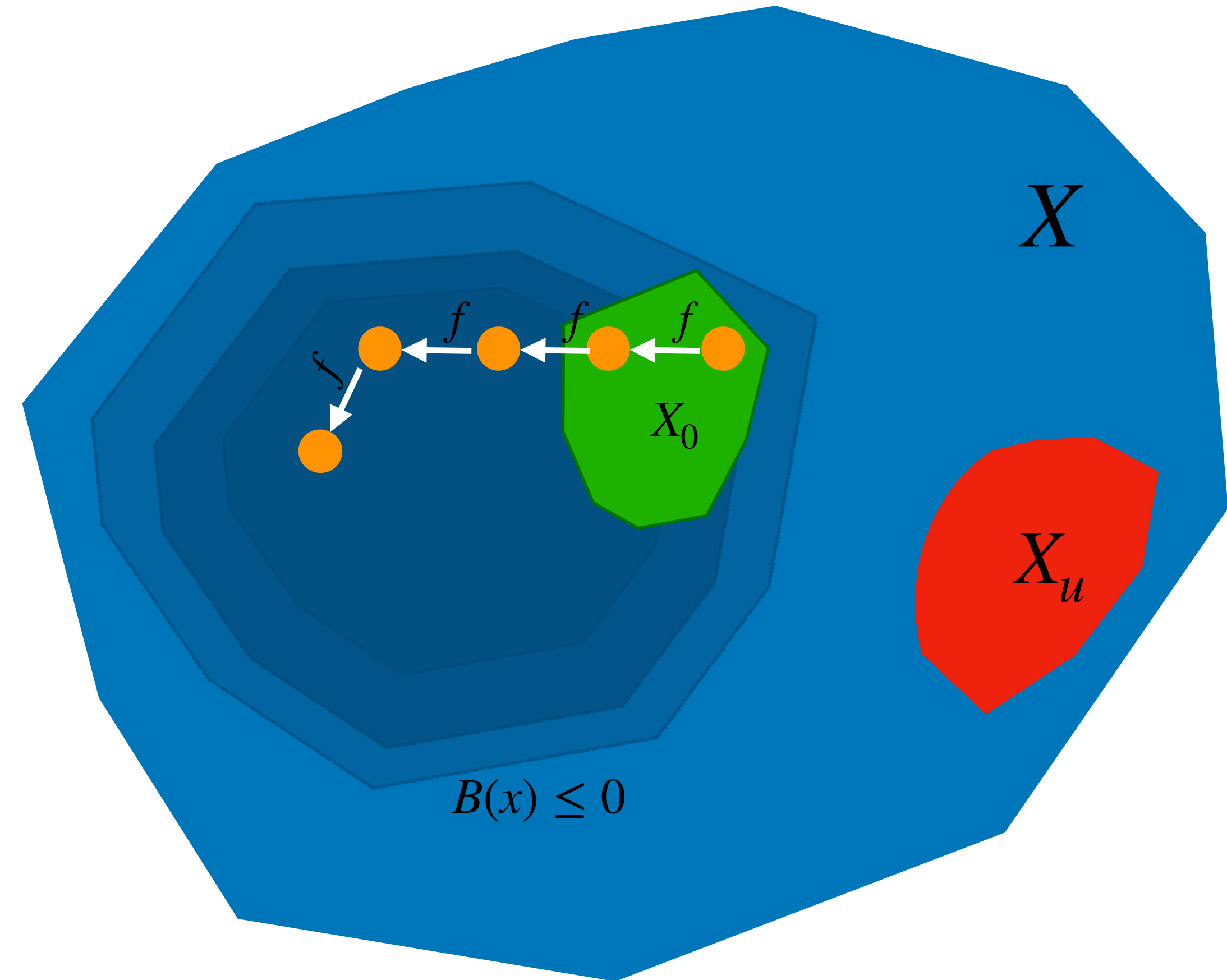
$$B(x) \leq 0, \text{ for } x \in X_0,$$

► B is **positive** in the unsafe states, i.e.,

$$B(x) > 0, \text{ for } x \in X_u, \text{ and}$$

► B **decreases** with transitions in every safe state, i.e.,

$$B(f(x)) \leq B(x) \text{ for all } x \in X \setminus X_u.$$



“Level-0 set of barrier functions characterize inductive state-invariants.”

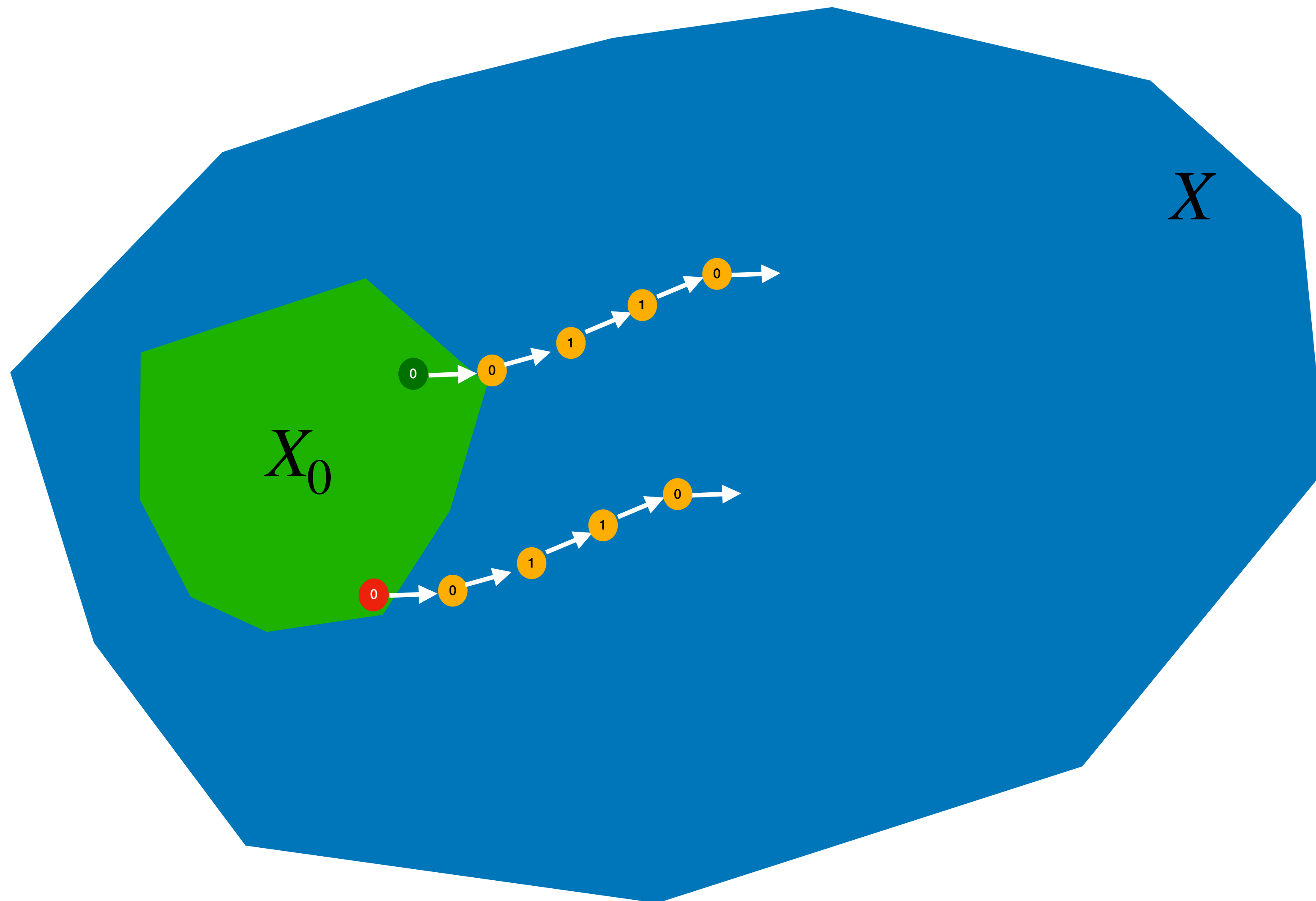
Theorem: Barrier Certificates and Safety.

1. The existence of a barrier certificate implies safety.
2. A number of methods exist for discovering *barrier certificates* including **sum-of-square** (SOS) relaxation, **optimization**, and counterexample-guided inductive synthesis, and machine learning.

$$\forall \pi \exists \pi' ((\text{Init}_\pi \wedge \text{Init}_{\pi'} \wedge \text{Secret}_\pi) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_\pi \iff \mathcal{O}_{\pi'}))$$

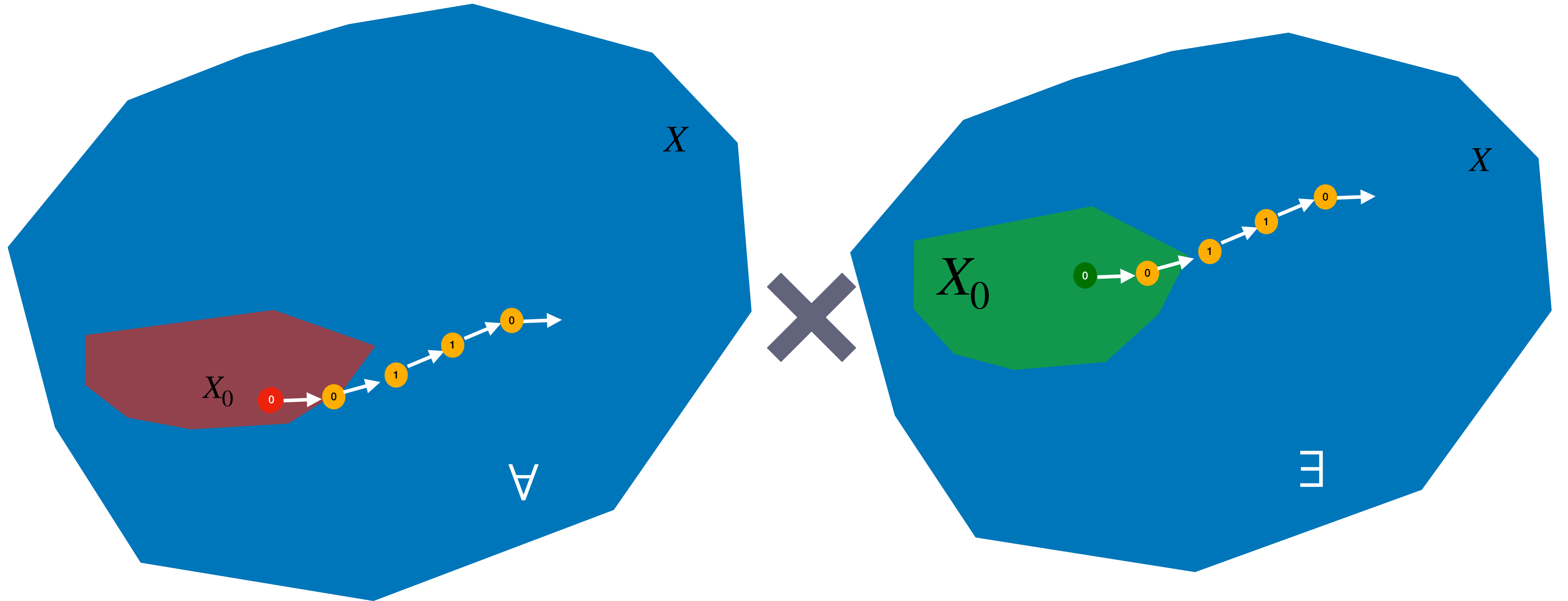
Verifying Hypersafety using Barrier Certificates

$$\forall \pi \exists \pi' ((\text{Init}_\pi \wedge \text{Init}_{\pi'} \wedge \text{Secret}_\pi) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_\pi \iff \mathcal{O}_{\pi'}))$$



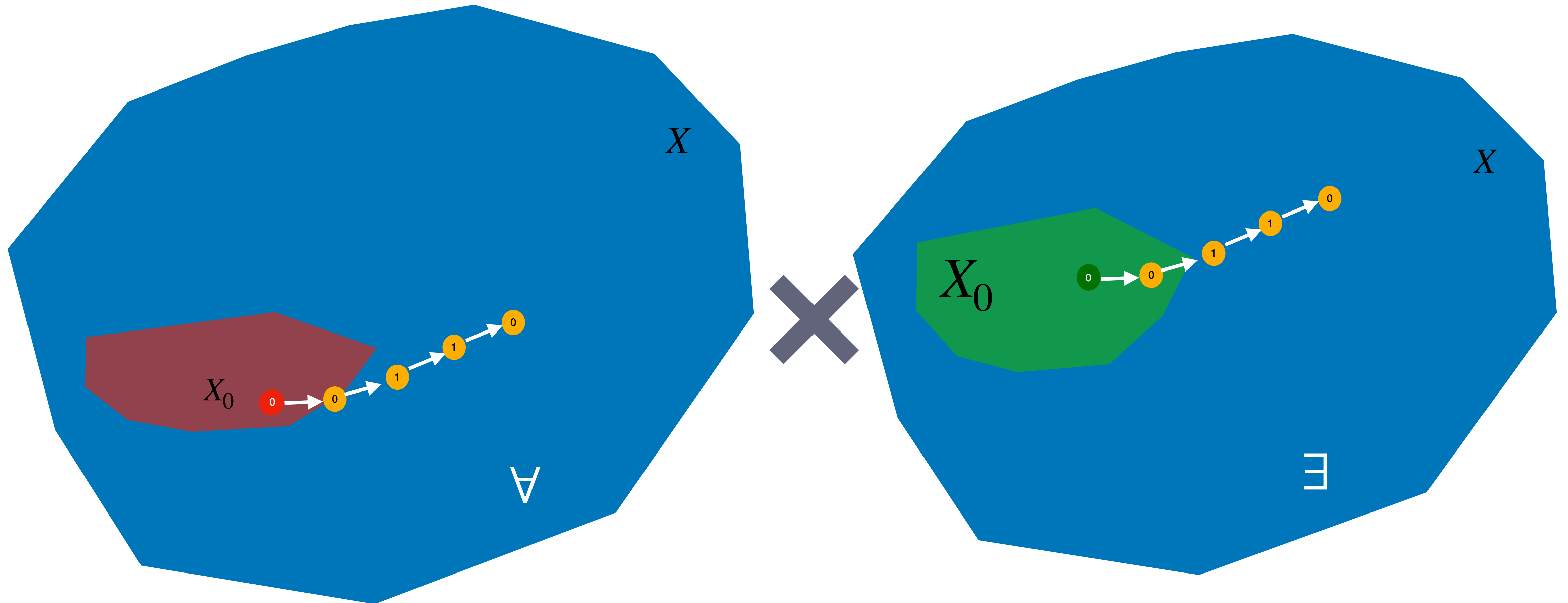
- Non-secret initial state
- Secret initial state
- 0 Observation 0
- 1 Observation 1

$$\forall \pi \exists \pi' ((\text{Init}_\pi \wedge \text{Init}_{\pi'} \wedge \text{Secret}_\pi) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_\pi \iff \mathcal{O}_{\pi'}))$$



HyperLTL semantics

$$\forall \pi \exists \pi' ((\text{Init}_\pi \wedge \text{Init}_{\pi'} \wedge \text{Secret}_\pi) \implies \neg \text{Secret}_{\pi'} \wedge G(\mathcal{O}_\pi \iff \mathcal{O}_{\pi'}))$$



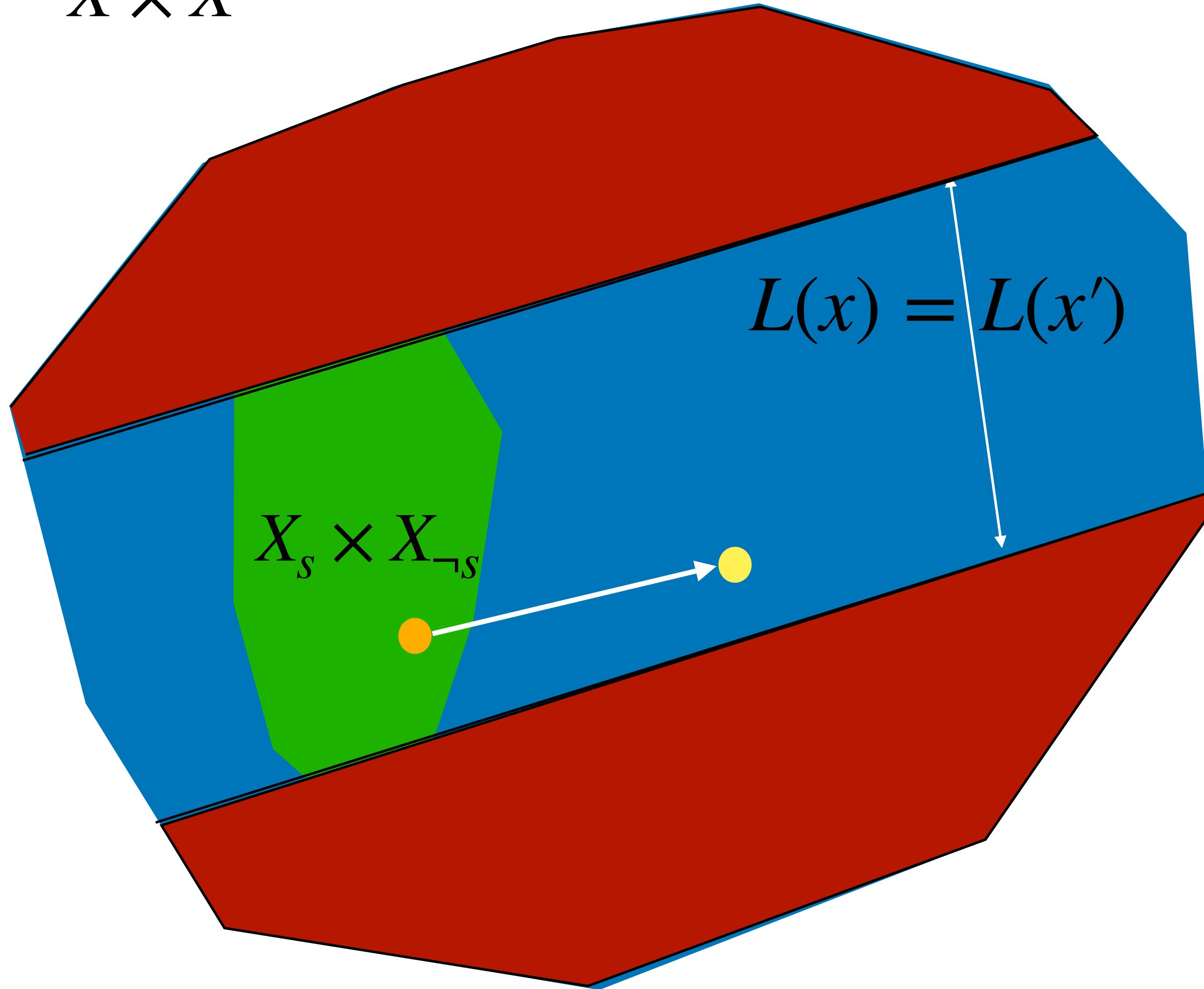
Winning strategy for $\exists$ -player under turn-based semantics implies one under standard semantics.

Turn-based semantics

Augmented barrier certificates (ABCs)

Idea. On the self-composition, for every state-pair with matching observations (**safe** state-pairs), $\exists$ -player can match $\forall$ -player's move to keep the system in such safe states.

$X \times X$



$$B : X \times X \rightarrow \mathbb{R}$$

For all $x \in X_s, y \in X_{\neg s}$,

where $L(x) = L(y)$,

$$B(x, y) \leq 0,$$

For all $x, y \in X$,

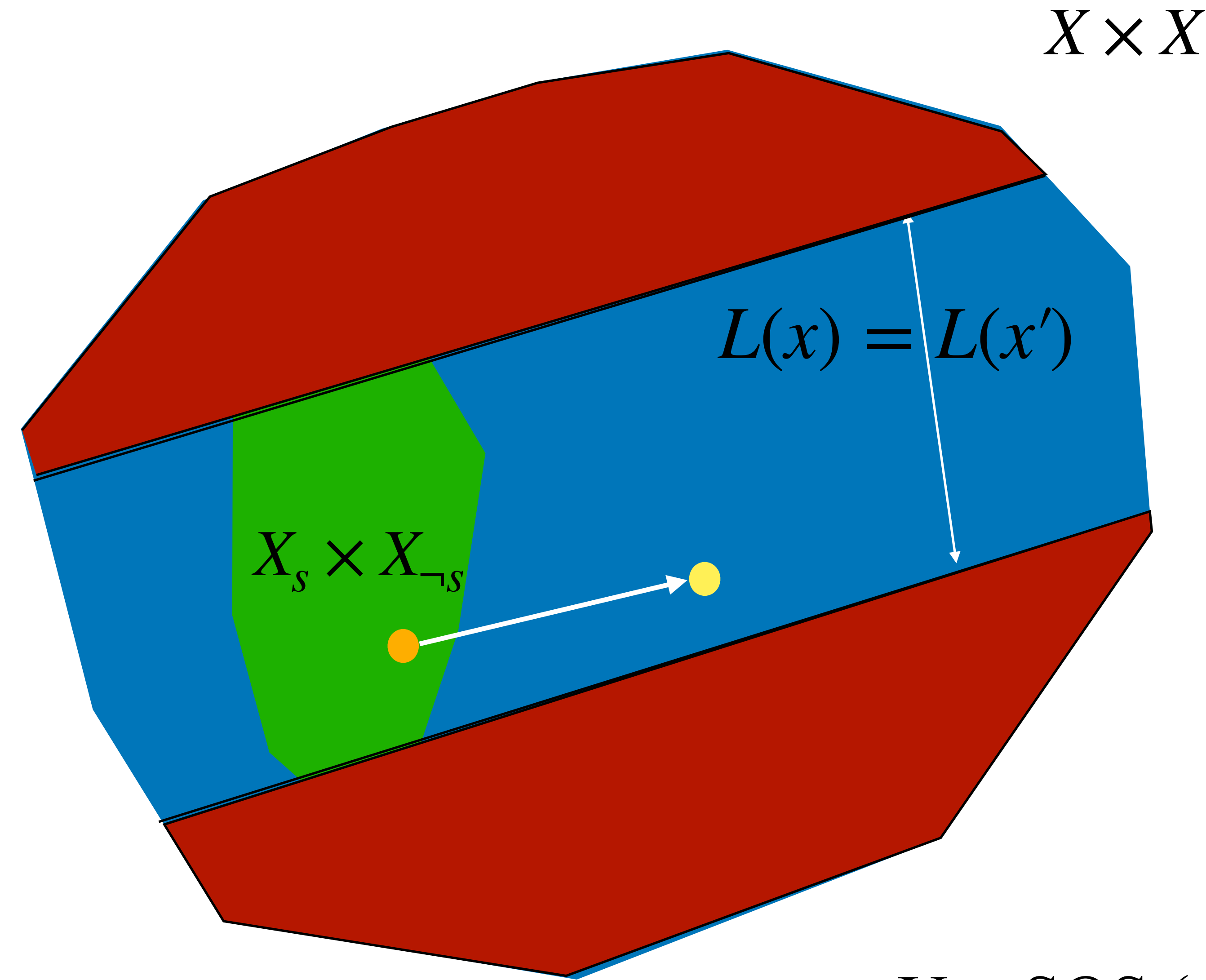
where $L(x) \neq L(y)$,

$$B(x, y) > 0,$$

For all $x, y \in X, x' \in f(x)$, there exists $y' \in f(y)$

$$(B(x, y) \leq 0) \implies (B(x', y') \leq 0).$$

Augmented Barrier Certificates



$$B : X \times X \rightarrow \mathbb{R}$$

For all $x \in X_s, y \in X_{\neg s},$

where $L(x) = L(y),$

$$B(x, y) \leq 0,$$

For all $x, y \in X,$

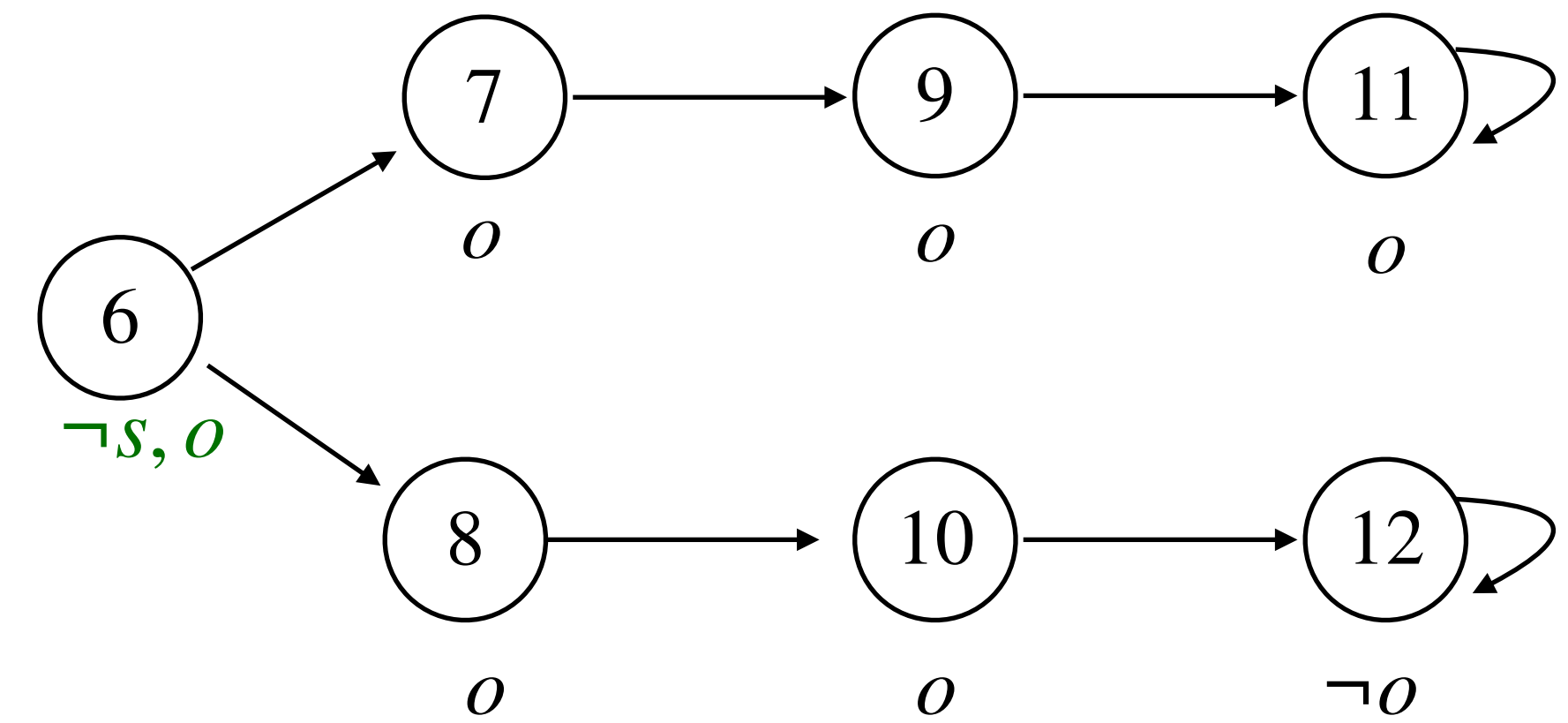
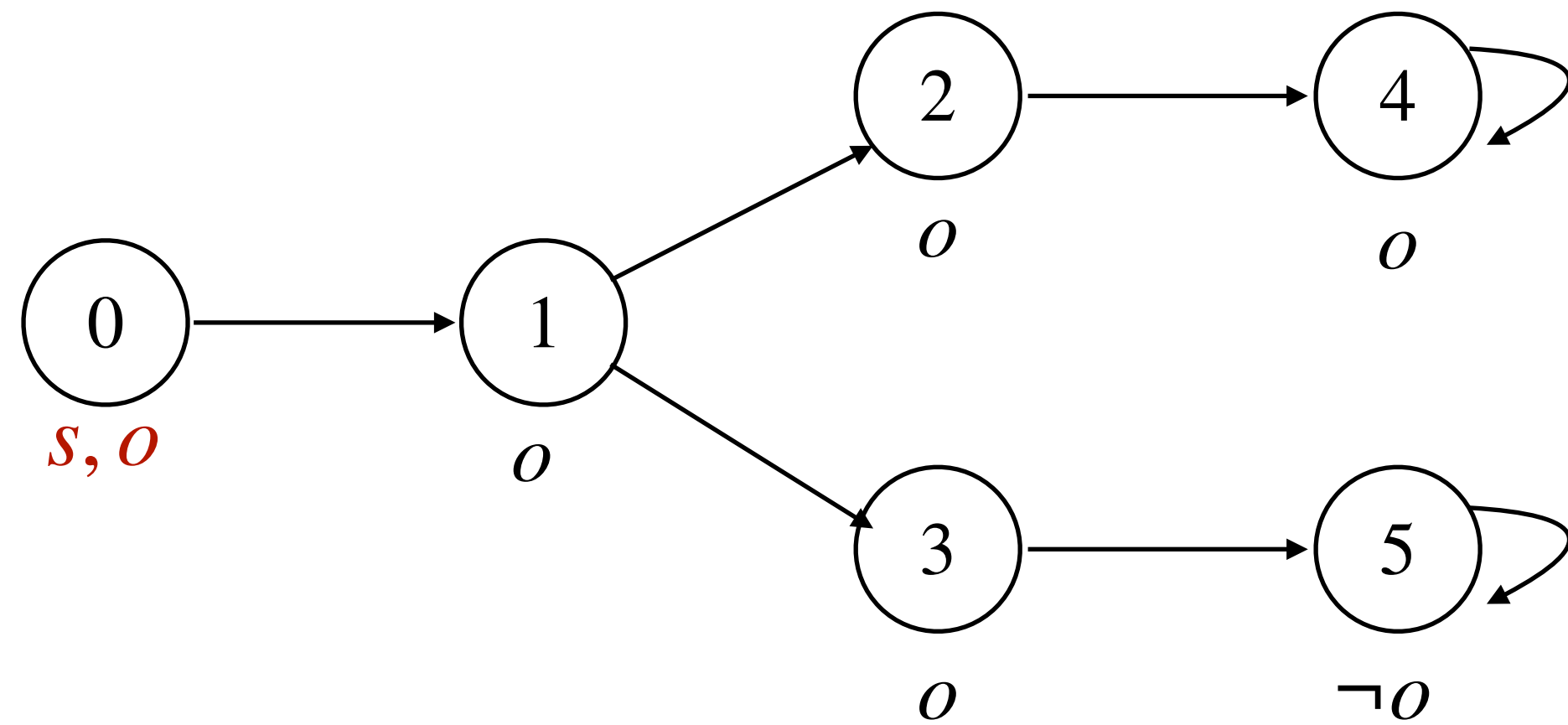
where $L(x) \neq L(y),$

$$B(x, y) > 0,$$

For all $x, x \in X, x' \in f(x),$ there exists $y' \in f(y)$

$$B(x', y') \leq B(x, y).$$

Use SOS (conservative) or CEGIS (expensive)



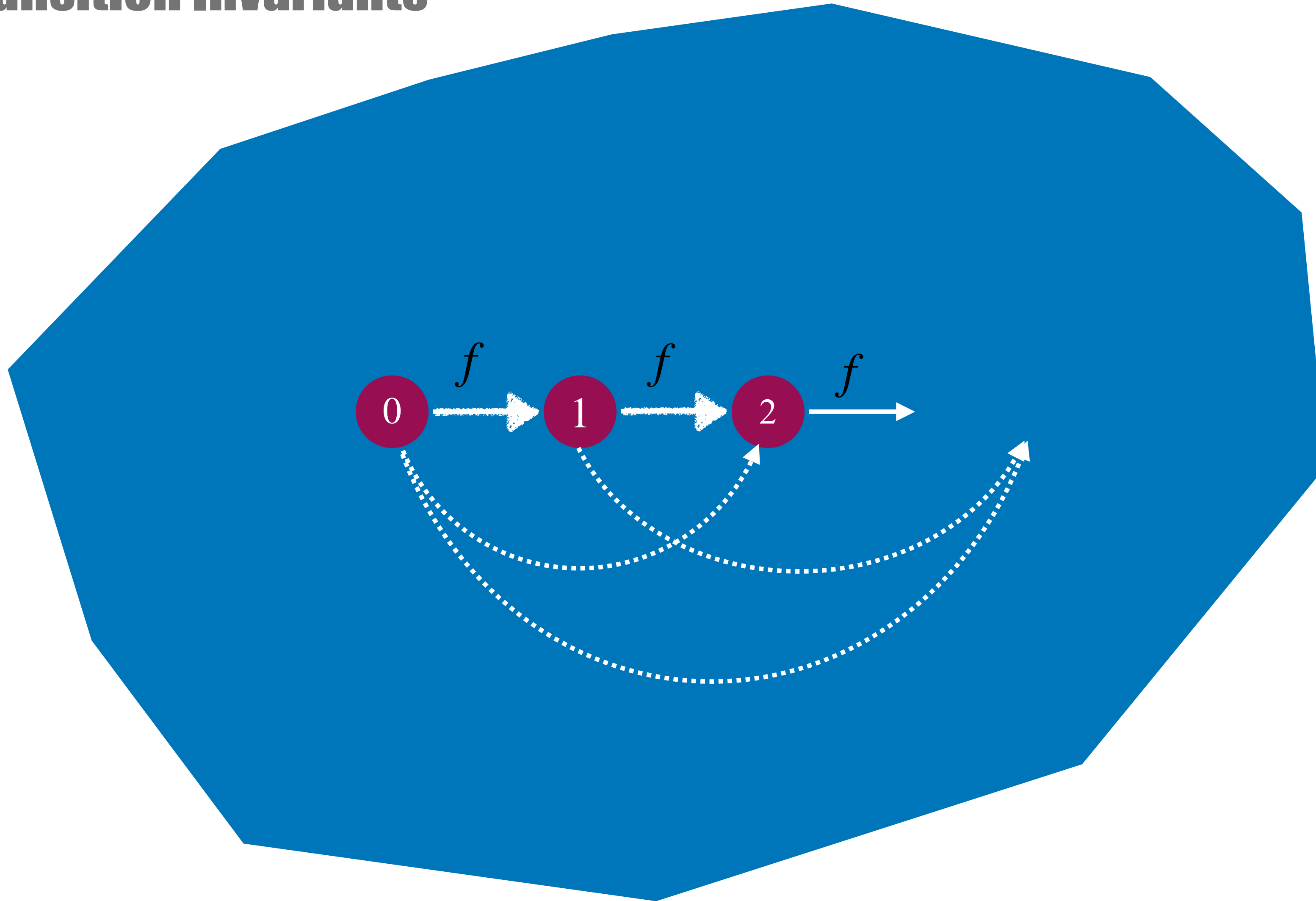
Augmented barrier certificates approach is not complete.

$\exists$ -player may require an arbitrary lookahead to choose his actions.
Can we capture this long term connectivity using function certificates?

Idea. Transition-closure and Closure Certificates

Inductive Transition Invariants and Closure Certificates

Inductive Transition Invariants



For any $x \in X, (x, f(x)) \in T_{inv}$

If $(f(x), y) \in T_{inv}$, then $(x, y) \in T_{inv}$

Closure Certificates: Finite Visit

$$T : X \times X \rightarrow \mathbb{R}$$

For all $x, y \in X$

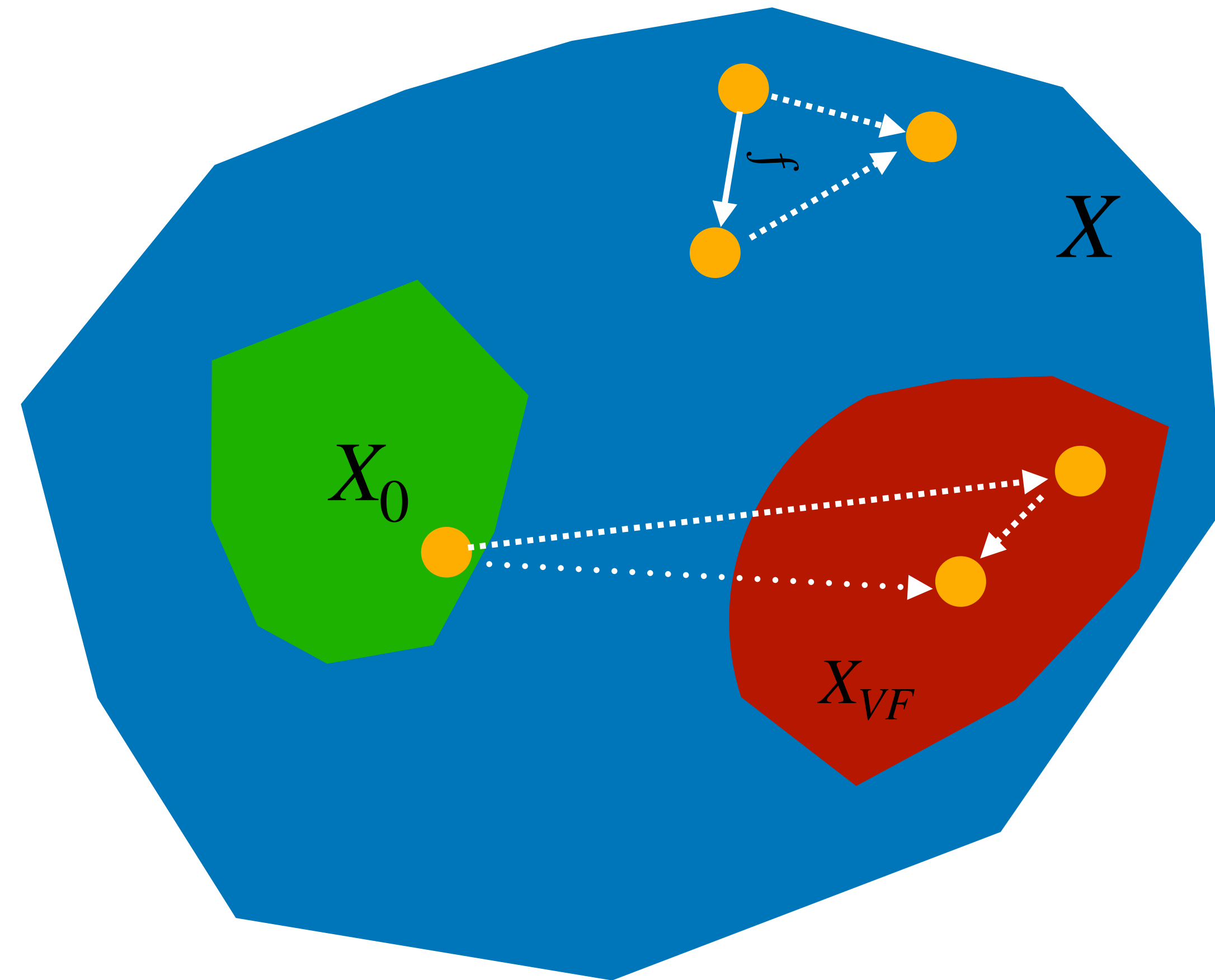
$$T(x, f(x)) \geq 0$$

$T(f(x), y) \geq 0$, implies $T(x, y) \geq 0$,

For all $x_0 \in X_0, y, y' \in X_{VF}$

$T(x_0, y) \geq 0$, and $T(y, y') \geq 0$, implies

$$T(x_0, y') \leq T(x_0, y) - \delta$$

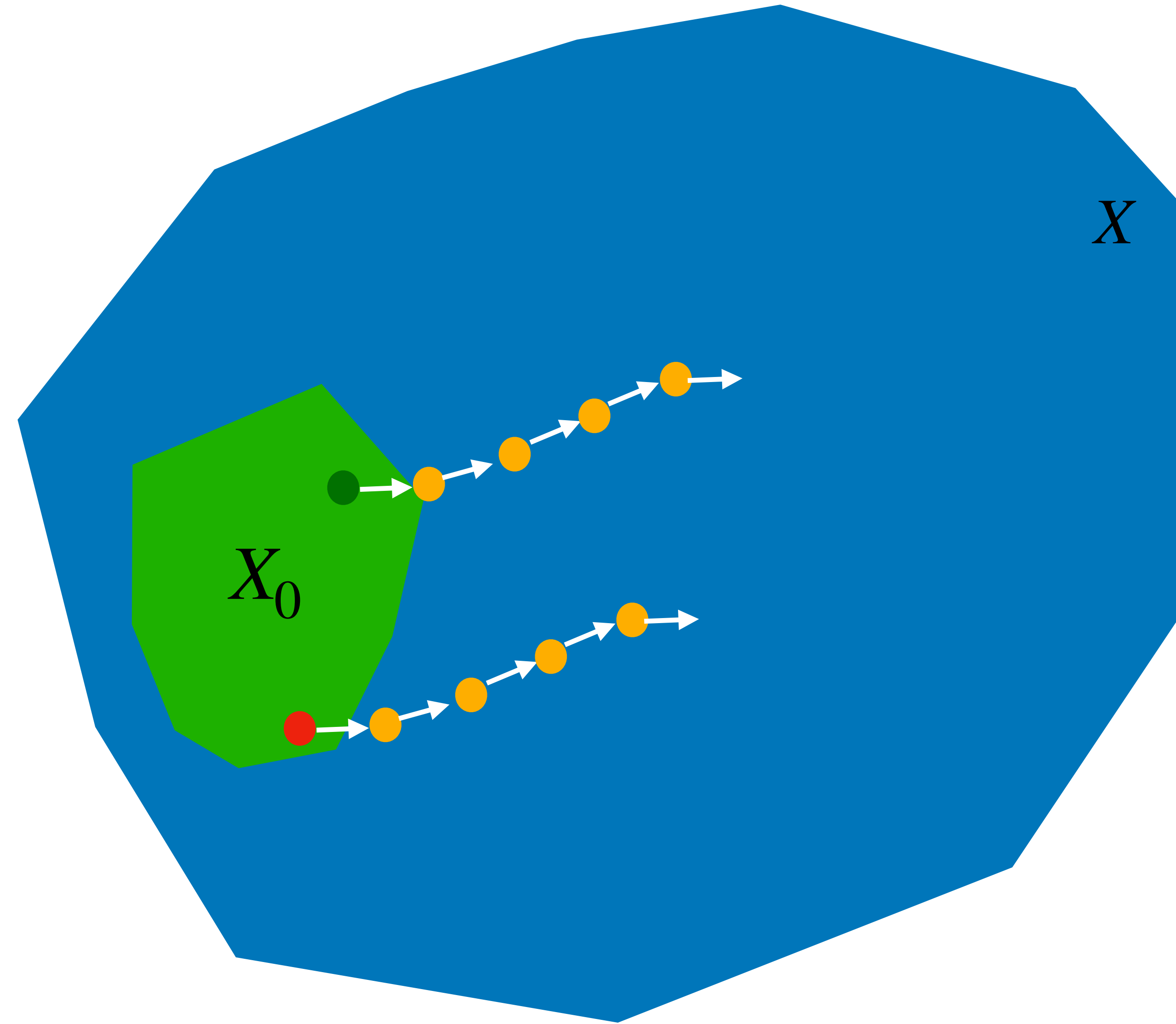


Transition Invariants for Opacity

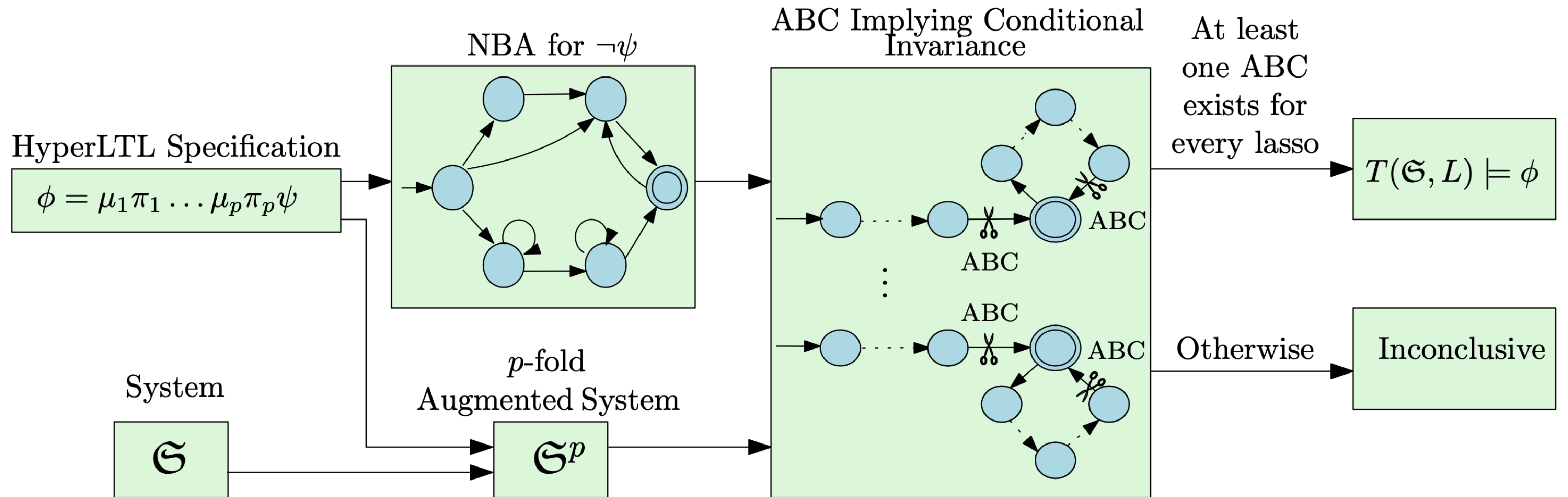
Relatively complete?

If $\exists$ player depends on arbitrary lookahead **yes**

If $\exists$ player depends on ω -lookahead - **unknown**



Verifying Hyperproperties using Barrier Certificates



Mahathi Anand, Vishnu Murali, Ashutosh Trivedi, Majid Zamani:
Formal Verification of Control Systems against Hyperproperties via Barrier Certificates. CoRR abs/2105.05493 (2021) IEEE, TAC (2024).

Tichakorn Wongpiromsarn, Ufuk Topcu , Andrew G. Lamperski 
Automata Theory Meets Barrier Certificates: Temporal Logic Verification of Nonlinear Systems. IEEE Trans. Autom. Control. 61(11): 3344-3355 (2016)

Future Directions

- under-approximations of transitive closure
- combine under- and over- approximations for hyperproperties
- Asynchronous Hyperproperties



Safe and Secure CPS



Fairness Specifications



Legal Specifications



Fairness Specifications



Saeid Tizpaz-Niari
UT El Paso



Verya Monjezi
UT El Paso



Gang Tan
Penn State University

Verya Monjezi, Ashutosh Trivedi, Gang Tan, Saeid Tizpaz-Niari:
Information-Theoretic Testing and Debugging of Fairness Defects in Deep Neural Networks. ICSE 2023: 1571-1582

Saeid Tizpaz-Niari, Ashish Kumar, Gang Tan, Ashutosh Trivedi:
Fairness-aware Configuration of Machine Learning Libraries. ICSE 2022: 909-920

Verya Monjezi , Ashish Kumar , Gang Tan , Ashutosh Trivedi , Saeid Tizpaz-Niari 
Causal Graph Fuzzing for Fair ML Software Development. ICSE Companion 2024: 402-403





State of Wisconsin Department of Corrections

COMPAS

DOC uses the Correctional Offender Management Profiling for Alternative Sanctions tool, commonly known as COMPAS, for criminogenic assessments and unified case planning. This actuarial risk assessment system contains offender information specifically designed to detect and inform dynamic case plans that will guide the offender throughout his or her lifecycle in the criminal justice system.

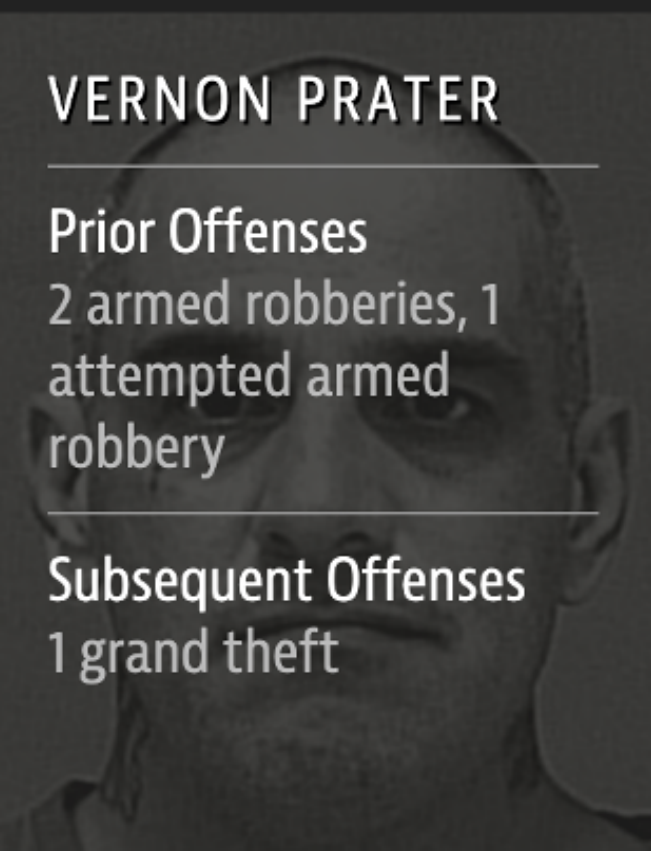
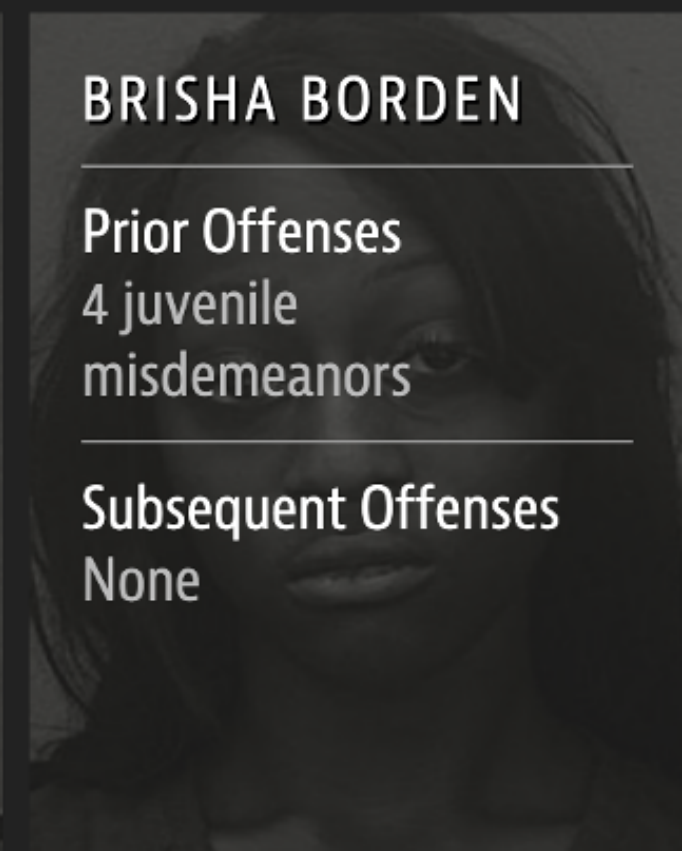
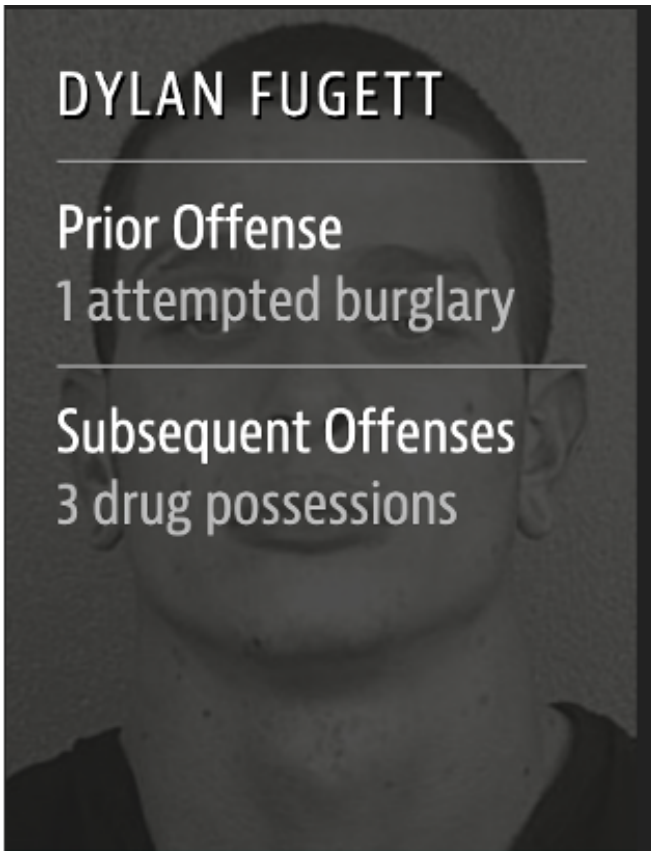
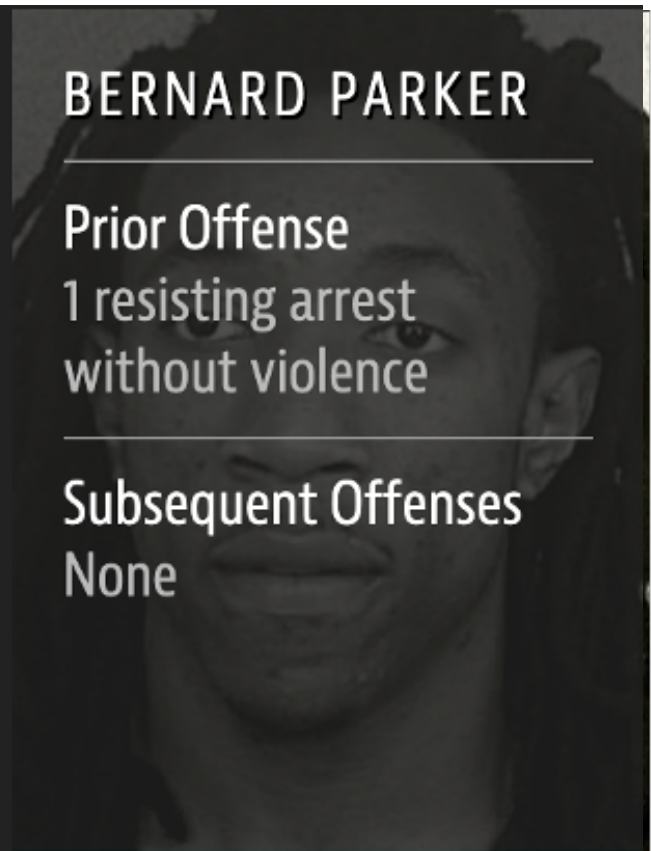


Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Two Petty Theft Arrests		Two Drug Possession Arrests	
 VERNON PRATER Prior Offenses 2 armed robberies, 1 attempted armed robbery Subsequent Offenses 1 grand theft	 BRISHA BORDEN Prior Offenses 4 juvenile misdemeanors Subsequent Offenses None	 DYLAN FUGETT Prior Offense 1 attempted burglary Subsequent Offenses 3 drug possessions	 BERNARD PARKER Prior Offense 1 resisting arrest without violence Subsequent Offenses None
LOW RISK 3	HIGH RISK 8	LOW RISK 3	HIGH RISK 10

ProPublica's main charge:
Black defendant face higher false positive rates

Fairness

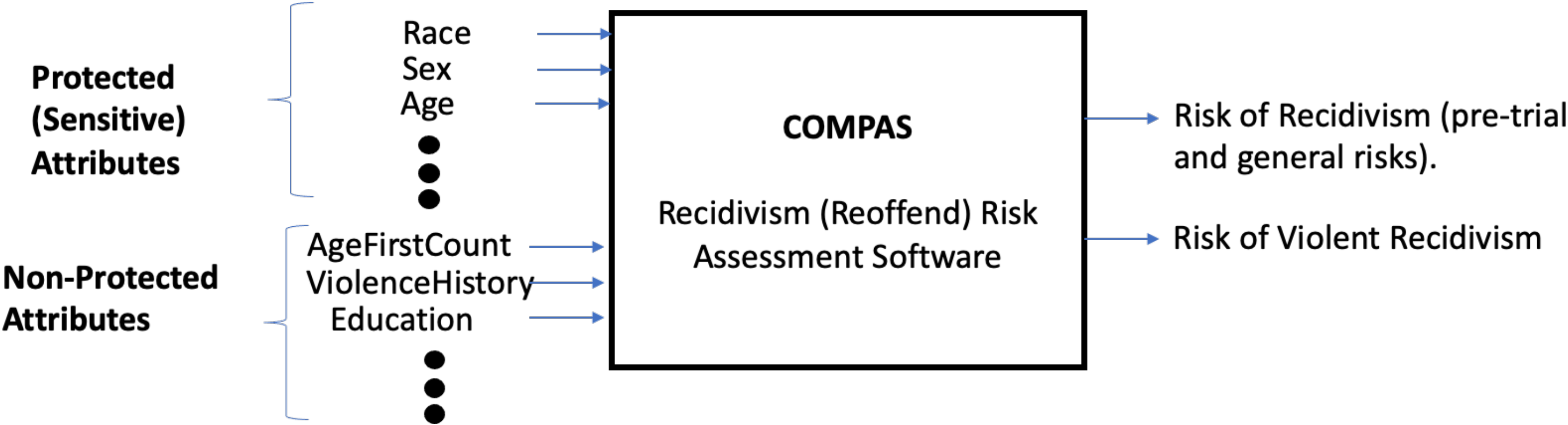
Individuals differing only in protected features, receive equivalent outcomes.

$\forall p_1, p_2 \text{ and } \forall n \text{ we have that } \mathcal{O}(p_1, n) = \mathcal{O}(p_2, n)$

In a letter, it criticized ProPublica's methodology and defended the accuracy of its test:
"Northpointe does not agree that the results of your analysis, or the claims being made based upon that analysis, are correct or that they accurately reflect the outcomes from the application of the model."

A debate over fairness semantics

How to define fairness?

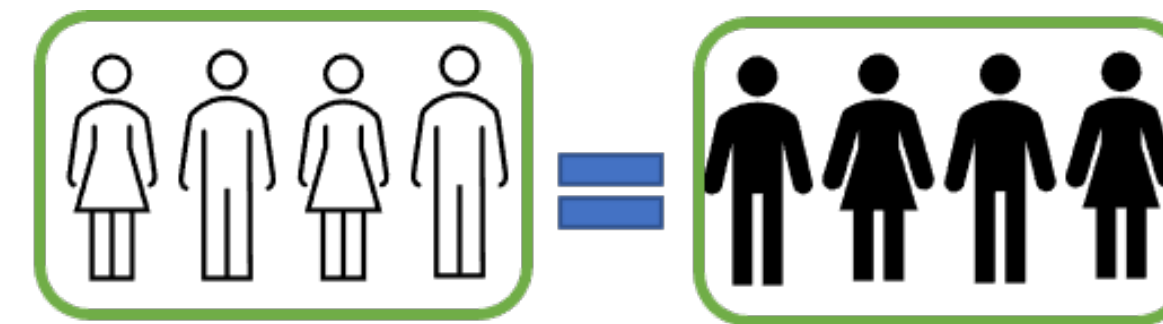


How to define fairness?

Individual Fairness:

- ▶ Fairness through unawareness (FTU):
 - ▶ Masking protected attributes during training
 - ▶ Correlation of protected attributes with non-protected ones (race and zip-code))
- ▶ Fairness through awareness (FTA):
 - ▶ Example: “Two individuals with similar qualifications should receive similar outcomes.”
 - ▶ $\forall x, y \text{ (Qualification}_x \equiv \text{Qualification}_y) \implies \text{Pred}_x \equiv \text{Pred}_y$.
 - ▶ Measuring qualification is hard!
- ▶ Causal Fairness:
 - ▶ Assuming everything else stays the same, changing a protected attribute shouldn't change outcome.
 - ▶ $\forall x, y \text{ (} x \equiv_{\text{Sex, Race}} y) \implies \text{Pred}_x \equiv \text{Pred}_y$.
 - ▶ Such counterfactuals may be unrealistic and conservative.

How to define fairness?



Group Fairness: Requires the statistics of outcomes for two groups to remain similar.

- ▶ Statistical parity difference
 - ▶ Difference in probability of favorable outcomes between privileged and under-privileged groups.
- ▶ Disparate Impact:
 - ▶ the ratio in the probability of favorable outcomes between the unprivileged and privileged groups.
- ▶ Equal Opportunity Difference (EOD):
 - ▶ Difference between true-positive rates of different groups
- ▶ Average Odd Difference (AOD):
 - ▶ The average of difference between false positive rates and true positive rates between two groups.

State of the art: Fairness Testing and Mitigation

Themis
FSE'17

Fairness Testing:
Testing Software for Discrimination

Sainyam Galhotra Yuriy Brun Alexandra Meliou

AEQUITAS
ASE'18

Automated Directed Fairness Testing

Sakshi Udeshi Pryanshu Arora Sudipta Chattopadhyay

Black-box
FSE'19

Black Box Fairness Testing of Machine Learning Models

Aniya Aggarwal Pranay Lohia Seema Nagar Kuntal Dey Diptikalyan Saha

ADT
ICSE'20

White-box Fairness Testing through Adversarial Sampling

Peixin Zhang Jingyi Wang* Jun Sun

Adversarial
Debiasing
AIES'18

Mitigating Unwanted Biases with Adversarial Learning

Brian Hu Zhang Blake Lemoine Margaret Mitchell

Prejudice
Remover
ICDM'12

Decision Theory for Discrimination-aware Classification

Faisal Kamiran*, Asim Karim[†], and Xiangliang Zhang*

Exponentiated
Gradient
ICML'18

A Reductions Approach to Fair Classification

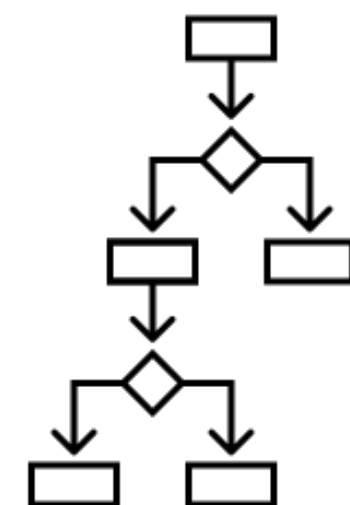
Alekh Agarwal¹ Alina Beygelzimer² Miroslav Dudík¹ John Langford¹ Hanna Wallach¹

61

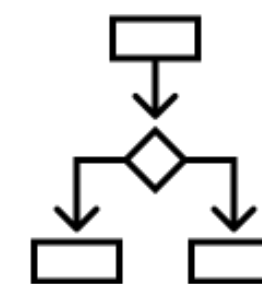
Can we improve fairness without compromising on accuracy?

Hyperparameters

Max. depth = 2



Max. depth = 1

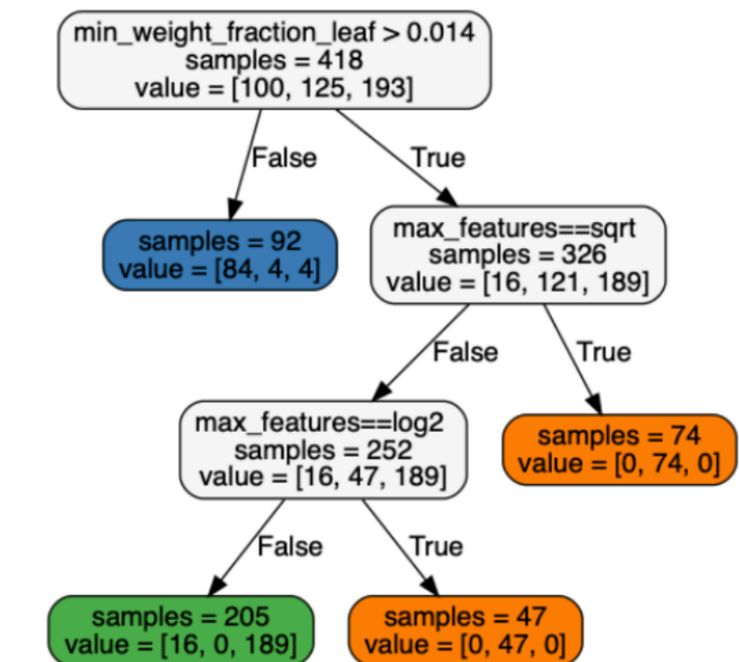
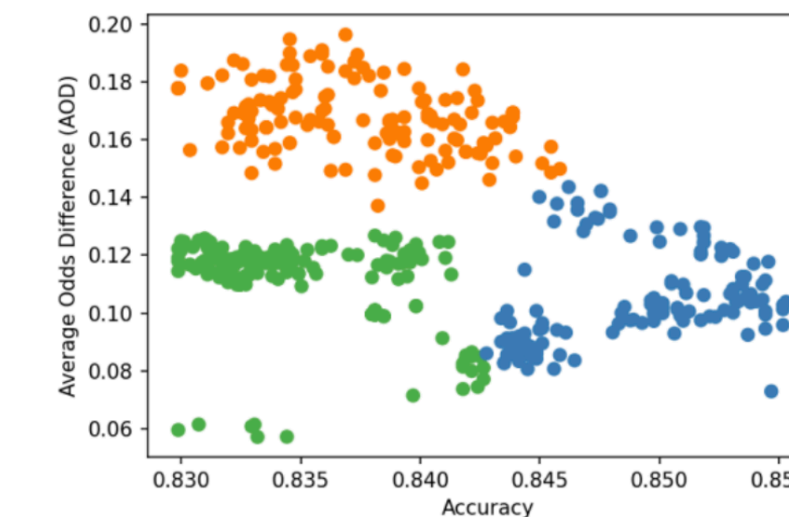
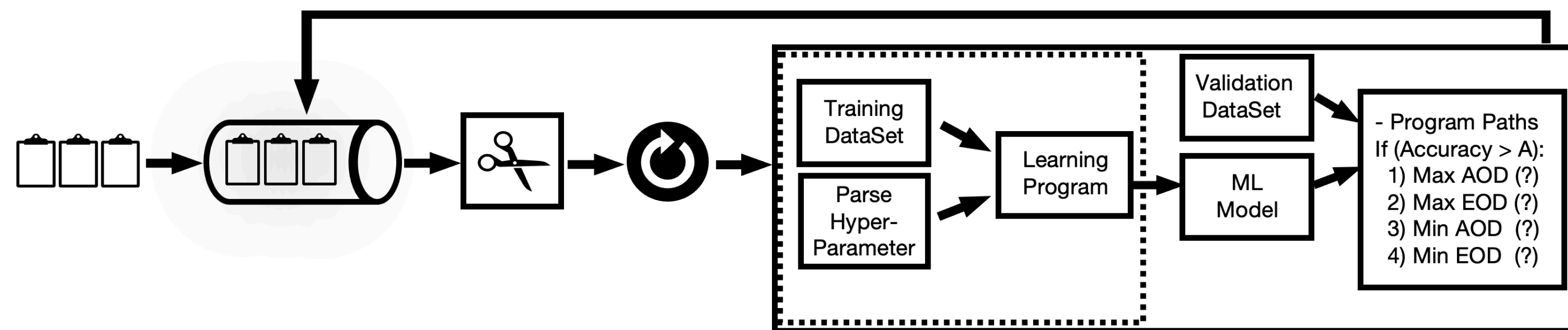


Some prominent examples of hyperparameters:

- The maximum depth of a decision tree
- The tolerance (tol) of optimization algorithm
- The number of layers/neurons in deep neural networks
- L1 vs. L2 loss function in support vector machines

Fairness-aware Configuration of Machine Learning Libraries

<https://arxiv.org/pdf/2202.06196.pdf>



Challenges:

- Testing individual models
- Modification of datasets and learning algorithms
- Low precision and Scalability.

RQ1: Hyperparameter can aggravate or suppress present biases in the dataset.

RQ2: Mutation-based Evolutionary algorithms are effective search technique.

RQ3: Some hyperparameters systematically introduce biases.

RQ4: Parfait-ML outperforms Exp. Gradients and Fairway mitigation techniques.



Reusable



Available

Our **open-source** tool **Parfait-ML** and all experimental subjects are **publicly available**:

<https://github.com/Tizpaz/Parfait-ML>



Safe and Secure CPS



Fairness Specifications



Legal Specifications



Legal Specifications



Morgan Wagner
UT El Paso



Krystia Reed
UT El Paso



Saeid Tizpaz-Niari
UT El Paso



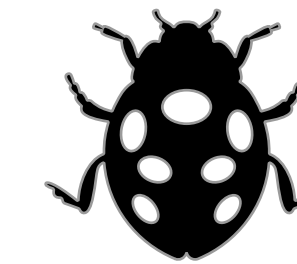
Verya Monjezi
UT El Paso



Siva Darian
CU Boulder

Saeid Tizpaz-Niari, Verya Monjezi, Morgan Wagner, Shiva Darian , Krystia Reed, Ashutosh Trivedi:
Metamorphic Testing and Debugging of Tax Preparation Software. ICSE (SEIS) 2023: 138-149

Dananjay Srinivas, Rohan Das, Saeid Tizpaz-Niari, Ashutosh Trivedi, Maria Leonor Pacheco:
On the Potential and Limitations of Few-Shot In-Context Learning to Generate Metamorphic Specifications for Tax Preparation Software. CoRR abs/2311.11979 (2023)



How to detect bugs in a **legal-critical system** such as tax-preparation software?

Absence of **oracle**

Lack of reliable datasets

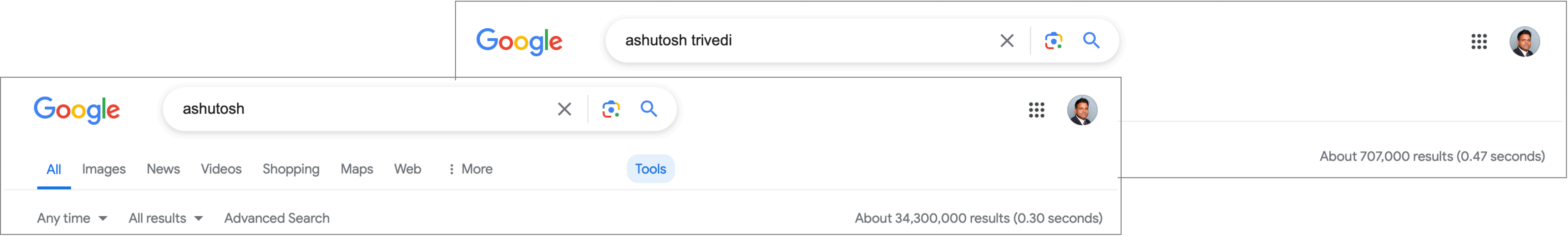
Computationally infeasible



Stare decisis (let the decision stand): courts and judges should honor “precedent” —or the decisions, rulings, and opinions from prior cases.

Corollary. Correctness properties for legal software can be defined in terms of relating similar cases!

$\forall c_1, c_2$ we have that $(c_1 \equiv c_2) \implies \mathcal{O}(c_1) \equiv \mathcal{O}(c_2)$



Metamorphic Testing

sin(pi - 5 radians) =

-0.95892427466

Rad | Deg

x!

(

)

%

AC

Inv

sin

ln

7

8

9

÷

$$\sin(x) = \sin(\pi - x)$$

sin(5 radians) =

-0.95892427466

Rad | Deg

x!

(

)

%

AC

Inv

sin

ln

7

8

9

÷

Earned Income Credit (EIC)

For use in preparing
2021 Returns



Get forms and other information faster and easier at:
• [IRS.gov](#) (English)
• [IRS.gov/Spanish](#) (Español)
• [IRS.gov/Chinese](#) (中文)
• [IRS.gov/Korean](#) (한국어)
• [IRS.gov/Russian](#) (Русский)
• [IRS.gov/Vietnamese](#) (Tiếng Việt)

Jan 10, 2022

Contents

What's New for 2021	3
Reminders	3
Chapter 1. Rules for Everyone	4
Rule 1—Adjusted Gross Income (AGI) Limits	4
Rule 2—You Must Have a Valid Social Security Number (SSN)	5
Rule 3—If You Are Separated From Your Spouse and Not Filing a Joint Return, You Must Meet Certain Rules	5
Rule 4—You Must Be a U.S. Citizen or Resident Alien All Year	6
Rule 5—You Cannot File Form 2555	6
Rule 6—Your Investment Income Must Be \$10,000 or Less	6
Rule 7—You Must Have Earned Income	8
Chapter 2. Rules if You Have a Qualifying Child	10
Rule 8—Your Child Must Meet the Relationship, Age, Residency, and Joint Return Tests	10
Rule 9—Your Qualifying Child Cannot Be Used by More Than One Person To Claim the EIC	13
Rule 10—You Cannot Be a Qualifying Child of Another Taxpayer	16
Chapter 3. Rules If You Do Not Have a Qualifying Child	17
Rule 11—You Must Meet the Age Requirements	17
Rule 12—You Cannot Be the Dependent of Another Person	18
Rule 13—You Cannot Be a Qualifying Child of Another Taxpayer	18
Rule 14—You Must Have Lived in the United States More Than Half of the Year	19
Chapter 4. Figuring and Claiming the EIC	19
Rule 15—Earned Income Limits	19
IRS Will Figure the EIC for You	21
How To Figure the EIC Yourself	21
Chapter 5. Disallowance of the EIC	22
Chapter 6. Detailed Examples	24
How To Get Tax Help	28
EIC Eligibility Checklist	32
Index	43

Future Developments

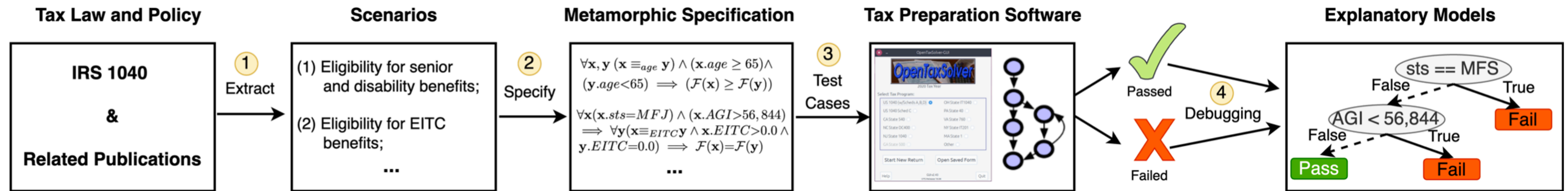
For the latest information about developments related to Pub. 596, such as legislation enacted after it was published, go to [IRS.gov/Pub596](#).

A senior (over age of 65) must receive similar or better tax benefits when compared to a person without the seniority who is similar in every other aspect (due to higher standard deductions for seniors).

$$\forall x,y((x \equiv_{\text{age}} y) \wedge (x.\text{age} \geq 65) \wedge (y.\text{age} < 65)) \implies \mathcal{F}(x) \geq \mathcal{F}(y).$$

Id	Domain	Metamorphic Property
1	Disability	$\forall \mathbf{x}, \mathbf{y}((\mathbf{x} \equiv_{age} \mathbf{y}) \wedge (\mathbf{x}.age \geq 65) \wedge (\mathbf{y}.age < 65)) \vee ((\mathbf{x} \equiv_{blind} \mathbf{y}) \wedge (\mathbf{x}.blind \wedge \neg \mathbf{y}.blind)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
2	Disability	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{s_age} \mathbf{y}) \wedge (\mathbf{x}.s_age \geq 65) \wedge (\mathbf{y}.s_age < 65)) \vee ((\mathbf{x} \equiv_{s_blind} \mathbf{y}) \wedge (\mathbf{x}.s_blind \wedge \neg \mathbf{y}.s_blind)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
3	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L27} \mathbf{y} \wedge \mathbf{x}.L27 > 0.0 \wedge \mathbf{y}.L27 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
4	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \wedge (\mathbf{x}.AGI > 56,844) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L27} \mathbf{y} \wedge \mathbf{x}.L27 > 0.0 \wedge \mathbf{y}.L27 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
5	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{AGI} \mathbf{y} \wedge \mathbf{x}.AGI \leq 56,844 \wedge \mathbf{y}.AGI > 56,844) \vee (\mathbf{x} \equiv_{L27} \mathbf{y} \wedge \mathbf{x}.L27 > 0.0 \wedge \mathbf{y}.L27 = 0.0) \vee (\mathbf{x} \equiv_{QC} \mathbf{y} \wedge \mathbf{x}.QC \geq 1 \wedge \mathbf{y}.QC = 0) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
6	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \wedge (\mathbf{x}.AGI \leq 56,844) \wedge (\mathbf{x}.QC \leq 3) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{L27} \mathbf{y}) \wedge \mathbf{x}.L27 \geq \mathbf{y}.L27) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
7	CTC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \wedge (\mathbf{x}.AGI \leq 200k) \forall \mathbf{y}((\mathbf{x} \equiv_{L19} \mathbf{y}) \wedge (\mathbf{x}.L19 \geq \mathbf{y}.L19)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
8	CTC	$\forall \mathbf{x}, \mathbf{x}'(\mathbf{x}.sts = \mathbf{x}'.sts = MFJ) \wedge (\mathbf{x}.AGI < 400k) \wedge (\mathbf{x}'.AGI \geq 400k) \wedge \lceil \mathbf{x}'.AGI - 400k \rceil_{1k} * 0.05 < \mathbf{x}'.QC * 2k + \mathbf{x}.OD * 0.5k \implies \forall \mathbf{y}, \mathbf{y}'(\mathbf{x} \equiv_{\{QC, OD\}} \mathbf{y}) \wedge (\mathbf{x}' \equiv_{\{QC, OD\}} \mathbf{y}') \wedge (0 \leq \mathbf{y}.QC = \mathbf{y}'.QC \leq \mathbf{x}.QC = \mathbf{x}'.QC \leq 10) \wedge (0 \leq \mathbf{y}.OD = \mathbf{y}'.OD \leq \mathbf{x}.OD = \mathbf{x}'.OD \leq 10) \implies (\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})) \geq (\mathcal{F}(\mathbf{x}') - \mathcal{F}(\mathbf{y}'))$
9	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \wedge \mathbf{x}.L29 > 0.0 \wedge \mathbf{y}.L29 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
10	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \wedge (\mathbf{x}.AGI \geq 180k) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \wedge \mathbf{x}.L29 > 0.0 \wedge \mathbf{y}.L29 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
11	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \wedge (\mathbf{x}.AGI \leq 160k) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \wedge \mathbf{x}.L29 \geq \mathbf{y}.L29) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
12	ETC	$\forall \mathbf{x}, \mathbf{x}'(\mathbf{x}.sts = \mathbf{x}'.sts = MFJ) \wedge (\mathbf{x}.AGI \leq 160k) \wedge (160k < \mathbf{x}'.AGI < 180k) \implies \forall \mathbf{y}, \mathbf{y}'((\mathbf{x} \equiv_{L29} \mathbf{y}) \wedge (\mathbf{x}' \equiv_{L29} \mathbf{y}') \wedge (\mathbf{x}.L29 = \mathbf{x}'.L29 \geq \mathbf{y}.L29 = \mathbf{y}'.L29)) \implies (\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})) \geq (\mathcal{F}(\mathbf{x}') - \mathcal{F}(\mathbf{y}'))$
13	ID	$\forall \mathbf{x}, \mathbf{y}(\mathbf{x} \equiv_{MDE} \mathbf{y}) \wedge (\mathbf{x}.MDE \leq \mathbf{x}.AGI * 7.5\%) \wedge (\mathbf{y}.MDE = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
14	ID	$\forall \mathbf{x}(\neg \mathbf{x}.iz) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{MDE} \mathbf{y} \wedge \mathbf{x}.MDE > 0.0 \wedge \mathbf{y}.MDE = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
15	ID	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{iz, L12} \mathbf{y}) \wedge (\mathbf{x}.iz \wedge \neg \mathbf{y}.iz) \wedge (\mathbf{x}.L12 \leq 24.8k \wedge \mathbf{y}.L12 = 0.0)) \implies \mathcal{F}(\mathbf{x}) \leq \mathcal{F}(\mathbf{y})$
16	ID	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{iz, L12} \mathbf{y}) \wedge (\mathbf{x}.iz \wedge \neg \mathbf{y}.iz) \wedge (\mathbf{x}.L12 > 24.8k \wedge \mathbf{y}.L12 = 0.0)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$

Metamorphic Testing and Debugging of Tax Preparation Software (<https://arxiv.org/abs/2205.04998>)



Forensic DNA Software

- New York City's Office of Chief Medical Examiner (OCME) for thousands of criminal cases between 2011 and 2017
- Undisclosed data dropping method `CheckFrequencyForRemoval()`
- Falsely skew results toward false inclusion for individuals whose DNA was not present.

"Do I Qualify?" Screening Software

- Poverty management systems in Pennsylvania (Check Eligibility)
- Comparative implementation of benefit eligibility handbook
- Errors in the eligibility checking: Exclude the most vulnerable families from receiving the essential aids



Safe and Secure CPS



Fairness Specifications



Legal Specifications

Study Suggests It's More Dangerous For Black Pedestrians To Cross The Street

Study Suggests It's More Dangerous For Black Pedestrians To Cross The Street

By Kim Bellware

Jun 6, 2014, 05:04 PM EDT

Updated Jun 9, 2014



LEAVE A COMMENT



Study: Drivers Less Likely to Yield for Black Pedestrians



By Angie Schmitt

1:00 PM EDT on May 23, 2014

Bias Behind the Wheel: Fairness Analysis of Autonomous Driving Systems

XINYUE LI, Peking University, China

ZHENPENG CHEN, Nanyang Technological University, Singapore

JIE M. ZHANG, King's College London, United Kingdom

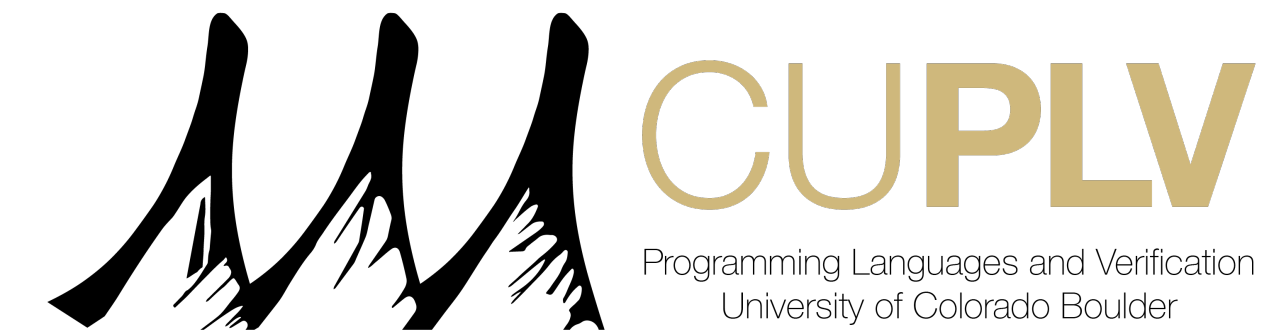
FEDERICA SARRO, University College London, United Kingdom

YING ZHANG, Peking University, China

XUANZHE LIU, Peking University, China

This paper analyzes fairness in automated pedestrian detection, a crucial but under-explored issue in autonomous driving systems. We evaluate eight state-of-the-art deep learning-based pedestrian detectors across demographic groups on large-scale real-world datasets. To enable thorough fairness testing, we provide extensive annotations for the datasets, resulting in 8,311 images with 16,070 gender labels, 20,115 age labels, and 3,513 skin tone labels. Our findings reveal significant fairness issues, particularly related to age. The undetected proportions for children are 20.14% higher compared to adults. Furthermore, we explore how various driving scenarios affect the fairness of pedestrian detectors. We find that pedestrian detectors demonstrate significant gender biases during night time, potentially exacerbating the prevalent societal issue of female safety concerns during nighttime out. Moreover, we observe that pedestrian detectors can demonstrate both enhanced fairness and superior performance under specific driving conditions, which challenges the fairness-performance trade-off theory widely acknowledged in the fairness literature. We publicly release the code, data, and results to support future research on fairness in autonomous driving.

Acknowledgements



PhD Advisees

Devendra Bhave (IIT Bombay, 2020)

Taylor Dohmen

John Komp

Tianhan Lu

Mateo Perez

Saeid Tizpaz-Niari (2020)

Vishnu Murali

David Baines

Alireza Nadali

Amin Falah

Student Collaborators

Jordan Perr-Sauer (MS)

Kalyani Dole (PhD, IITB)

Noah Topper (BS, UCF)

Milad Kazemi (PhD, Newcastle)

Ismail Alkhouri (PhD, UCF)

Mahathi Anand (PhD, TUM)

Shiva Darian (PhD, CUB)

Sankalp Gambhir (BTech, IITB)

Aman Singh (BTech, IITB)

Ian McKibben (BS, CUB)

John Paul Martin (MS, CUB)

Juraj Culak (MS, CUB)

Shemal Somil Lalaji (MS, 2019)

Anniruddha Phatak (MS/PhD)

Aniket Lata (MS, 2016)

Ramdas Diwakaran (MS, 2017)

Zachary Mckevitt (BS)

Emily Milican (BS)

Mateo Perez (BS, 2020)

Abolfazl Lavei (LMU Munich)

Aviral Kumar (IIT Bombay)

Vrunda Dave (IIT Bombay)

Lakshmi Manasa (IIT Bombay)

Salar Moarref (UPenn)

Faculty/Researcher Collaborators

Fabio Somenzi

Majid Zamani

Bor-Yuh Evan Chang

Sriram Sankaranarayanan

Tamara Lehman

Gowtham Kaki

Alvaro Velasquez

Pavol Černý (TU Wien)

Behrouz Touri (UCSD)

Ernst Moritz Hahn (UTwente)

Sven Schewe (Liverpool)

Dominik Wojtczak (Liverpool)

Krishna S. (IIT Bombay)

Sadegh Soudjani (Newcastle)

George K. Atia (UCF)

Andre Beckus (AFRL)

Emmanuel Filiot (ULB)

Jean-Francois Raskin (ULB)

Sadegh Soudjani (Newcastle)

R. Phawade (IIT Dharwad)

Shibashis Guha (TIFR)

Gethin Norman (Glasgow)

Vojtech Forejt (Diffblue)

Akshay S. (IIT Bombay)

Supratik Chakraborty (IITB)

K Subramani (West Virginia)

Alvaro Velasquez (AFRL)

Jyotirmoy Deshmukh (USC)

Saeid Tizpaz-Niari (UTEP)

Krystia Reed (UTEP)

Image Credit

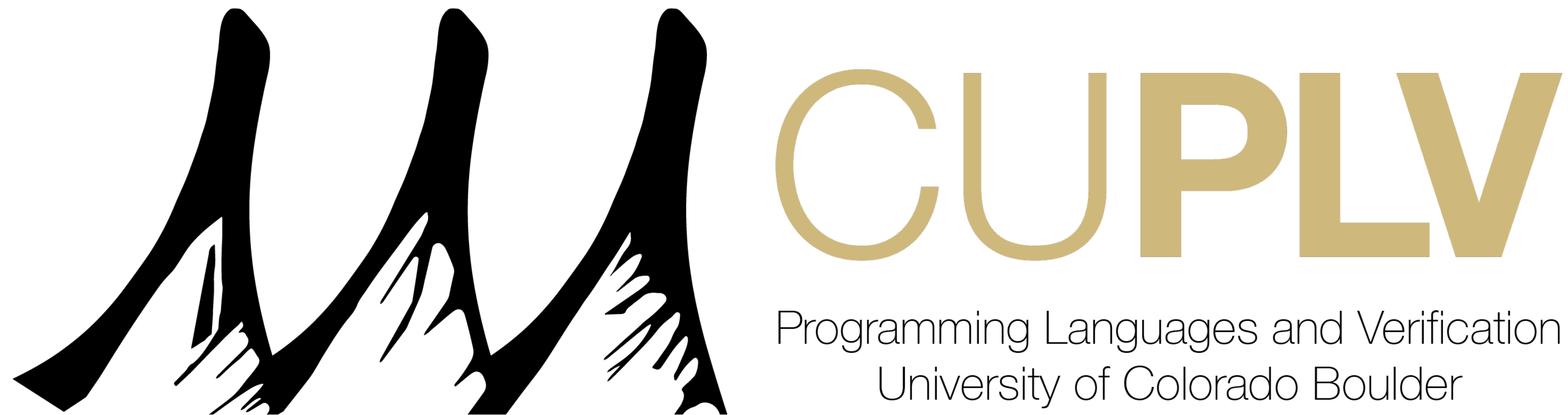
Photo by [Samuele Errico Piccarini](#) on [Unsplash](#)

Photo by [Matthew Henry](#) on [Unsplash](#)

Photo by [Tingey Injury Law Firm](#) on [Unsplash](#)

Photo by [Google DeepMind](#) on [Unsplash](#)

Photo by [Mr Cup / Fabien Barral](#) on [Unsplash](#)



[http://ashutoshrivedi
plv.colorado.edu](http://ashutoshrivedi.plv.colorado.edu)