Human-in-the-Loop Robotics: Specification, Verification, and Synthesis

Sanjit A. Seshia

Associate Professor

EECS, UC Berkeley

Students: D. Sadigh, A. Puggelli, W. Li, K. Driggs-Campbell
Collaborators: R. Bajcsy, A. Sangiovanni-Vincentelli, S. Shankar Sastry

July 2014

Human-in-the-Loop Robotics

Driver Assistance in Cars

Fly-by-wire Cockpit Interfaces

Robotic Surgery & Medicine

Semi-Autonomous Manufacturing

This Talk

FORMAL METHODS

Provable Guarantees
(correctness, performance, etc.)

New Verification/Synthesis Problems

- Interfaces + Control
- Quantitative req.
- Uncertainty

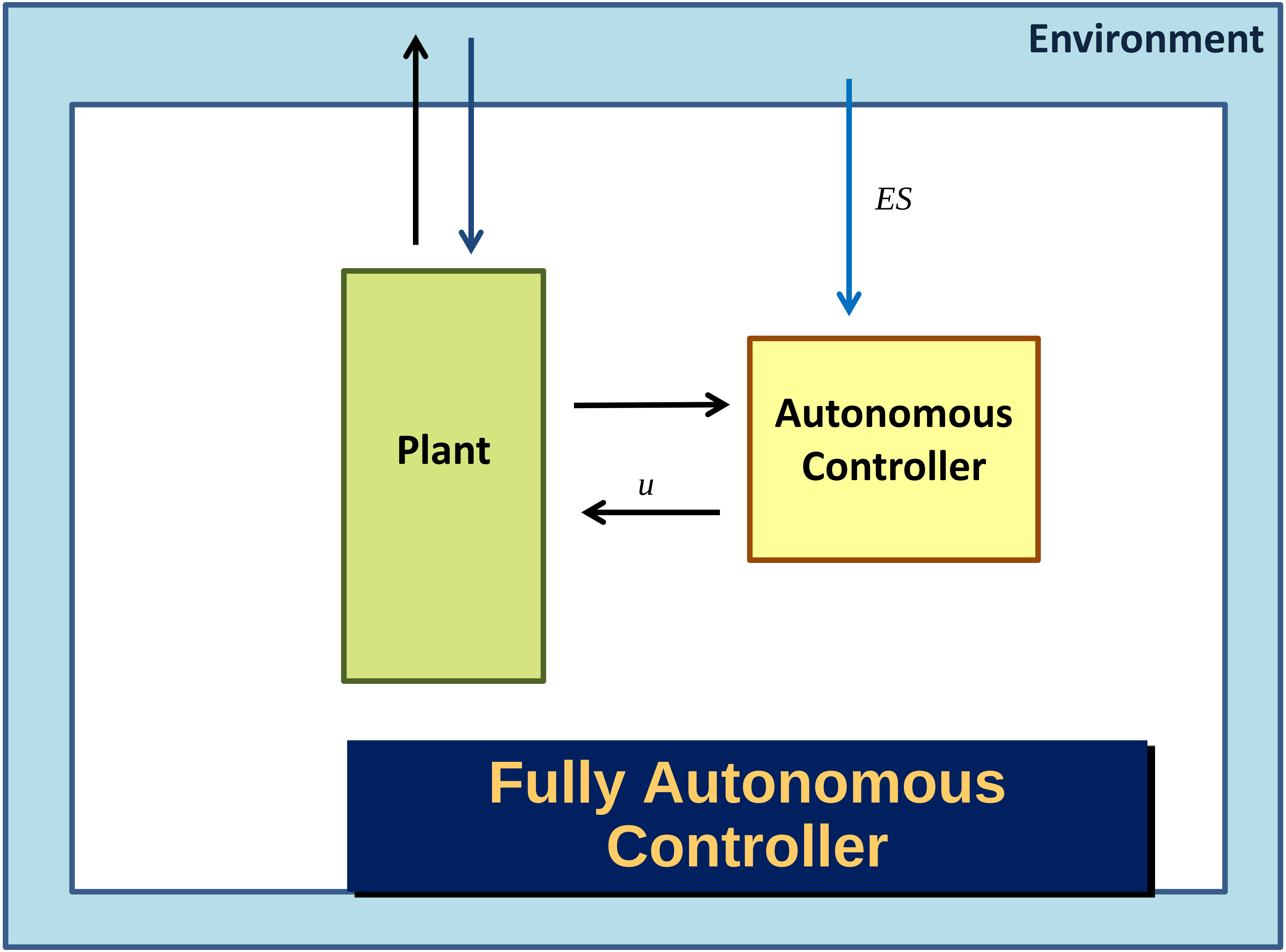
HUMAN-IN-THE-LOOP (HuIL) ROBOTICS

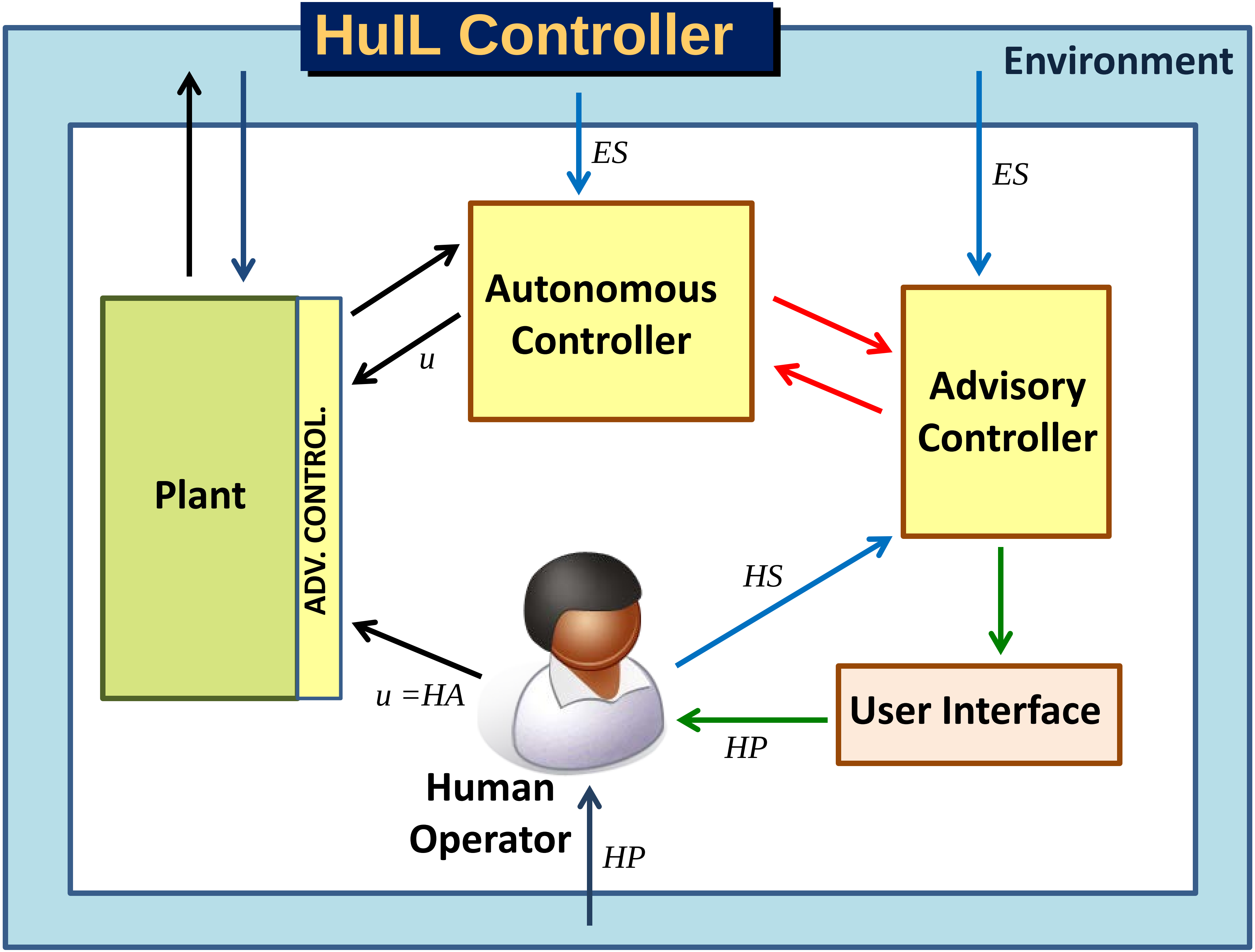
Outline

- **Formal Modeling for HuIL Robotics**
- **Specification**
- **Verification**
- **Synthesis**

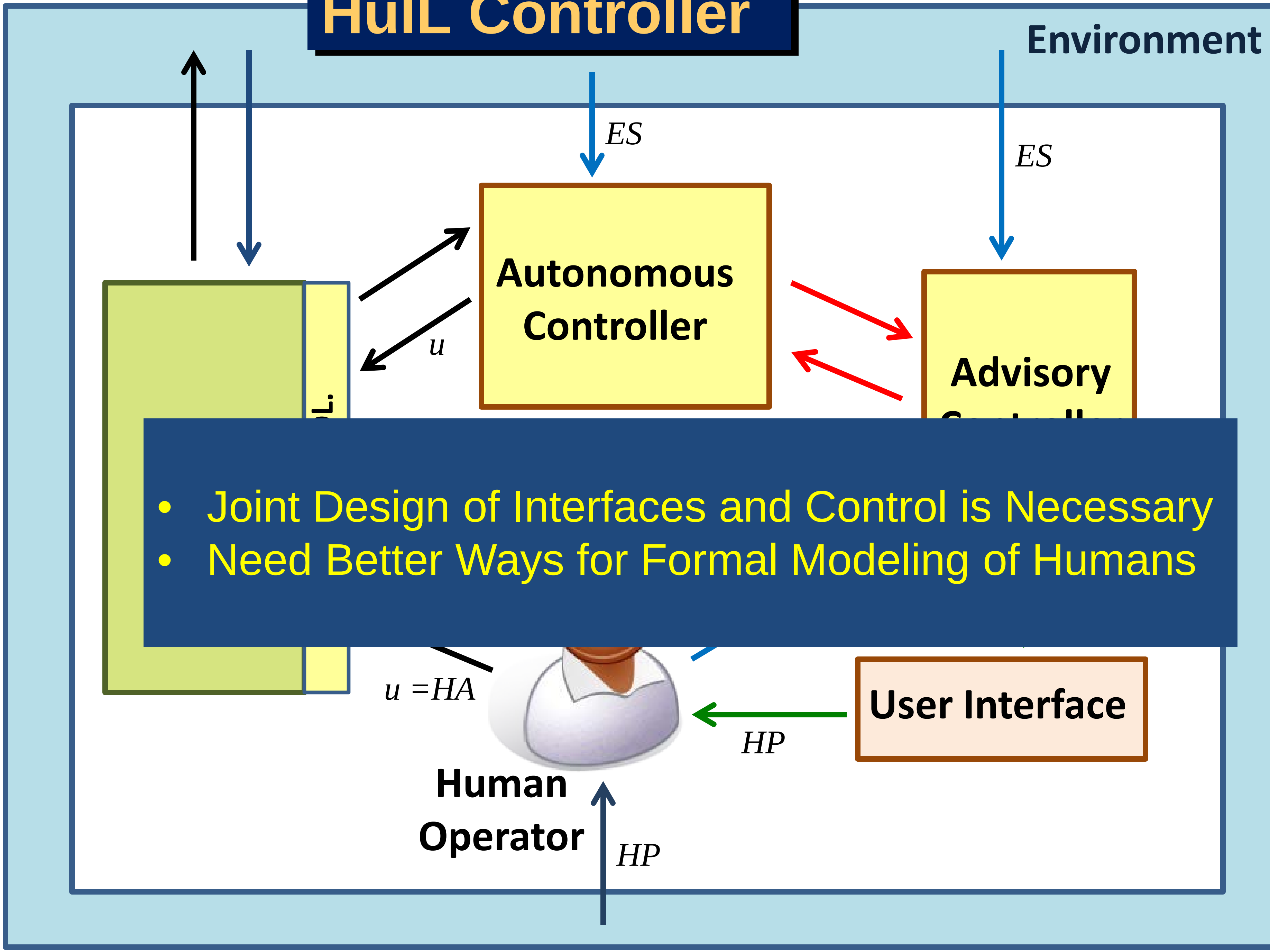
- Li, Sadigh, Sastry, Seshia, “Synthesis for Human-in-the-Loop Control Systems”, TACAS 2014.
- Sadigh et al., “Data-Driven Probabilistic Modeling and Verification of Human Driver Behavior”, 2014.

Modeling


Environment
ES
Plant
Autonomous Controller
u
Fully Autonomous Controller


HulL Controller
Environment
ES
ES
Autonomous Controller
Advisory Controller
Plant
ADV. CONTROL.
u
u =HA
HS
HP
User Interface
Human Operator
HP

HuIL Controller



- Joint Design of Interfaces and Control is Necessary
- Need Better Ways for Formal Modeling of Humans

Specification / Requirements

NHTSA Preliminary Policy Statement, May 2013

U.S. Department of Transportation Releases Policy on Automated Vehicle Development

NHTSA 14-13
Thursday, May 30, 2013
Contact: Karen Aldana, 202-366-9550, Public.Affairs@dot.gov

Provides guidance to states permitting testing of emerging vehicle technology

WASHINGTON - The U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA) today announced a new policy concerning vehicle automation, including its plans for research on related safety issues and recommendations for states related to the testing, licensing, and regulation of "autonomous" or "self-driving" vehicles. Self-driving vehicles are those in which operation of the vehicle occurs without direct driver input to control the steering, acceleration, and braking and are designed so that the driver is not expected to constantly monitor the roadway while operating in self-driving mode.

NHTSA
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION

- Levels of Automation in NHTSA document
  - Level 0: No Automation
  - Level 1: Function-Specific Automation
  - Level 2: Combined Function Automation
  - Level 3: Limited Self-Driving Automation
  - Level 4: Full Self-Driving Automation

Focus on Level 3: Limited Self-Driving Automation

"*Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time. The vehicle is designed to ensure safe operation during the automated driving mode*."

Specification for Level 3

[Li, Sadigh, Sastry, Seshia, TR 2013; TACAS'14]

4 Requirements common to all Level 3 systems:

- **Effective Monitoring**
  – **Should be able to monitor traffic & environment conditions relevant for correct operation**
- **Conditional Correctness**
  – **Should guarantee correct (safe) operation under those conditions**
- **Prescient Switching**
  – **Should request driver to take over well in advance (*T* sec advance warning)**
- **Minimally Intervening**
  – **Should *rarely* request driver intervention (only when there is high probability of imminent failure)**

Formal Specification for Level 3

[Li, Sadigh, Sastry, Seshia, TR 2013; TACAS'14]

4 Requirements common to all Level 3 systems:

- **Effective Monitoring**
  - **Sufficient Sensing**
- **Conditional Correctness**
  - **"Traditional" Formal Specification (e.g., in Linear Temporal Logic)**
- **Prescient Switching**
  - **Response Time Specification (bound *T*, or fine-grained model)**
- **Minimally Intervening**
  - **Cost Function**

Synthesis

Problem Formulation

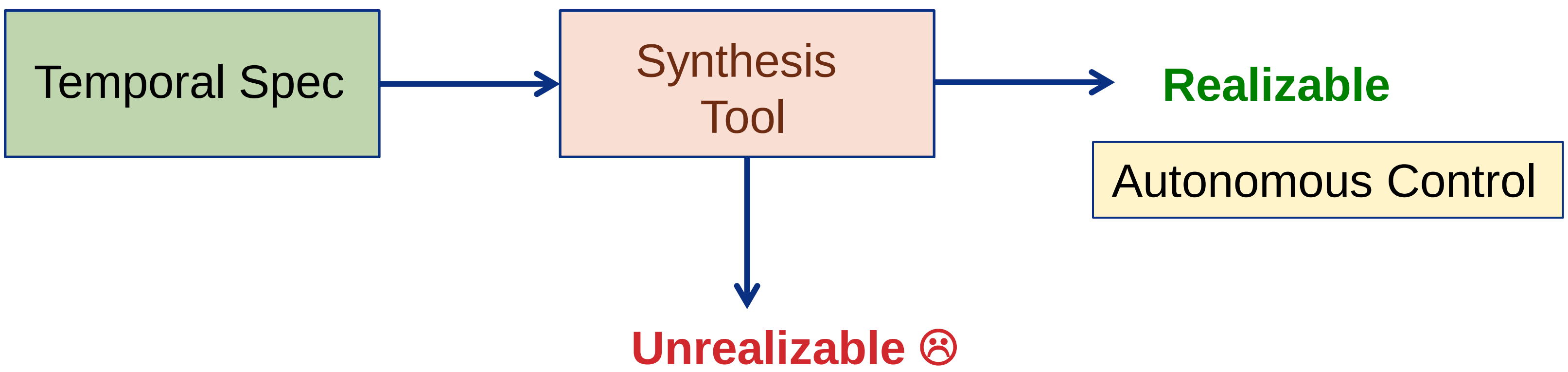
- Given **driver's response time** parameter T
- Given a **cost function** penalizing human's intervention
- Given a **high level specification** (GR1(1) formula)

Synthesize a **fully autonomous controller** satisfying the specification
Or a **Human in the Loop Controller** (composition of auto-controller, human operator, advisory controller) that is:

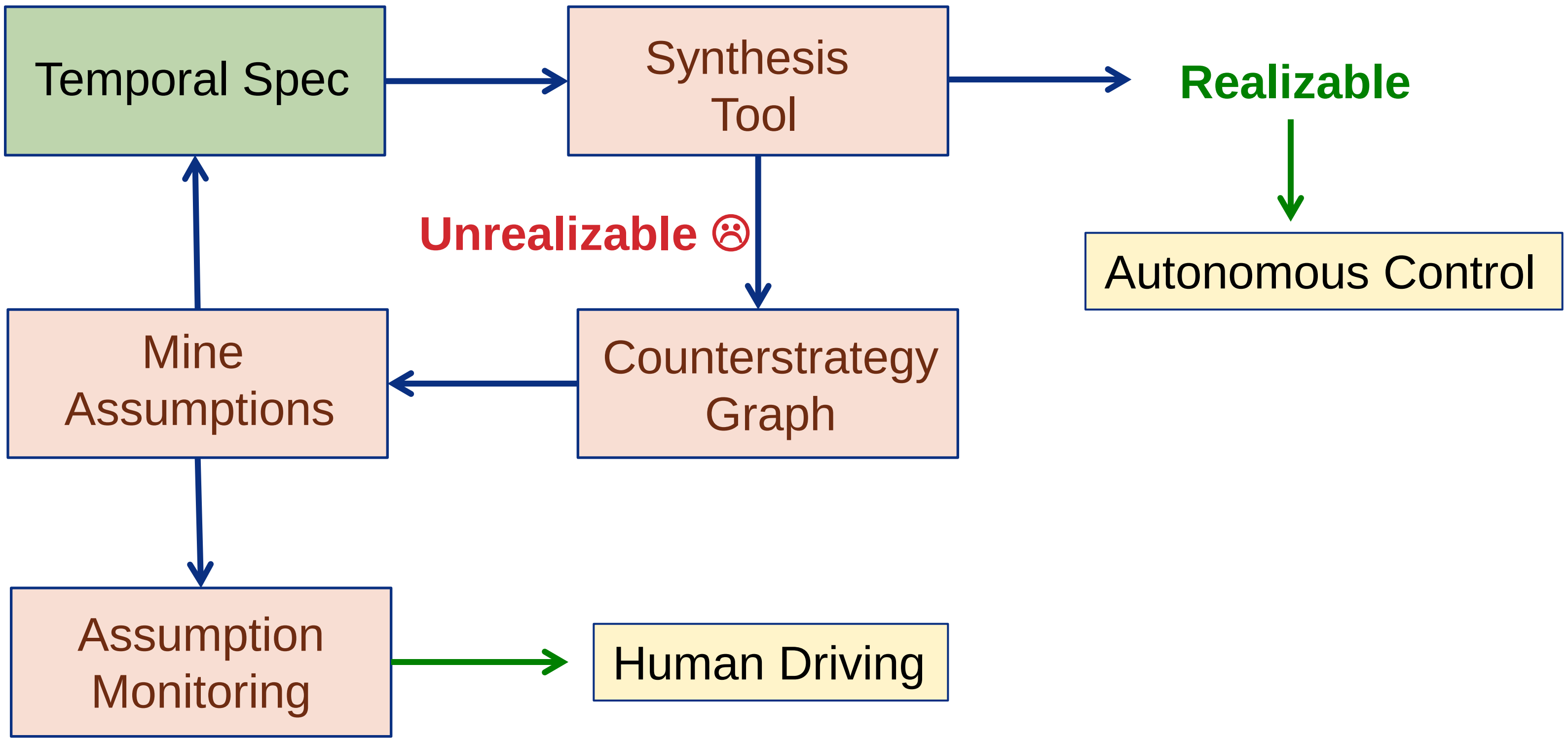
- Effectively Monitoring
- Prescient (with parameter T)
- Minimally intervening
- Conditionally correct

Li, Sadigh, Sastry, Seshia, "Synthesis for Human-in-the-Loop Control Systems", TACAS 2014.

Approach

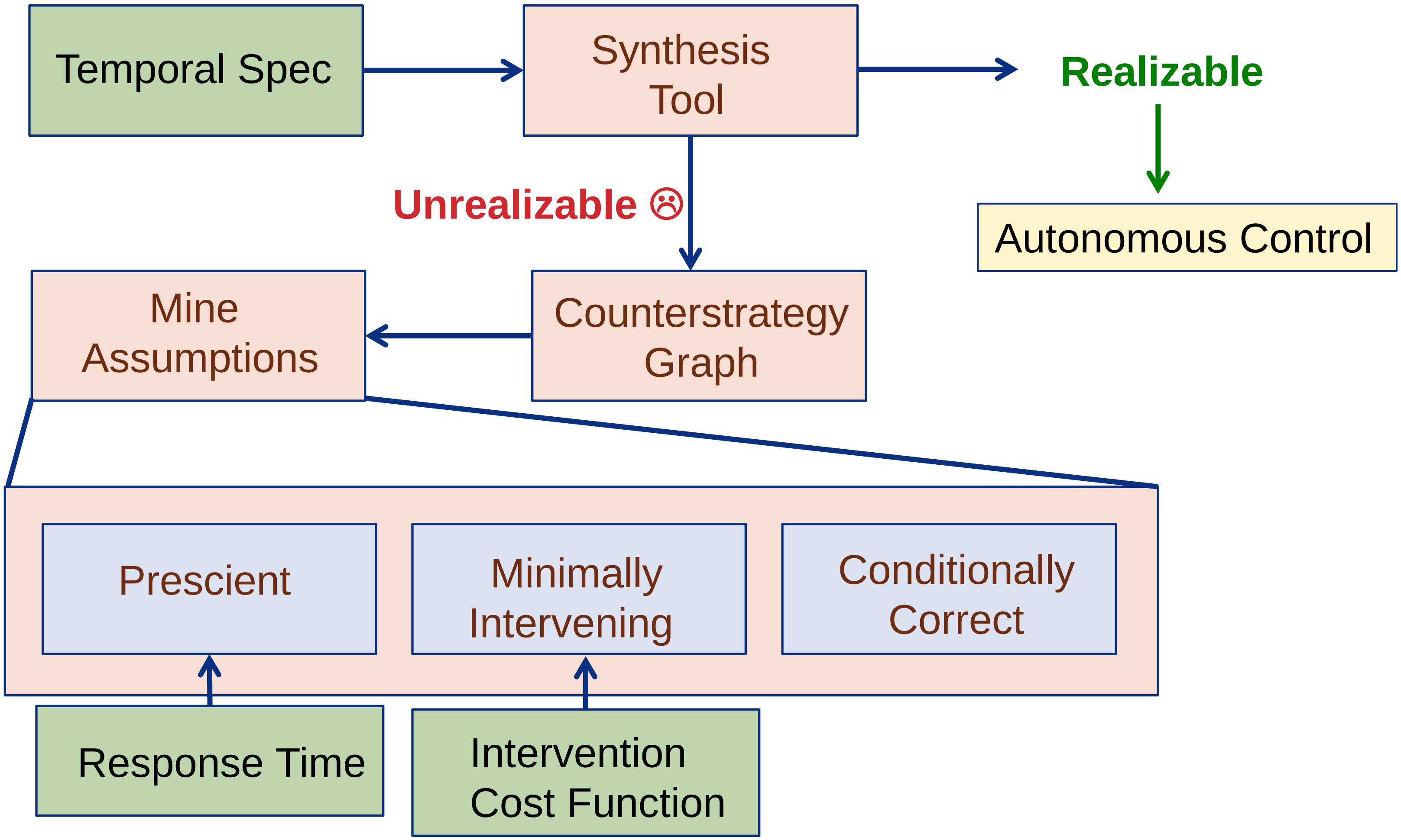

Temporal Spec
Synthesis Tool
Realizable
Autonomous Control
Unrealizable ☹

Approach



Li, Dworkin, Seshia, “Mining Assumptions for Synthesis”, MEMOCODE 2011.
Li, Sadigh, Sastry, Seshia, “Synthesis for Human-in-the-Loop Control Systems”, TACAS 2014.

Approach



Li, Sadigh, Sastry, Seshia, “Synthesis for Human-in-the-Loop Control Systems”, TACAS 2014.

Solution Sketch

Extract CounterStrategies

- Counterstrategy Graph G: similar to game graph, but represents winning strategies for environment

Infer Assumptions

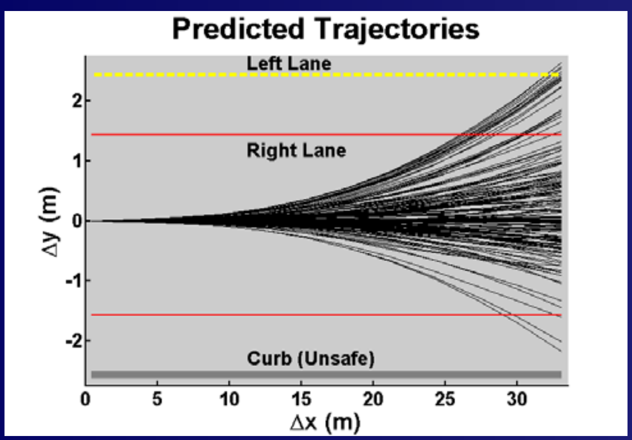
- Identify Winning Nodes / SCCs in G for environment
- Compute “min cut” in G
  - Minimize cost function capturing probability of intervention
  - Maintain distance of at least T steps from winning nodes for environment

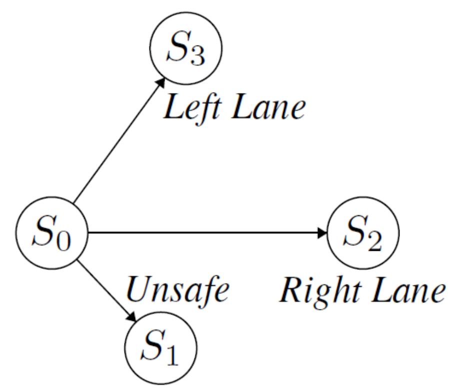
Verification

Probabilistic Verification with Uncertainties

Collect Driving Data ⇨ Learn Probabilistic Model that captures uncertainties (CMC, CMDP, etc.) ⇨ Probabilistic Model Checking → Yes/No + Probab.

Prob. Temporal Logic Property ↘ Probabilistic Model Checking





Transition	Transition Probability Interval
$S_0 \rightarrow S_1$	[0.019,0.021]
$S_0 \rightarrow S_2$	[0.890,0.980]
$S_0 \rightarrow S_1$	[0.048,0.053]

Papers: Puggelli et al., CAV 2013; Sadigh et al., AAAI Symp. 2014

Formal Verification + UI Testing

- **Blend formal verification of models of the system with human-in-the-loop UI testing**
- S. A. Seshia, “Verifying High-Confidence Interactive Systems: Electronic Voting and Beyond”, Jan 2013.

Conclusion

FORMAL METHODS

Provable Guarantees
(correctness, performance, etc.)

↓ ↑

New Verification/Synthesis Problems

- Interfaces + Control
- Quantitative req.
- Uncertainty

HUMAN-IN-THE-LOOP (HuIL) ROBOTICS