System-Theoretic Safety Analysis for Teams of Collaborative Controllers

2023 STAMP Workshop

Andrew Kopeikin*

6 June 2023

***PhD Candidate, MIT Lincoln Lab Technical Staff, USAF Officer (Reserve), CFII**

kopeikin@mit.edu

Human Team vs Human-Machine Interactions

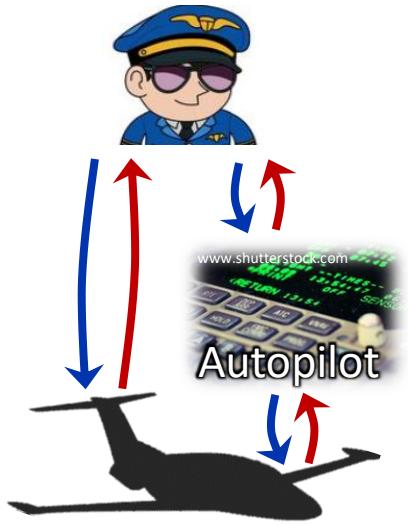
Interactions in current human-automation systems are simpler

Human as Supervisor

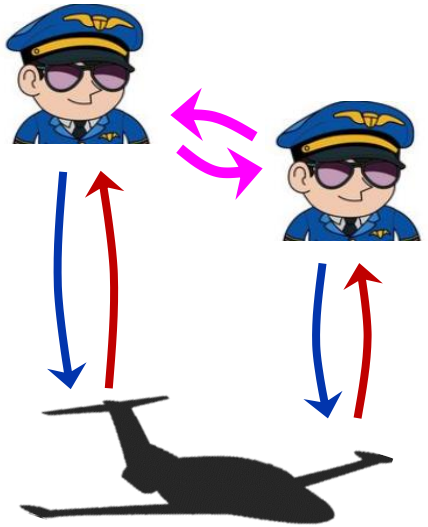
- sets control goal
- supervises
- intervenes

Automated Controller

- feedback control of aircraft only



Interactions in human teams are complex

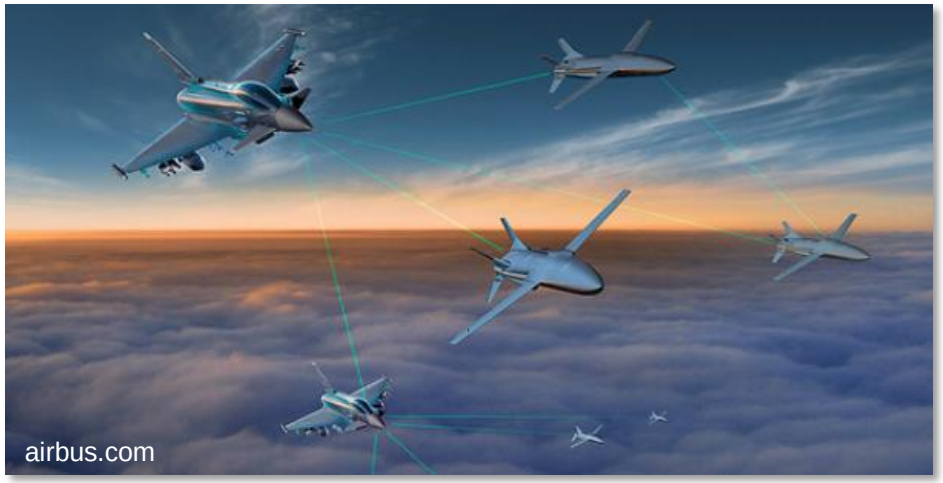


Collaborative Control

- establish roles
- change authorities
- team cognition
- coordination
- coupled in control loops

Seek to engineer systems with complex team-inspired interactions

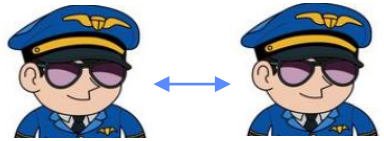
Aviation Concepts Seeking Team-Like Interactions



- Simplified Vehicle Operations (UAM*)
- Remote Supervisory Operations (UAM*)
- Single Pilot Operations (Airlines)

- Multi-UAS & Swarms
- Manned – Unmanned Aircraft Teaming
- Manned – Unmanned Aircrew

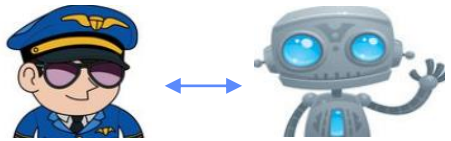
Human Teaming



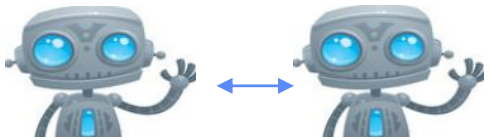
human-human

Inspires

New Interactions in Designed Systems



human-machine



machine-machine

Despite all of the interest – none of these systems have been fielded

**UAM: Urban Air Mobility*

Challenges Engineering Safe Collaborative Systems

Team-inspired interactions challenging

Many models,
but few for safety or
beyond system boundary

Need improved design techniques

Current processes are oversimplified or face drawbacks for safety

Lack effective safety assurance methods

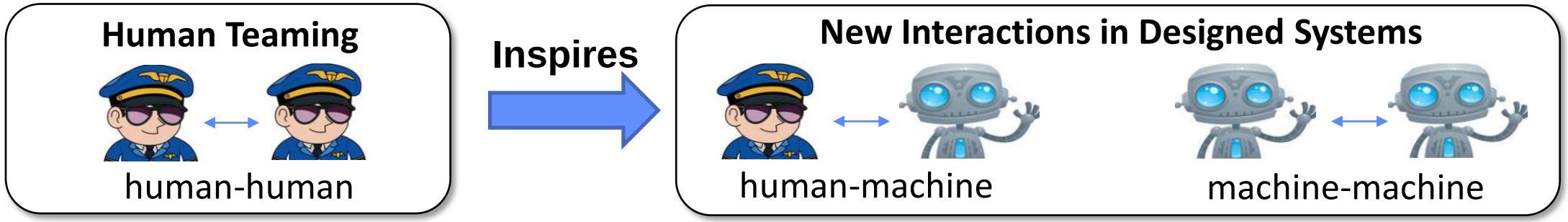
Current techniques applied too late & inadequate

Clear gap in hazard analysis capability

[Holbrook et al '20], [Mosier et al '17], [Pritchett et al '18], [Prinzel '19]
[NATO HFM '20], [Connors '17], [Kearns '18], & many more…

Beyond current modeling, analysis, design, and assurance methods for safety

Objective: Analyze Safety in Collaborative Systems



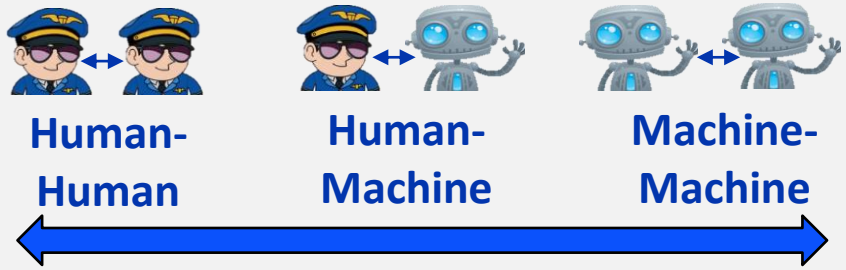
Objective: develop rigorous & systematic framework to analyze safety of collaborative control systems

Contributions:

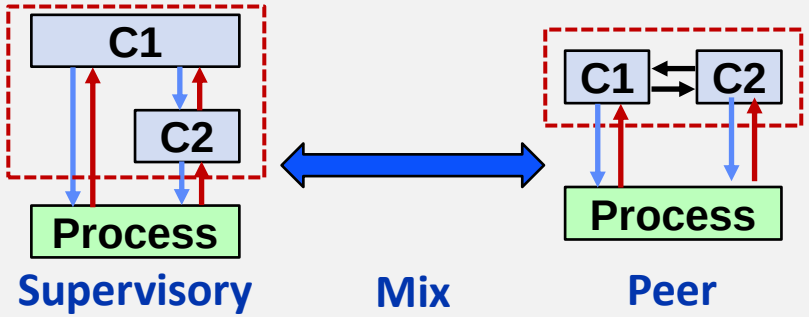
1. **Define** collaborative control interactions using Systems Theory
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Taxonomy of System Interaction Structure

Types of Controllers

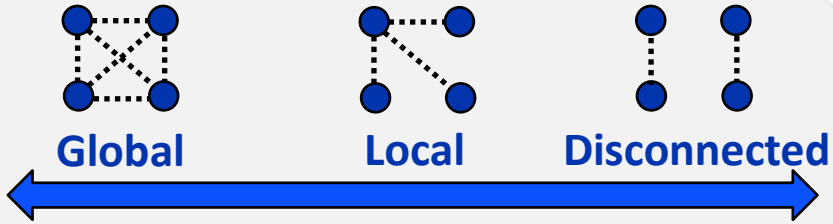


Hierarchal Structure



Behavioral Intent

Connectivity



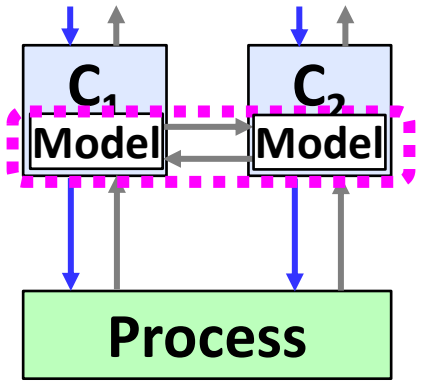
Information Exchange

Roles & Responsibilities

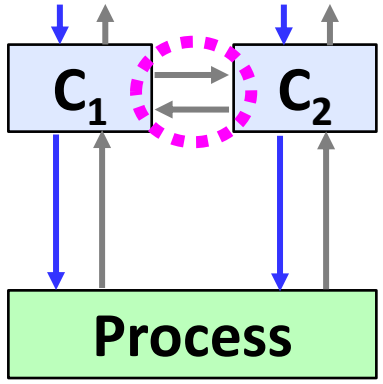
Developmental Origins

Structure influences the dynamics of controller interactions

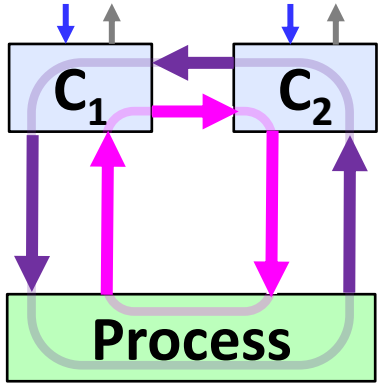
Collaborative Interactions to Address in Hazard Analysis



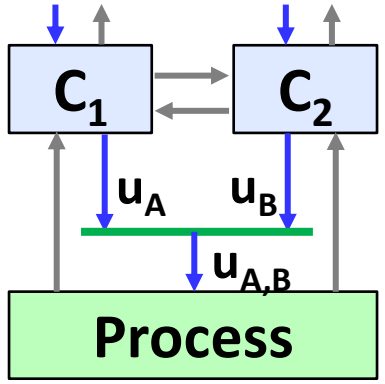
1. Cognitive Alignment



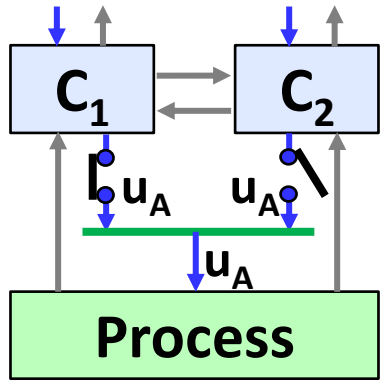
2. Lateral Coordination



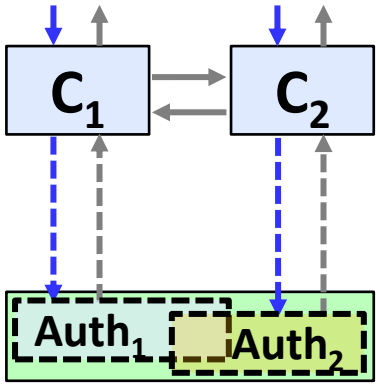
3. Mutually Closing Control Loops



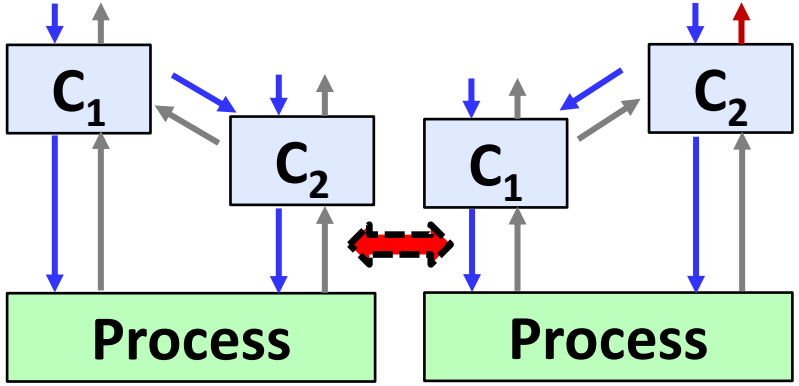
4. Shared Authority



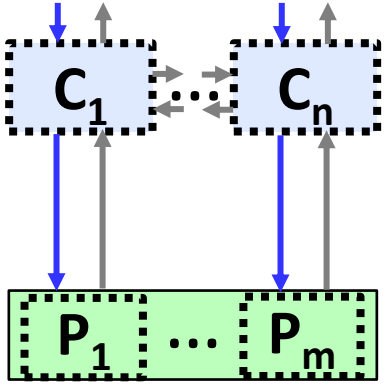
5. Transfer of Authority



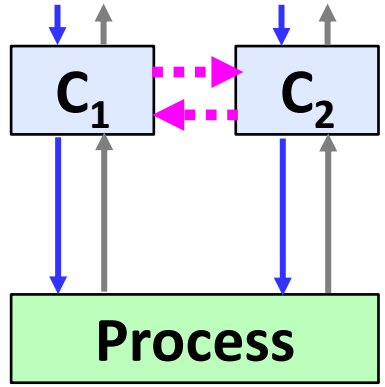
6. Dynamic Authority



7. Dynamic Hierarchy



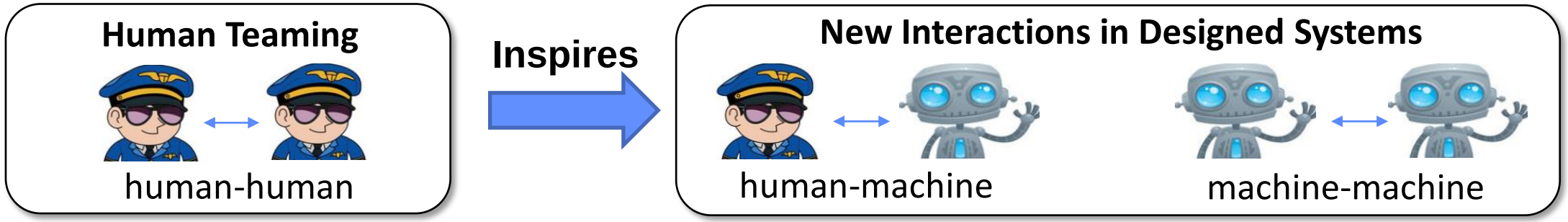
8. Dynamic Membership



9. Dynamic Connectivity

Evaluated framework on 101 interactions in aerospace systems:
Novel concepts seek more of these interactions than fielded systems

Objective: Analyze Safety in Collaborative Systems



Objective: develop rigorous & systematic framework to analyze safety of collaborative control systems

Contributions:

1. **Define** collaborative control interactions using Systems Theory
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Three STPA Extensions for Collaborative Control

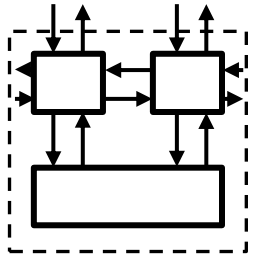
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios

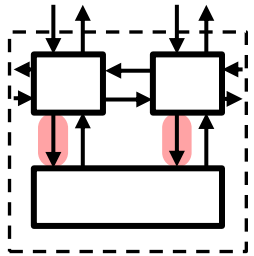
Goal: more systematically address collaborative control interactions in causal analysis

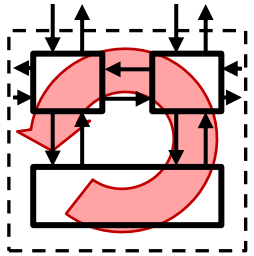
Generic Collaborative Control Structure

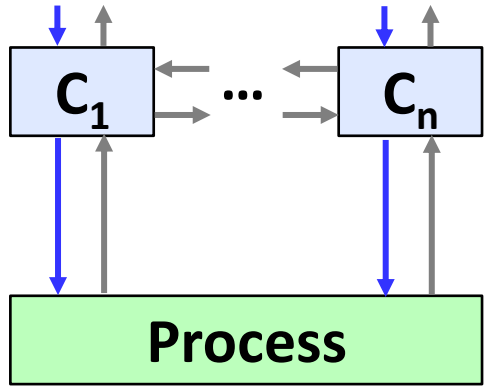
Expand how unsafe control found in collaborative control

Systematic causal scenario ID for collaborative control





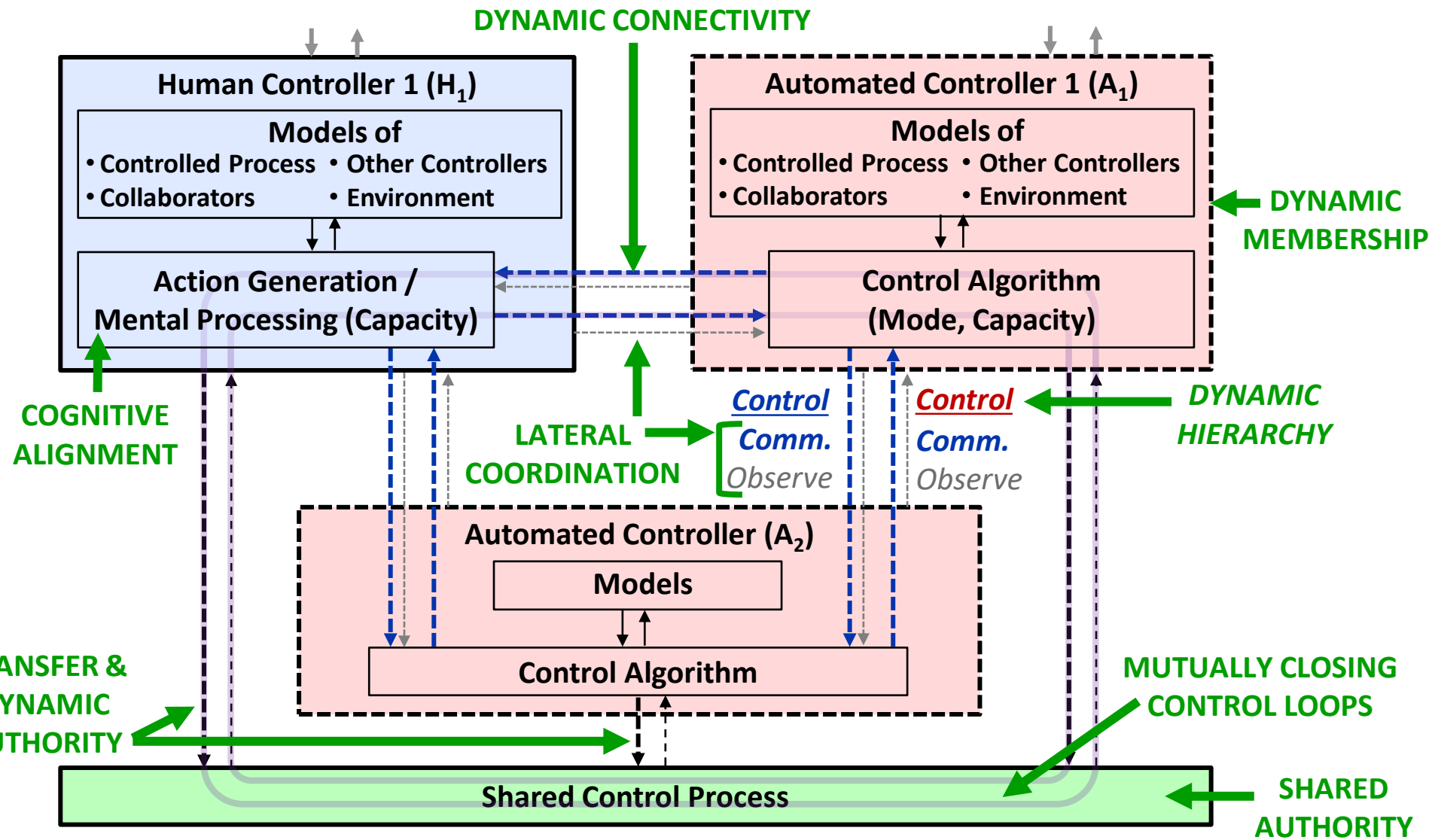




Collaborative Control System

Generic Collaborative Control Structure

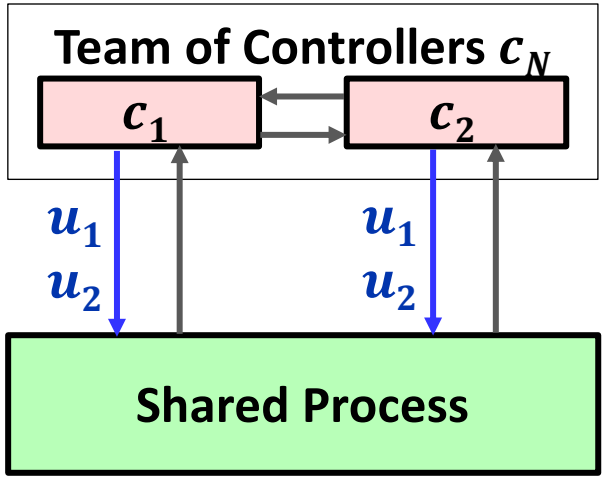
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios



Provides ability to express collaborative control dynamics in control structure

Unsafe Combinations of Control Actions (UCCA)

1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios



STPA Unsafe Control Action (UCA) Structure:

<Controller> **<UCA Type>** <Control Action> <*Context*> *[H]*

4 UCA Types:

1. **Provide**
2. **Not Provide**
3. **Provide Early / Late (start)**
4. **Apply too long / short (stop)**

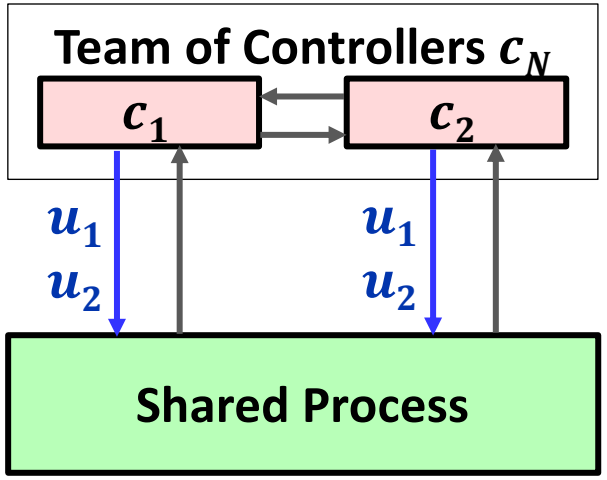
1. c_1 **does not provide** u_1, u_2; c_2 **does not provide** u_1, u_2 when… *[H]*
2. c_1 **does not provide** u_1, u_2; c_2 **does not provide** u_1 and **provides** u_2 when… *[H]*
3. c_1 **does not provide** u_1, u_2; c_2 **provides** u_1 and **does not provide** u_2 when… *[H]*
4. …

Type 1-2 UCCA

#	c_1		c_2		Context
1	$\neg u_1$	$\neg u_2$	$\neg u_1$	$\neg u_2$	
2	$\neg u_1$	$\neg u_2$	$\neg u_1$	u_2	
3	$\neg u_1$	$\neg u_2$	u_1	$\neg u_2$	
…	…	…	…	…	
16	u_1	u_2	u_1	u_2	

Unsafe Combinations of Control Actions (UCCA)

1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios



STPA Unsafe Control Action (UCA) Structure:

<Controller> **<UCA Type>** <Control Action> <*Context*> *[H]*

4 UCA Types:

1. **Provide**
2. **Not Provide**
3. **Provide Early / Late (start)**
4. **Apply too long / short (stop)**

1. c_1 **starts** u_1 before c_2 **starts** u_2 when… *[H]*
2. c_1 **starts** u_1 before c_2 **ends** u_2 when… *[H]*
3. c_1 **ends** u_1 before c_2 **starts** u_2 when… *[H]*
4. …

Type 1-2 UCCA

#	c_1		c_2		Context
1	$\neg u_1$	$\neg u_2$	$\neg u_1$	$\neg u_2$	
2	$\neg u_1$	$\neg u_2$	$\neg u_1$	u_2	
3	$\neg u_1$	$\neg u_2$	u_1	$\neg u_2$	
…	…	…	…	…	
16	u_1	u_2	u_1	u_2	

Type 3-4 UCCA $S(u)$ = Start u, $E(u)$ = End u

#	c_1	before c_2	Context
1	$S(u_1)$	$S(u_2)$	
2	$S(u_1)$	$E(u_2)$	
3	$E(u_1)$	$S(u_2)$	
…	…	…	
8	$E(u_2)$	$E(u_1)$	

Developed algorithm to manage combinatorial growth and automate part of UCCA identification

Causal Scenario Identification Process

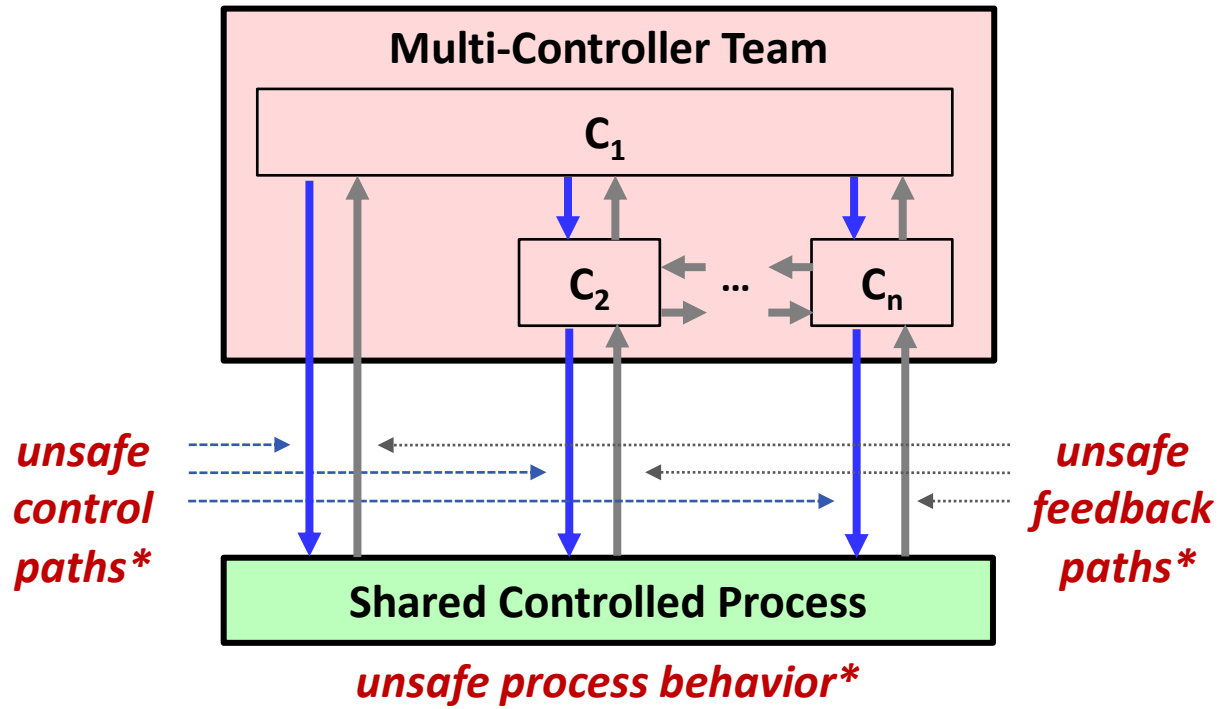
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios

Input: UCCA

1. **Reason about possible internal control**
2. **ID internal control factors**
3. **ID collaborative control factors**
4. **ID other causal factors**

Output: Scenarios *to derive traceable safety constraints*

Focus: ***unsafe (collective) controller behavior***



*Relatively unchanged from STPA

Goal: explain how unsafe combos of control actions can occur

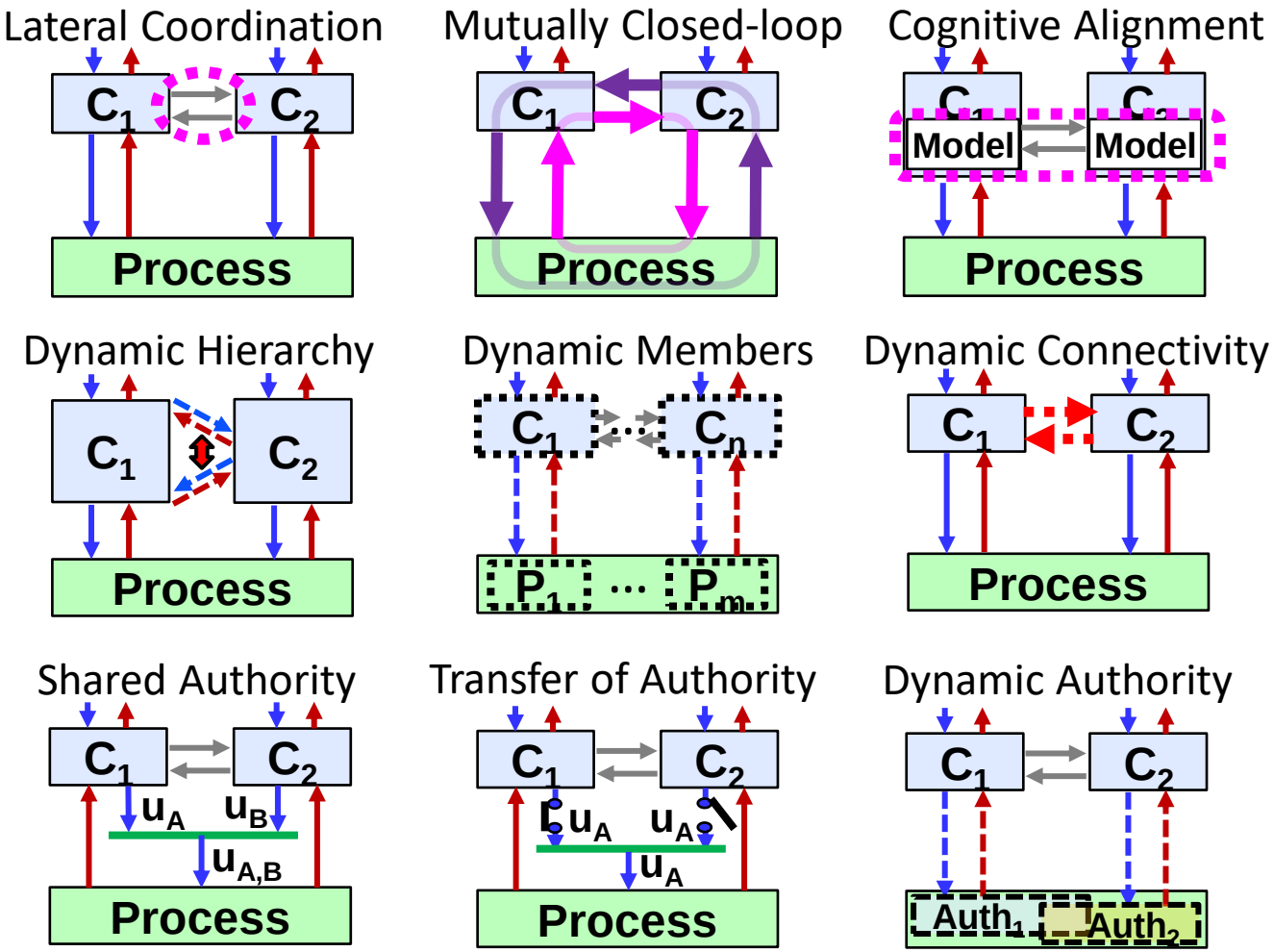
Causal Scenario Identification Process

1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios

Input: UCCA

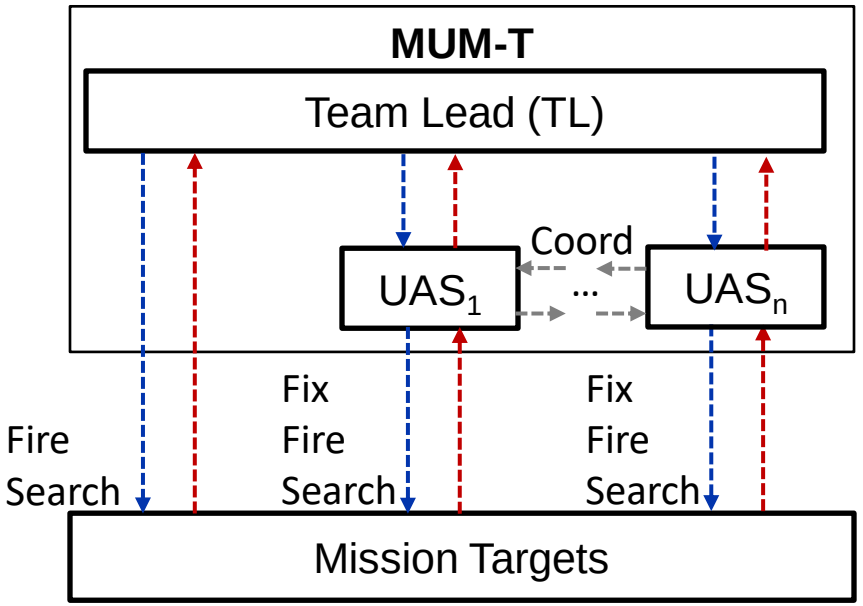
1. **Reason about possible internal control**
2. **ID internal control factors**
3. **ID collaborative control factors**
4. **ID other causal factors**

Output: Scenarios
to derive traceable safety constraints



Goal: explain how unsafe combos of control actions can occur

Case Study: Manned-Unmanned Teaming (MUM-T)



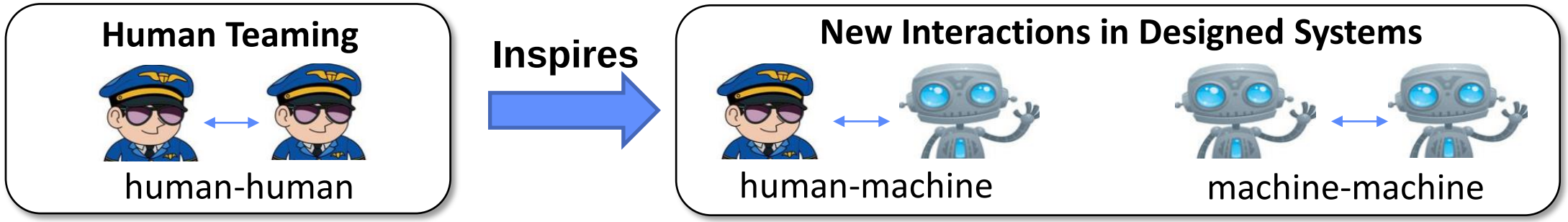
- **Baseline: STPA applied to MUM-T** *[Robertson, 19]*
- **Analyzed same system using extensions**

Causal Factors Found Related to Collaborative Control

Collaborative Control Dynamic	New: Not in Baseline	Found In Baseline
Lateral Coordination	74	33
Cognitive Alignment	29	5
Mutually Closing Control Loops	36	4
Dynamic Membership	25	6
Dynamic Connectivity	13	5
Transfer of Authority (only)	6	7
Dynamic Authority (only)	15	7
Shared Authority (only)	41	23
Total	239	90

Results: extended hazard analysis finds new unsafe controls and causal factors

Objective: Analyze Safety in Collaborative Systems

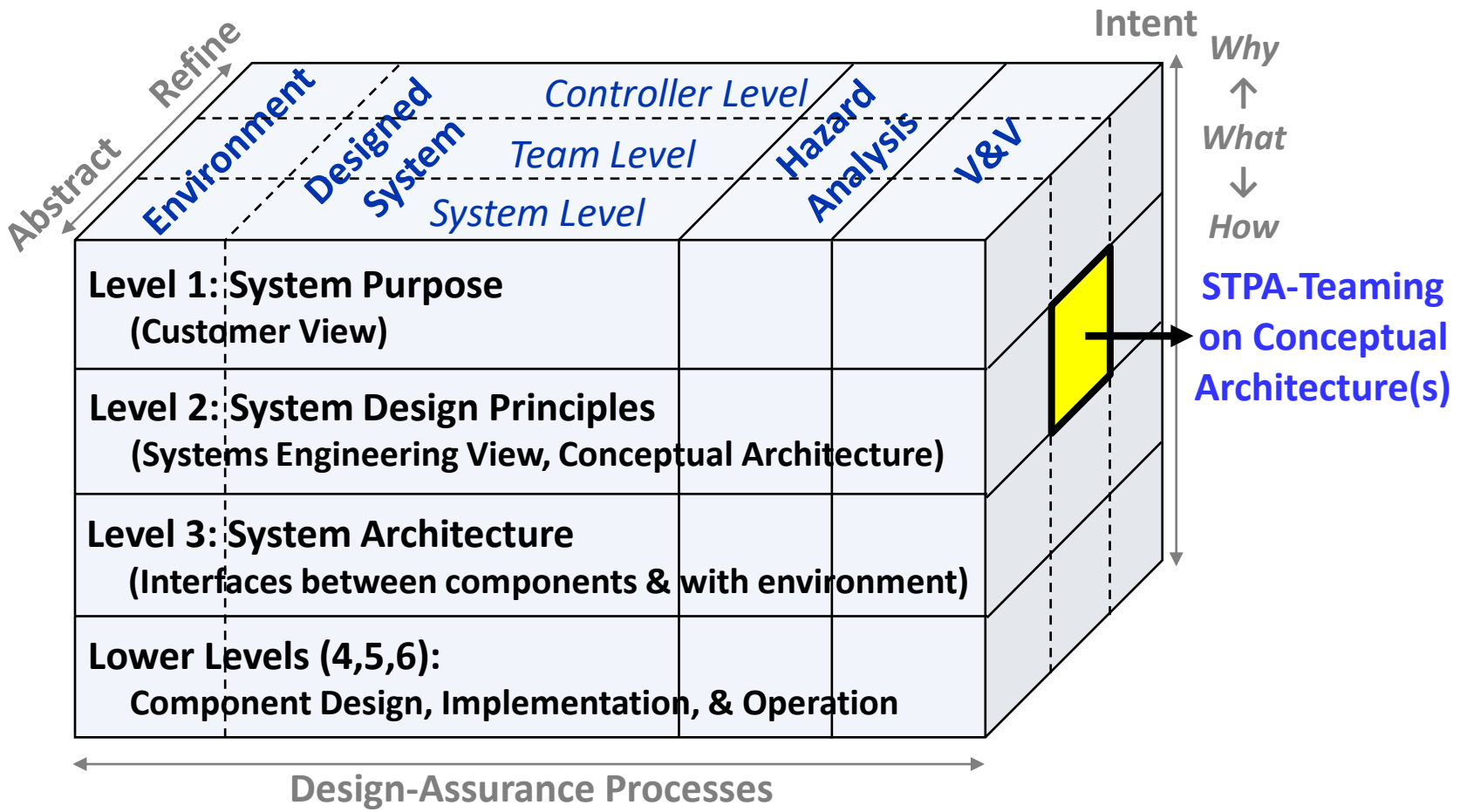


Objective: develop rigorous & systematic framework to analyze safety of collaborative control systems

Contributions:

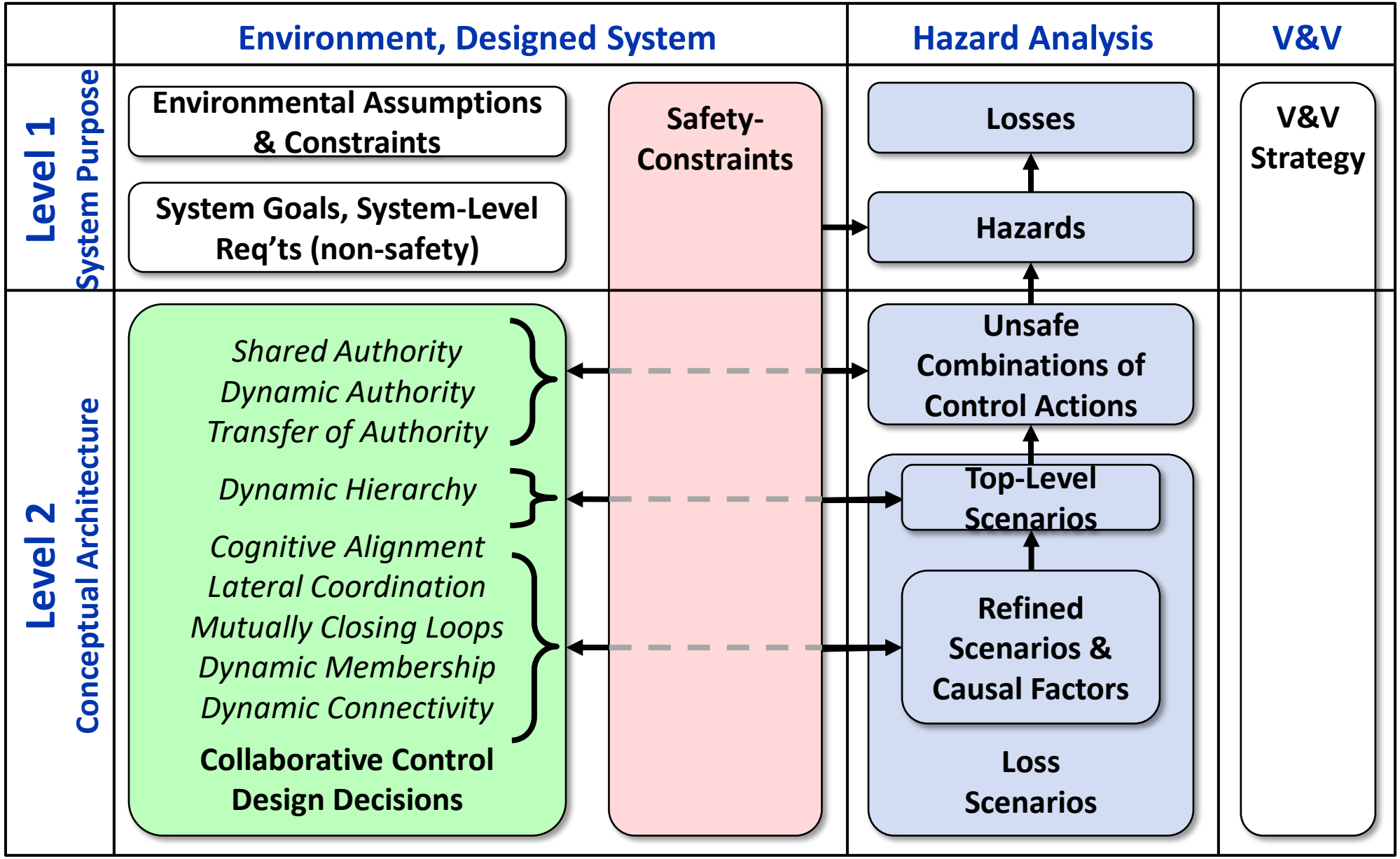
1. **Define** collaborative control interactions using Systems Theory
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Framework for Safety-Guided Design

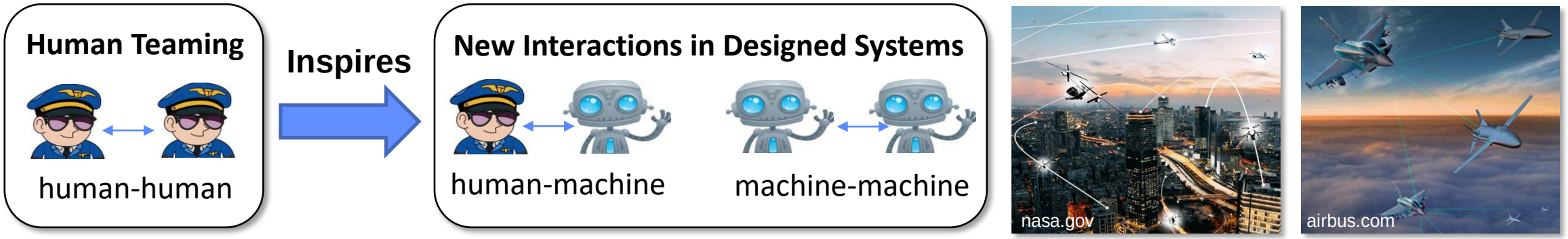


Overall goal: integrate safety-guided design with assurance through enhanced traceability

Traceability of Hazard Analysis Results to Design Decisions



Summary



Seek to engineer systems with complex team-inspired interactions

Beyond current modeling, analysis, design, and assurance methods

Objective: rigorous & systematic framework to analyze safety of collaborative systems

1. **Define** collaborative interactions using Systems Theory
2. **Extend** STAMP/STPA for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Kopeikin, Leveson, & Neogi 2023 (Prepub INCOSE IS)

PhD Dissertation to Follow (2023)

Andrew Kopeikin: kopeikin@mit.edu