33rd Annual INCOSE international symposium

hybrid event

Honolulu, HI, USA
July 15 - 20, 2023

Defining Collaborative Control Interactions Using Systems Theory

Andrew Kopeikin (MIT, MIT Lincoln Lab, USAF Reserve, CFII),
Prof Nancy Leveson (MIT), Dr. Natasha Neogi (NASA)

Human Team vs Human-Machine Interactions

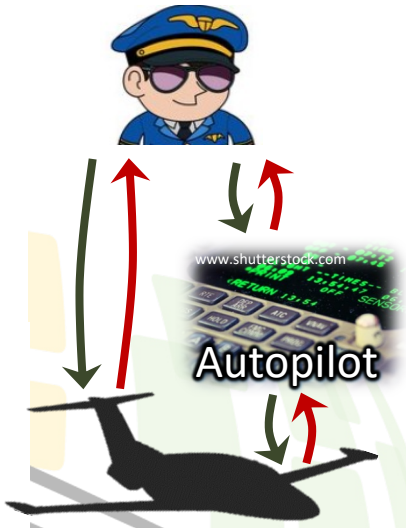
Interactions in current human-automation systems are simpler

Human as Supervisor

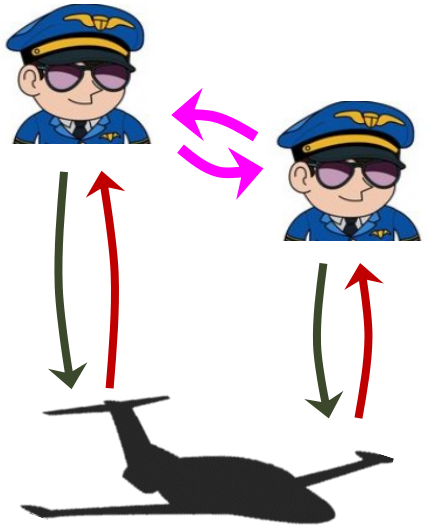
- sets control goal
- supervises
- intervenes

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

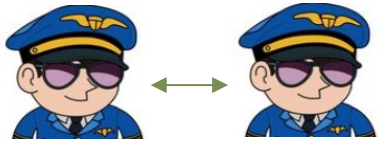
Seek to engineer systems with complex team-inspired interactions

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

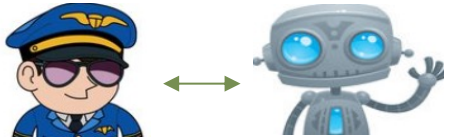
Human Teaming



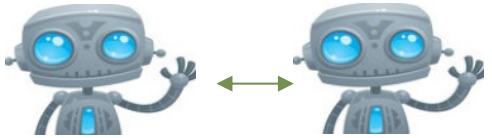
human-human

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 Control Systems

Rigorous & systematic framework to analyze safety & guide design:

Focus of Paper

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

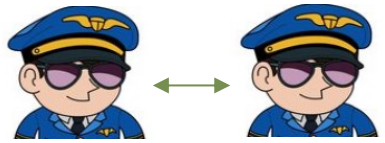
Objective: Analyze Safety in Collaborative Control Systems

Rigorous & systematic framework to analyze safety & guide design:

1. **Define** collaborative control interactions using Systems Theory
   - Relevant Literature
   - Framework: Taxonomy & Collaborative Dynamics
   - Analysis of systems using framework
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Theoretical Foundations of Teaming

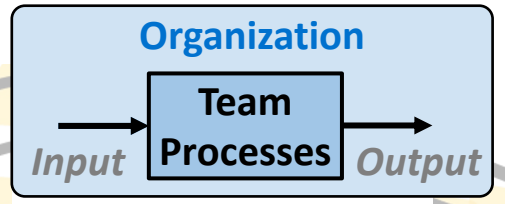
Human Teams



Many Team Models *[Salas '05]*

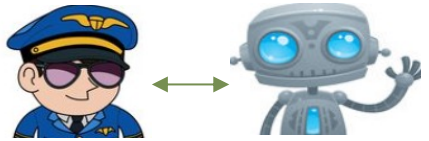
Shared & Distributed Cognition
[Endsley '99], [Stanton '06]

System-Theoretic View



[Paris '00], [Ilgen '99]

Human Machine Teams



Modeled after Human Teams

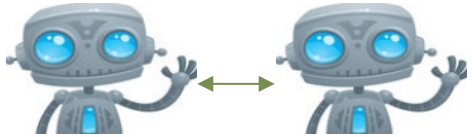
"Teammate" vs "Tool" *[Mosier '17]*

Human Machine Asymmetry
[Pritchett '18], [Klein '04]

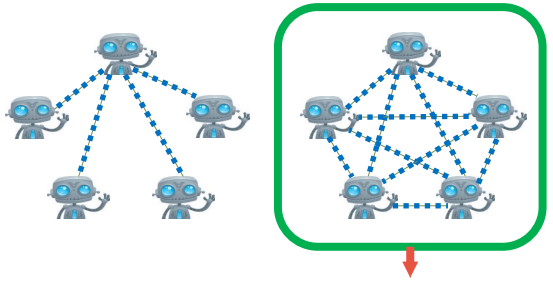
Trust *[Chancey '21]*

Human in the Loop *[Endsley '17]*

Machine Teams



Centralized vs Distributed



Attributes relate to teaming

Often overlook human

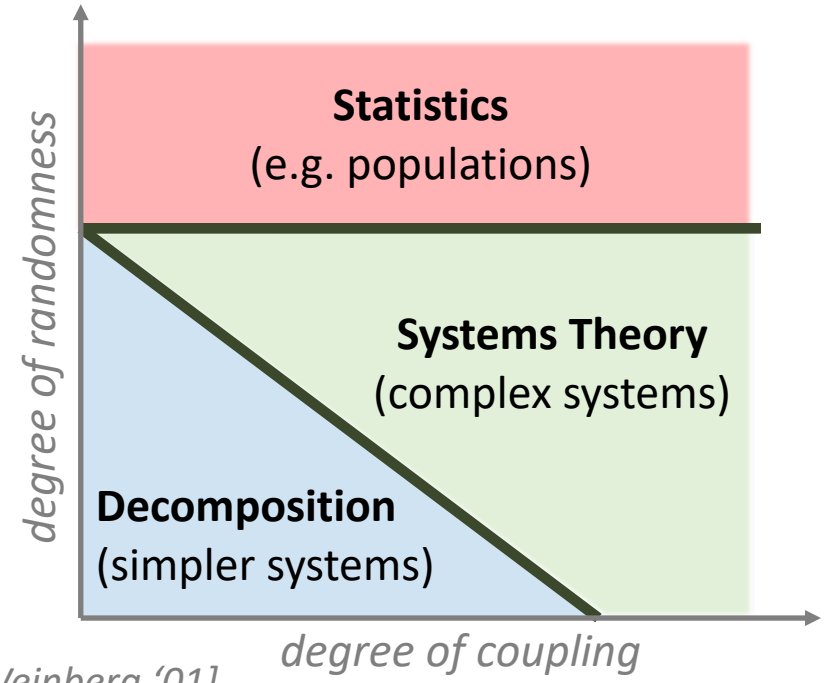
Key Takeaways: useful for control dynamics, but…

1. **Focus on performance vs safety**
2. **Lack guidance for analysis & design**
3. **Little consideration for larger socio-technical system**

Systems Theory

System: components act as whole to achieve common goal *[Leveson '13, '20]*

Emergent Properties | **"Open"** | **Recursive**



[Weinberg '01]

Two Key Principles *[Checkland '99]*

1. **Emergence & Hierarchy**
2. **Communication & Control**

Systems Theory augments where decomposition distorts analysis of behavior *[Leveson '21]*

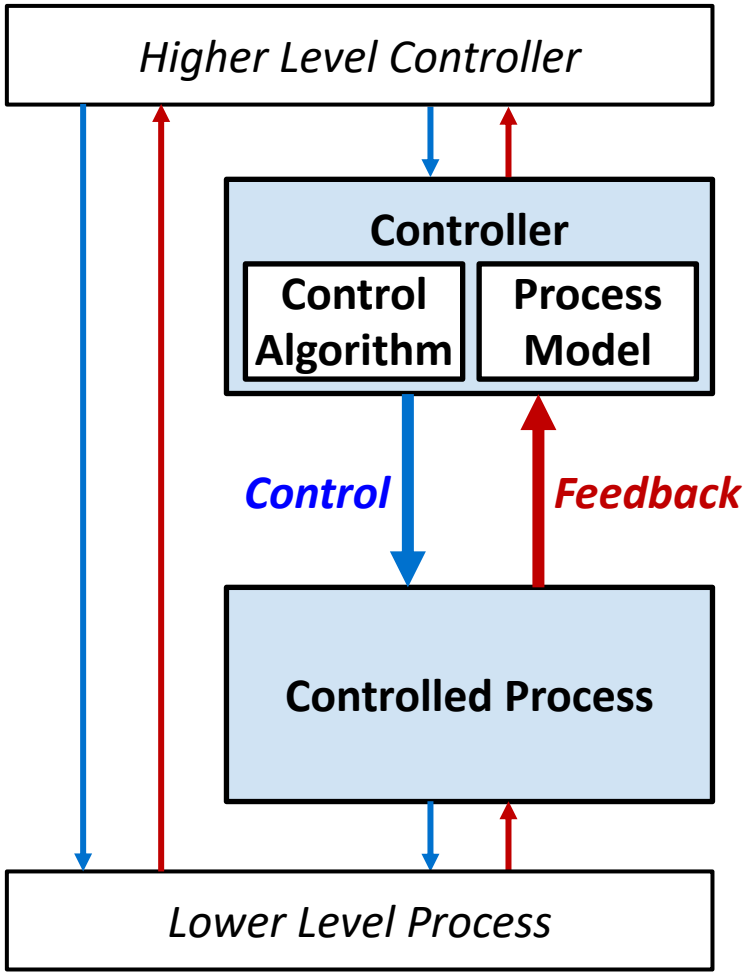
System Theoretic Accident Model & Processes (STAMP)

Accident Causality Model *[Leveson '11]*

- Grounded in Systems Theory
- Safety as control problem (vs reliability)
- Unsafe behaviors & interactions → Accidents
- Basis of analysis tools (ex: STPA*)

Well Suited for Teaming

- Non-linear causality: mutual component influence
- Interactions: hardware, software, humans
- Models complex socio-technical systems



But, more complex interactions (e.g., those in collaborative control) not fully defined in STAMP

**STPA: System Theoretic Process Analysis*

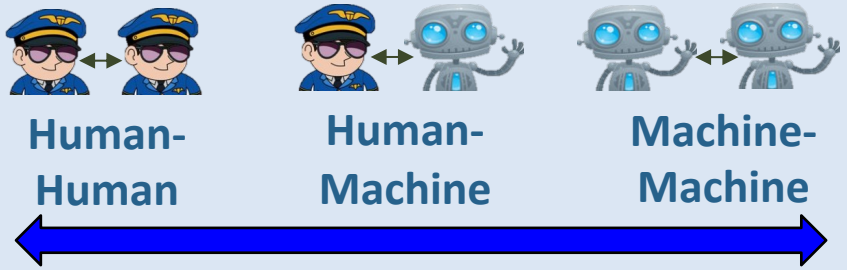
Objective: Analyze Safety in Collaborative Control Systems

Rigorous & systematic framework to analyze safety & guide design:

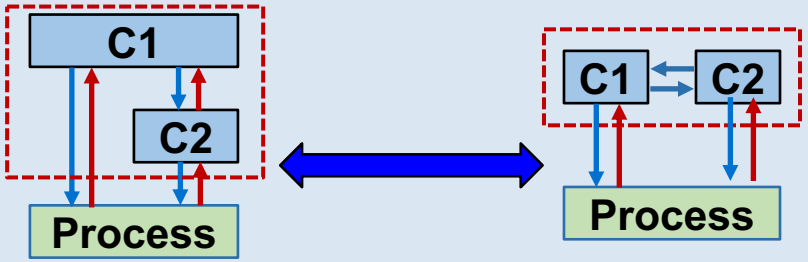
1. **Define** collaborative control interactions using Systems Theory
   - Relevant Literature
   - Framework: Taxonomy & Collaborative Dynamics
   - Analysis of systems using framework
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Taxonomy of System Interaction Structure

Types of Controllers

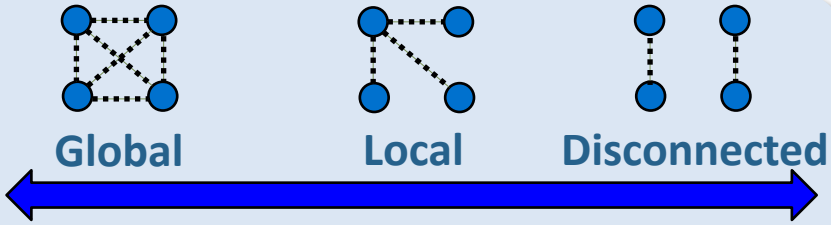


Hierarchal Structure



Behavioral Intent

Connectivity



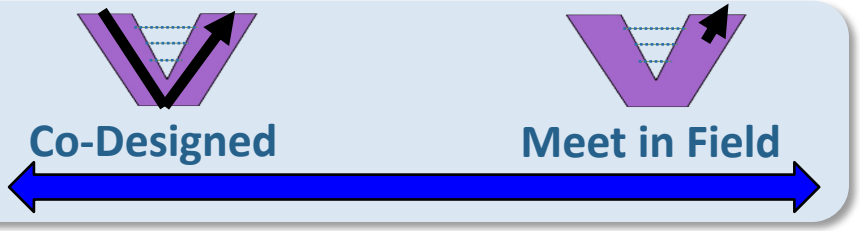
Information Exchange

Active (Messaging) — Passive (Observation) — None (Prediction)

Roles & Responsibilities

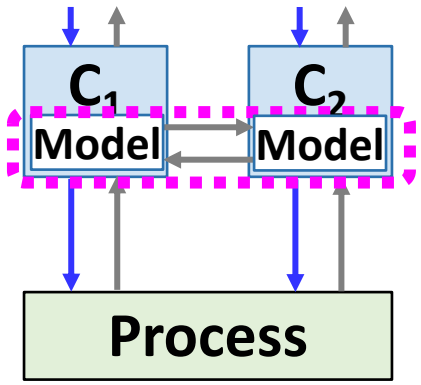
Prescribed — Dynamic — Ad-hoc

Developmental Origins

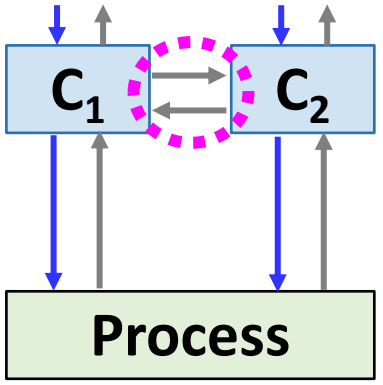


Structure influences the dynamics of controller interactions

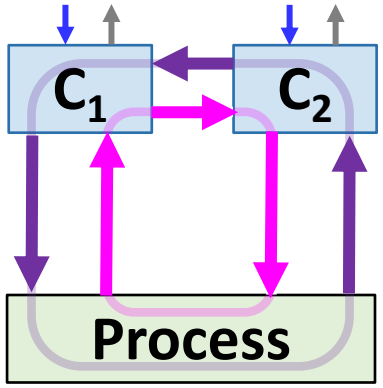
Collaborative Interactions to Address in Hazard Analysis



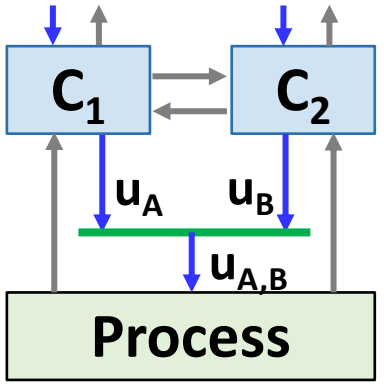
1. Cognitive Alignment



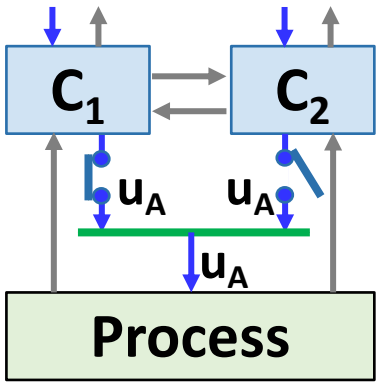
2. Lateral Coordination



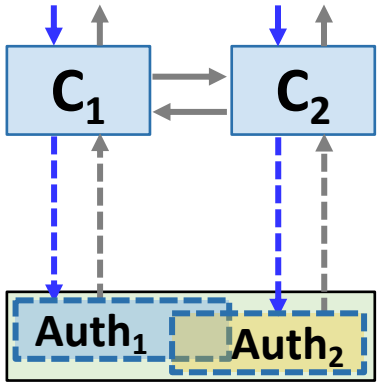
3. Mutually Closing Control Loops



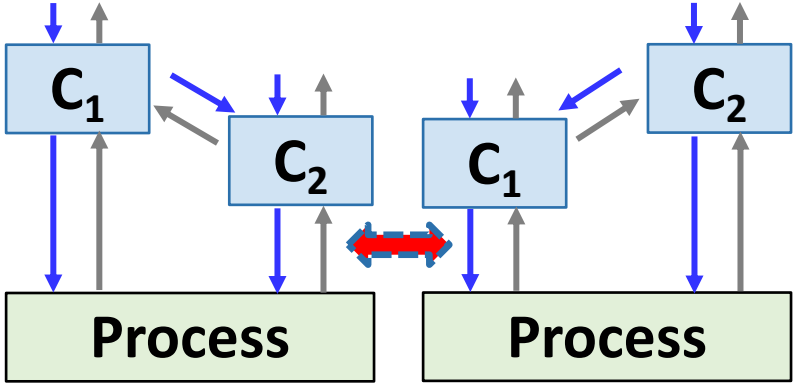
4. Shared Authority



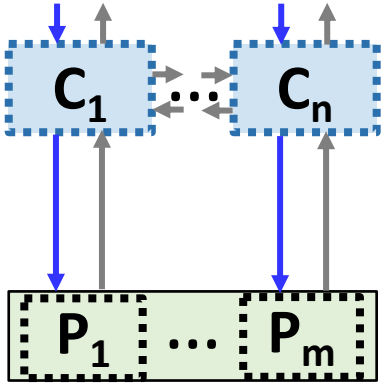
5. Transfer of Authority



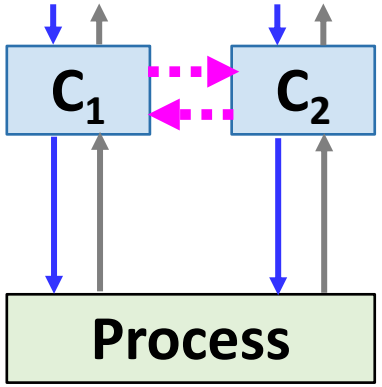
6. Dynamic Authority



7. Dynamic Hierarchy



8. Dynamic Membership



9. Dynamic Connectivity

Definitions grounded in Systems Theory & 3 parts* of STAMP

3 parts of STAMP: hierarchal control structures, process models, safety constraints

Objective: Analyze Safety in Collaborative Control Systems

Rigorous & systematic framework to analyze safety & guide design:

1. **Define** collaborative control interactions using Systems Theory
   - Relevant Literature
   - Framework: Taxonomy & Collaborative Dynamics
   - Analysis of systems using framework
2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

Categorized 101 Interactions from Aerospace Literature

Human-Machine & Multi-Machine (Not Fielded)

Human-Human (Not Fielded)

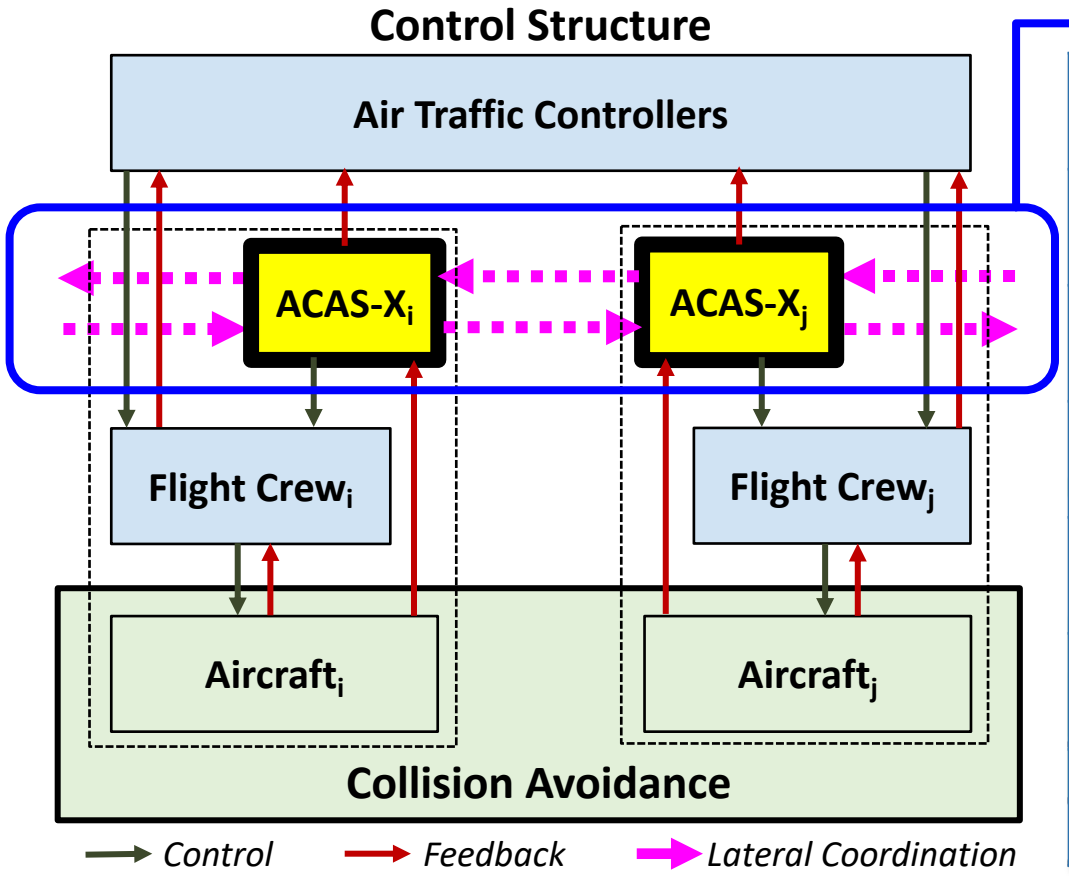
Human-Machine & Multi-Machine (Fielded)

Human-Human (Fielded)

Examples of systems included in this study

Categorized 101 Interactions from Aerospace Literature

Ex: ACAS-X Aircraft to ACAS-X Aircraft Interactions



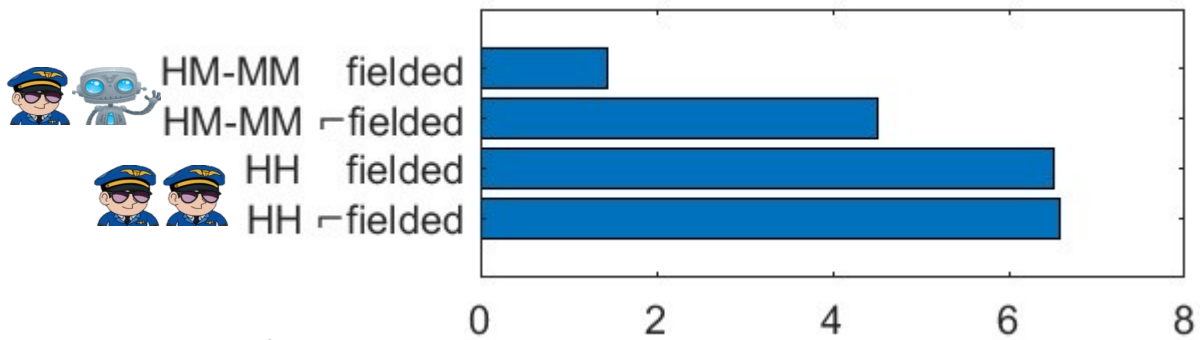
Structure of Interactions

Types of Controllers	Human — Human-Machine — Machine
Hierarchal Structure	Master-Slave — Mix — Peer
Behavioral Intent	Cooperative — Mix — Adversarial
Connectivity	Global — Local — Disconnect
Information Exchange	Active — Passive — None
Roles & Responsiblts	Prescribed — Dynamic — Ad-hoc
Development Origins	Co-Design — Meet-in-Field

Collaborative Dynamics

1. Lateral Coordination	✓
2. Mutually Close Loops	✓
3. Cognitive Alignment	✓
4. Shared Authority	✓
5. Transfer of Authority	✗
6. Dynamic Authority	✓
7. Dynamic Hierarchy	✗
8. Dynamic Membership	✓
9. Dynamic Connectivity	✓

Presence of Collaborative Dynamics in Analyzed Systems

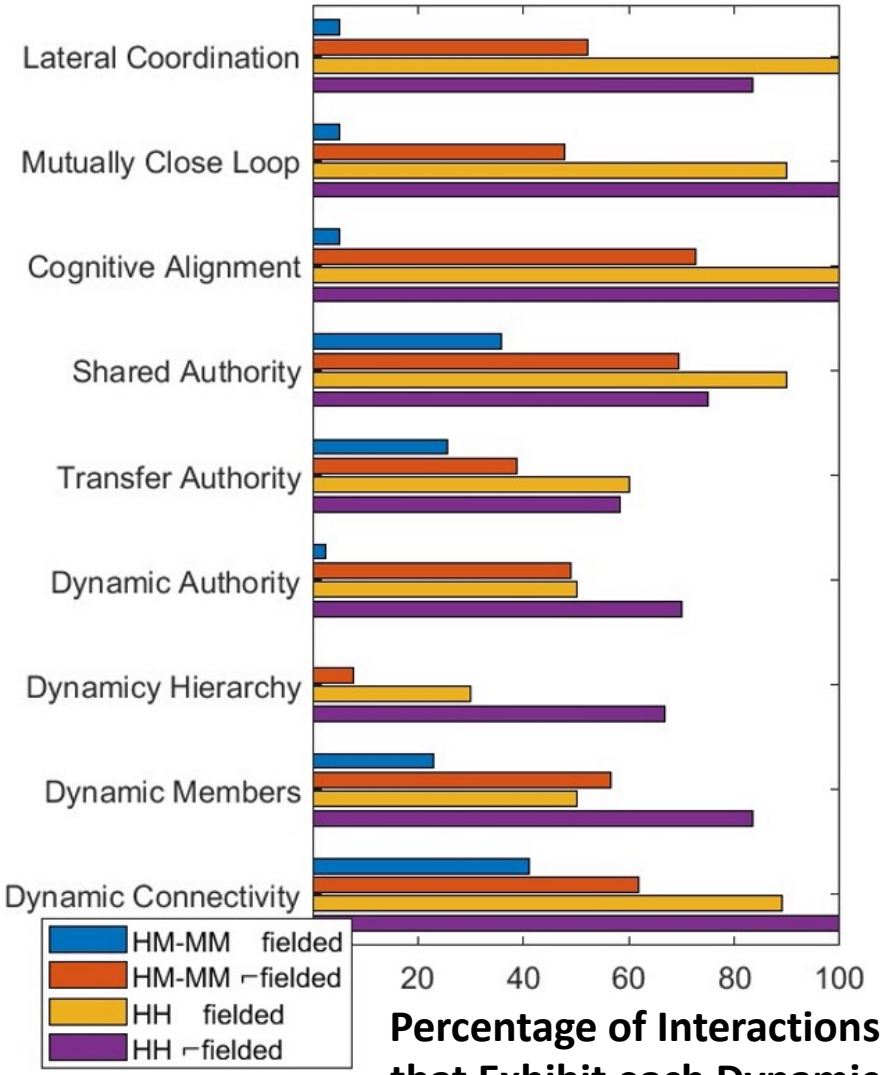


Mean # of Collaborative Control Dynamics Per Interaction

2 Important Takeaways:

1. **Systems are being designed with these collaborative control dynamics**
2. **In sample: systems not yet fielded exhibit more complex interactions those fielded**

Not a quantitative analysis representative of all systems



Percentage of Interactions that Exhibit each Dynamic

Objective: Analyze Safety in Collaborative Control Systems

Rigorous & systematic framework to analyze safety & guide design:

Focus of Paper ➡

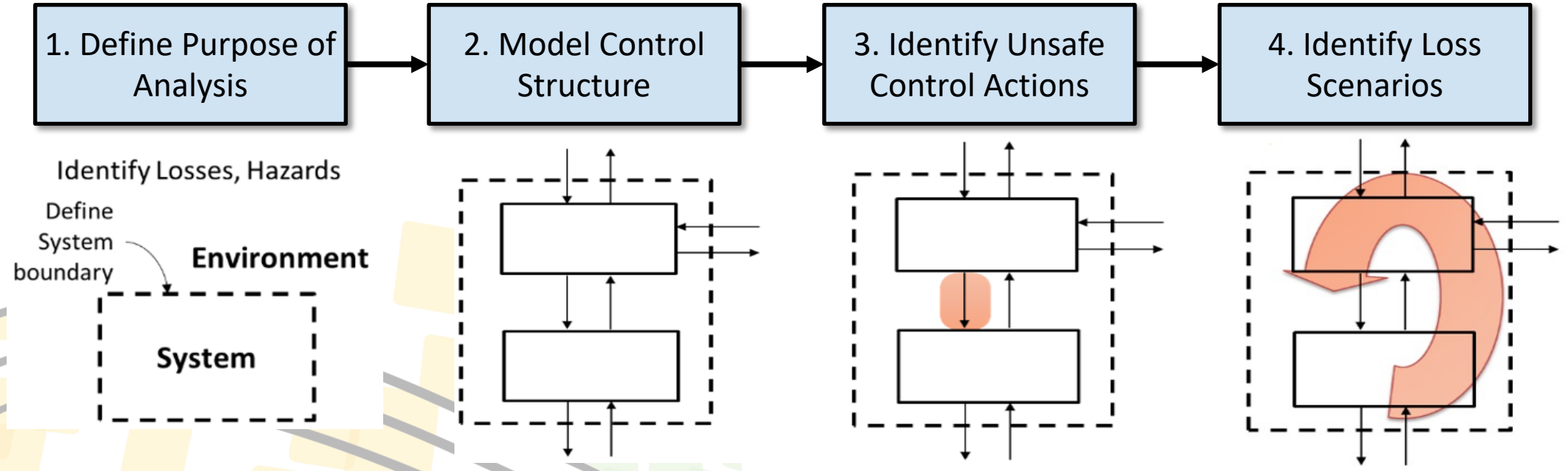
1. **Define** collaborative control interactions using Systems Theory
   - Relevant Literature
   - Framework: Taxonomy & Collaborative Dynamics
   - Analysis of systems using framework

Preview ➡

2. **Extend** state-of-art in hazard analysis for collaborative interactions
3. **Integrate** safety-guided design & assurance processes

System-Theoretic Process Analysis (STPA)

[Leveson & Thomas '18]



STPA: analysis method built on STAMP gaining popularity in many industries

System-Theoretic Process Analysis (STPA)

Hazard Analysis tool with STAMP as theoretical foundation

[Leveson & Thomas '18]

UCA Structure:

<Controller> <UCA Type> <Control Action> <*Context*>

Operator provides too late "Approve COA" *when* *COA no longer fulfills mission objectives*

Example from: *[Johnson, Kopeikin, Leveson '21]*

Find breakdowns in control

Fleet Controller **updates** COA too **frequently**.
Operator in cycle of **perpetual COA review**.

Design Requirement(s)

STPA powerful but needs more guidance to systematically handle collaborative interactions

STPA Extensions for Collaborative Control

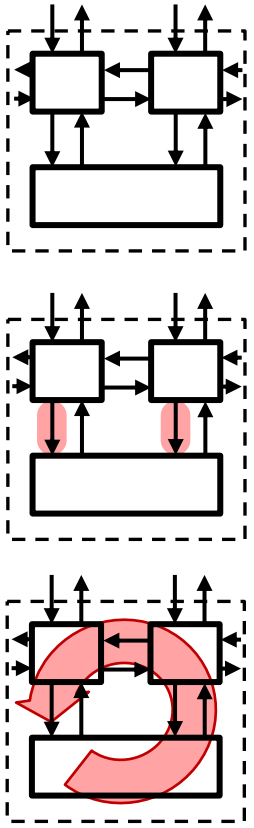
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios

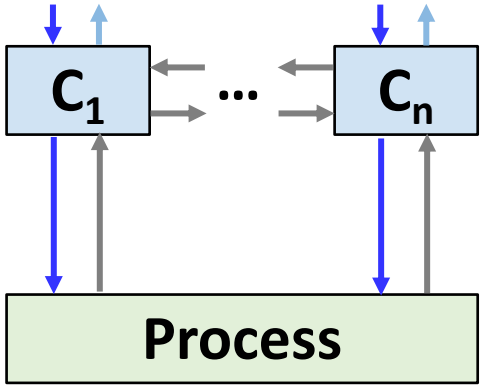
Goal: more systematically address collaborative control interactions in causal analysis

Generic Collaborative Control Structure

Expand how unsafe control found in collaborative control

Systematic causal scenario ID for collaborative control

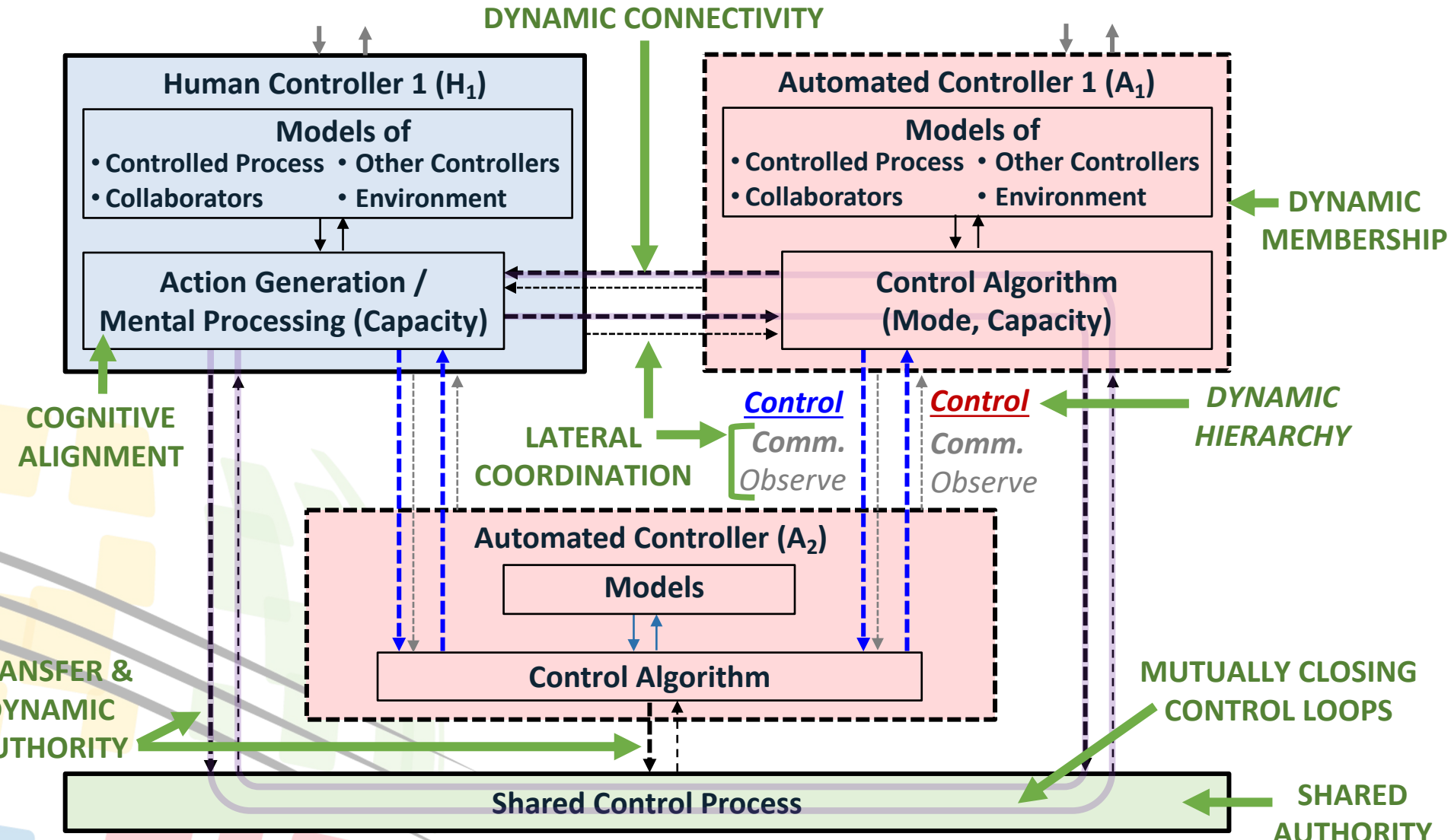




Collaborative Control System

Generic Collaborative Control Structure

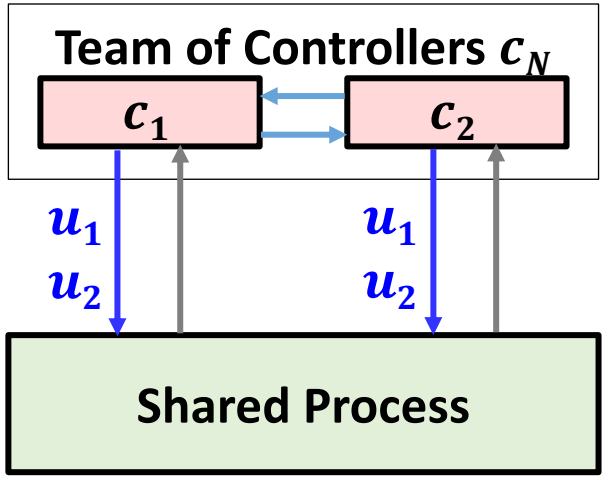
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios



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**
2. c_1 **does not provide**
3. c_1 **does not provide**
4. …

u_1, u_2; c_2 **does not provide** u_1, u_2 when… *[H]*

u_1, u_2; c_2 **does not provide** u_1 **and provides** u_2 when… *[H]*

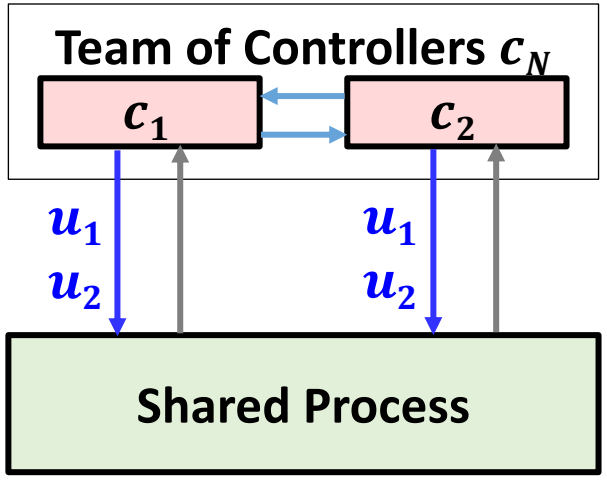
u_1, u_2; c_2 **provides** u_1 **and does not provide** u_2 when… *[H]*

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

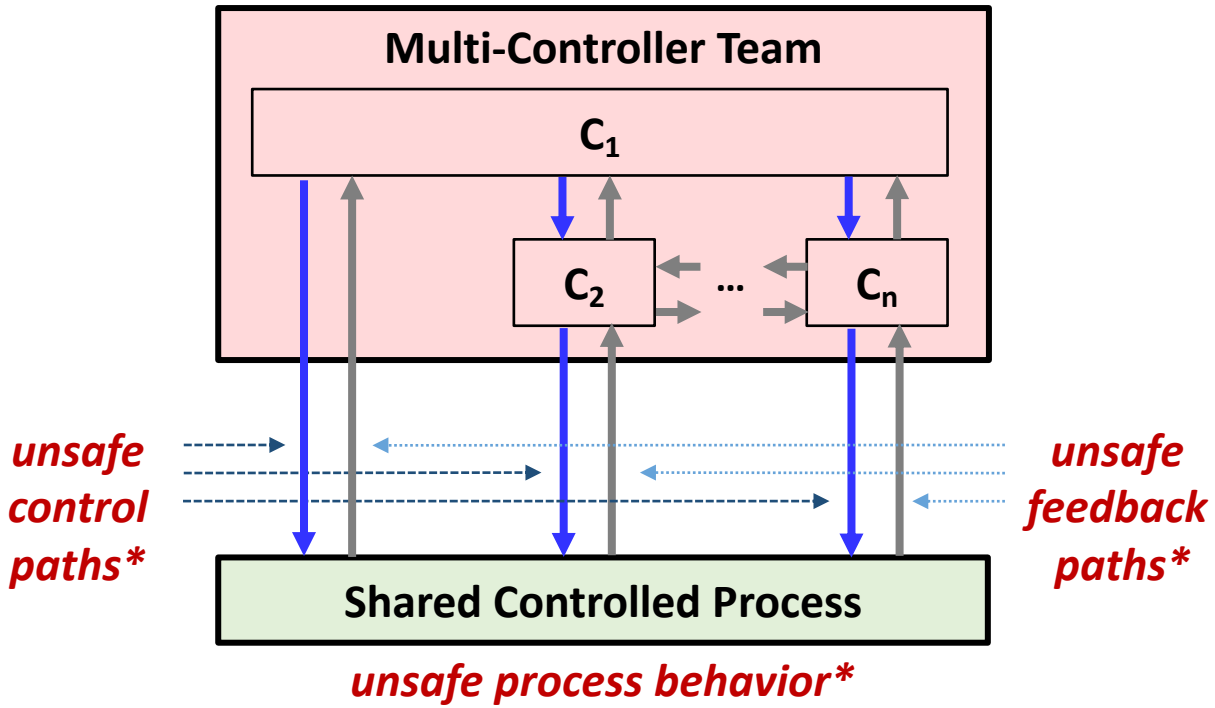
1. Define Purpose of Analysis
2. Model Control Structure
3. Identify Unsafe Control Actions
4. Identify Loss Scenarios

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

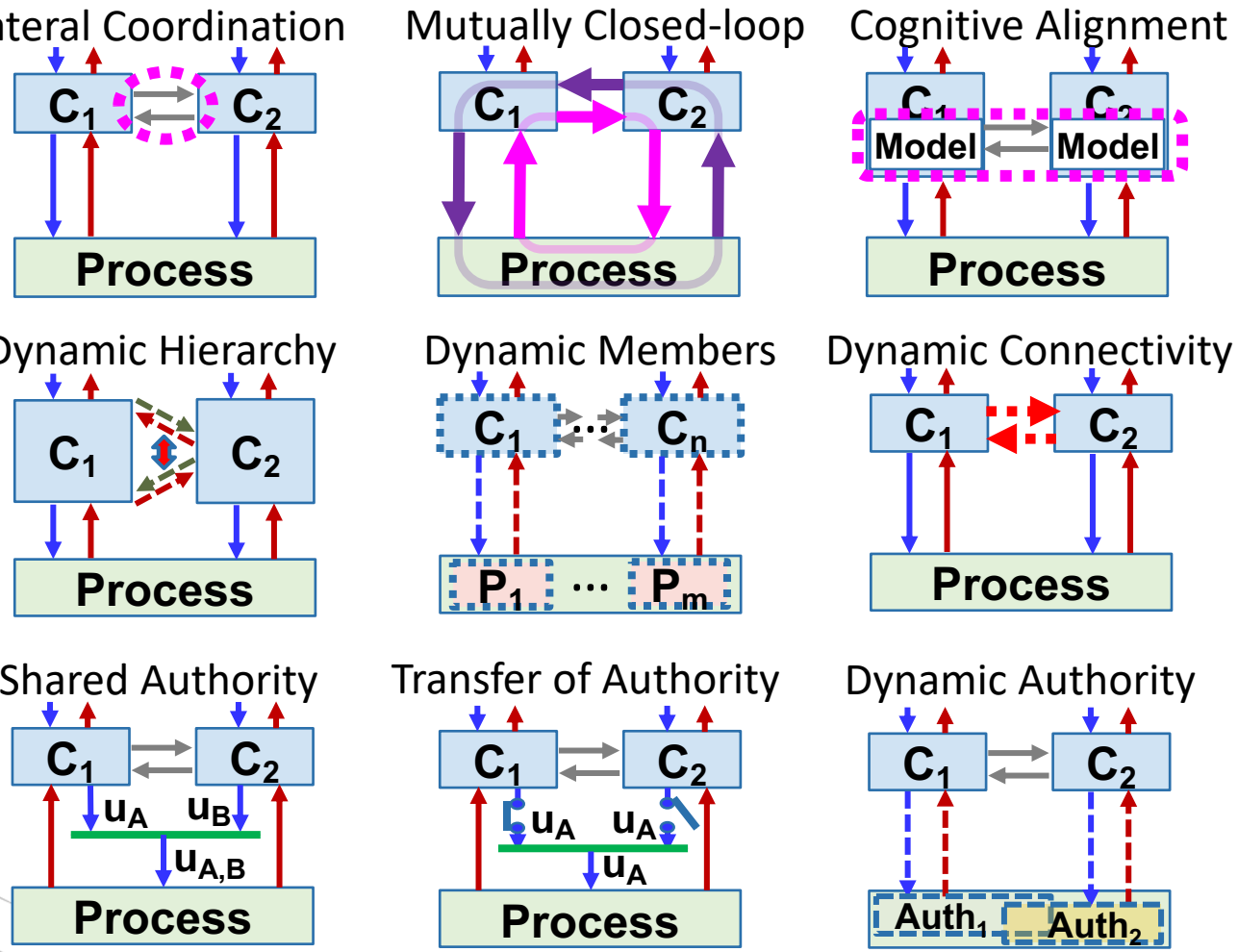
Causal Scenario Identification Process

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

Objective: Analyze Safety in Collaborative Control Systems

Rigorous & systematic framework to analyze safety & guide design:

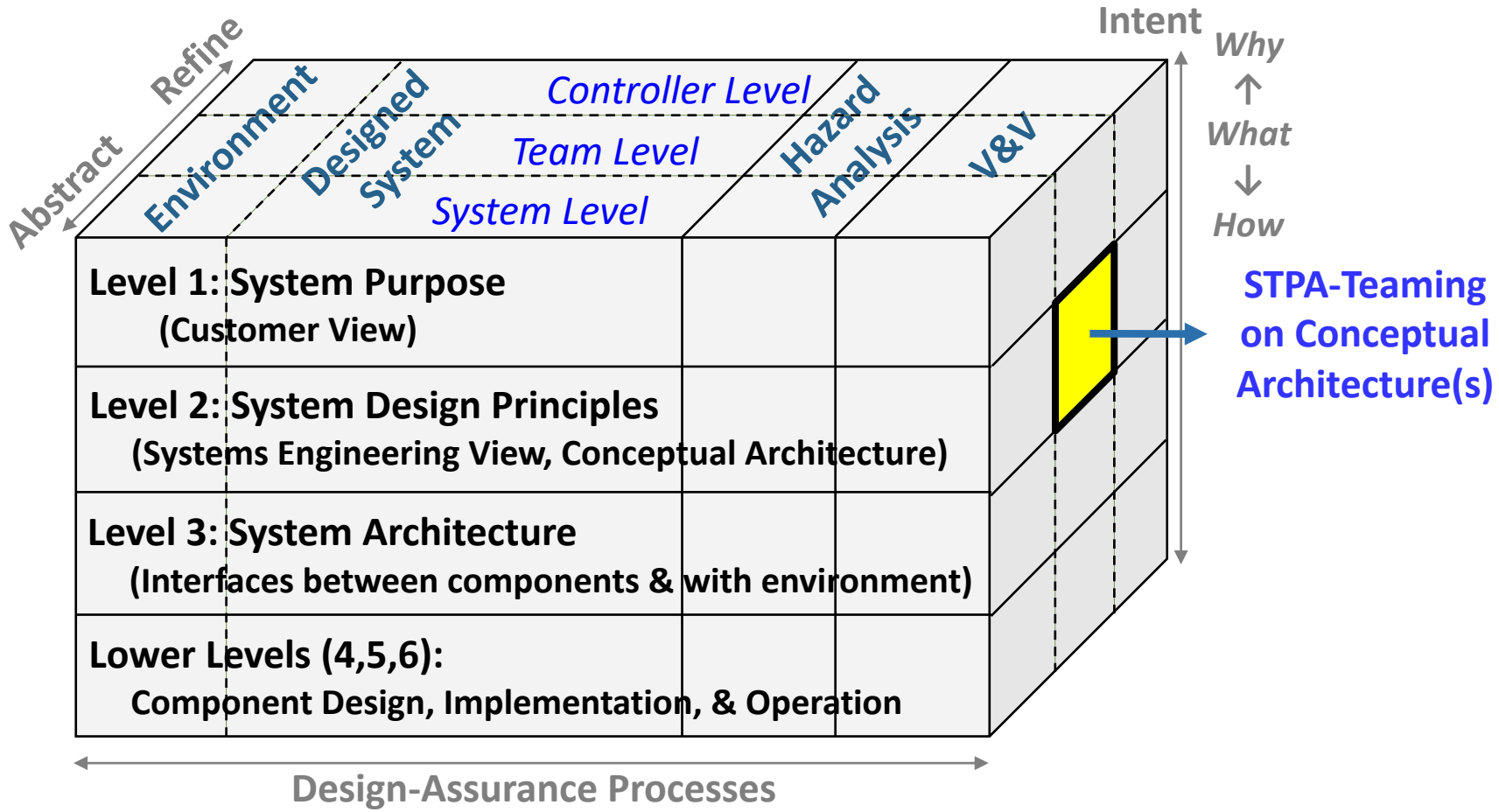
Focus of Paper ➡

1. **Define** collaborative control interactions using Systems Theory
   - Relevant Literature
   - Framework: Taxonomy & Collaborative Dynamics
   - Analysis of systems using framework
2. **Extend** state-of-art in hazard analysis for collaborative interactions

Preview ➡

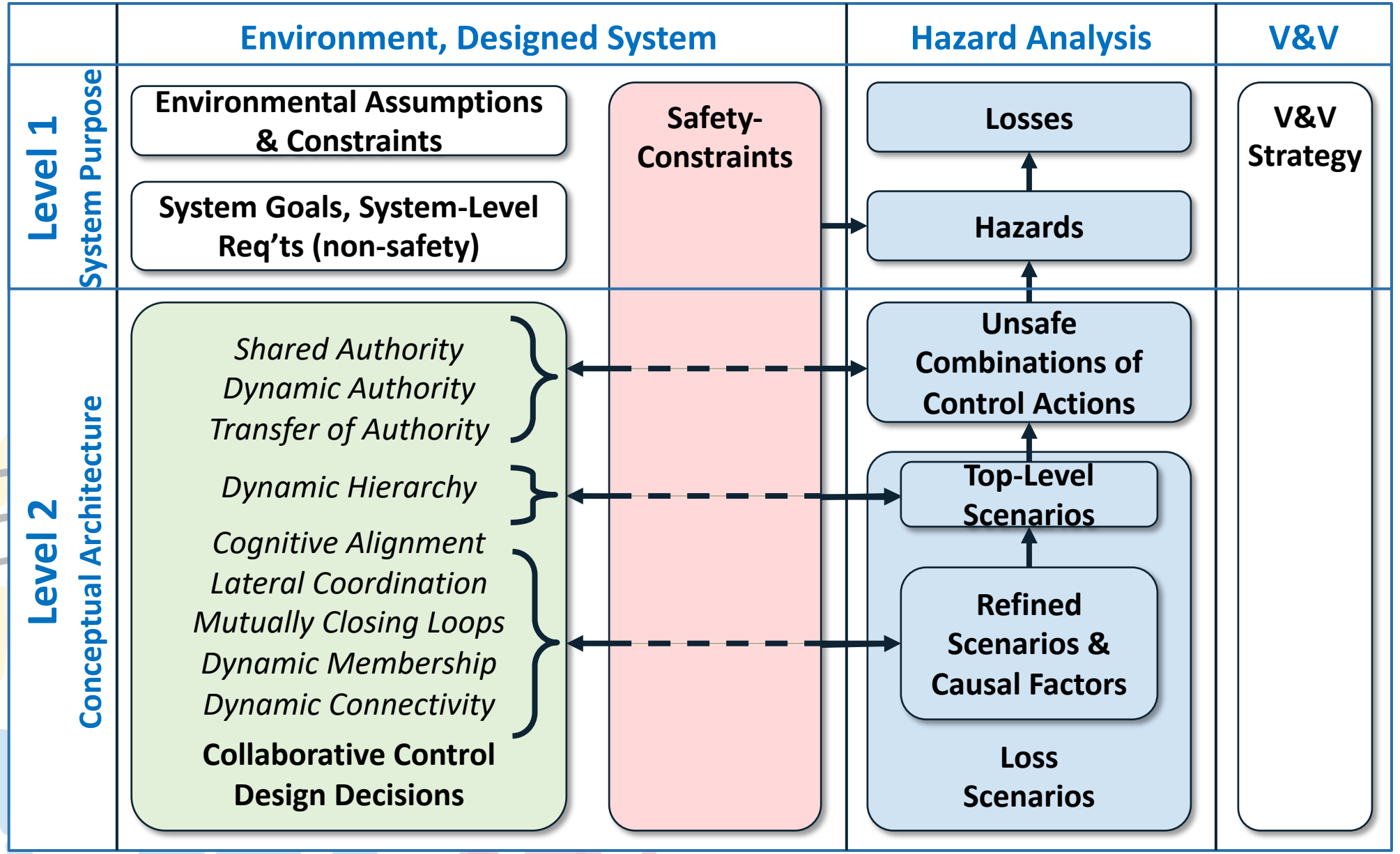
3. **Integrate** safety-guided design & assurance processes

Framework for Safety-Guided Design

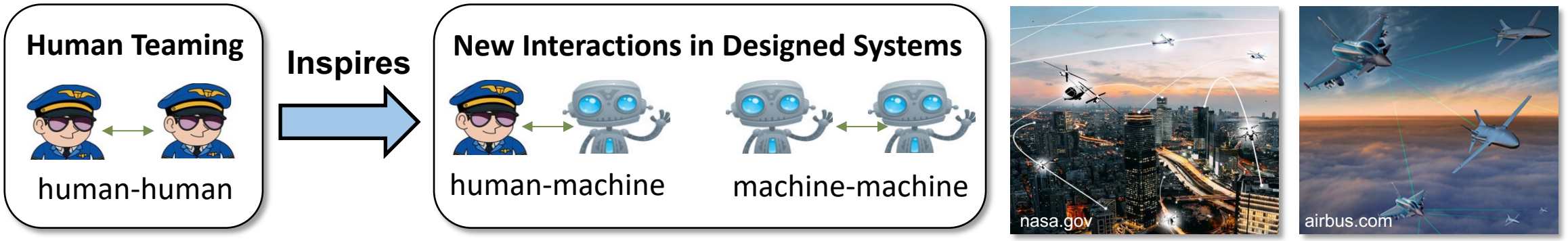


Overall goal: integrate safety-guided design with assurance through enhanced traceability

Traceability of Hazard Analysis Results to Design Decisions



Summary



Seek to engineer systems with complex team-inspired interactions

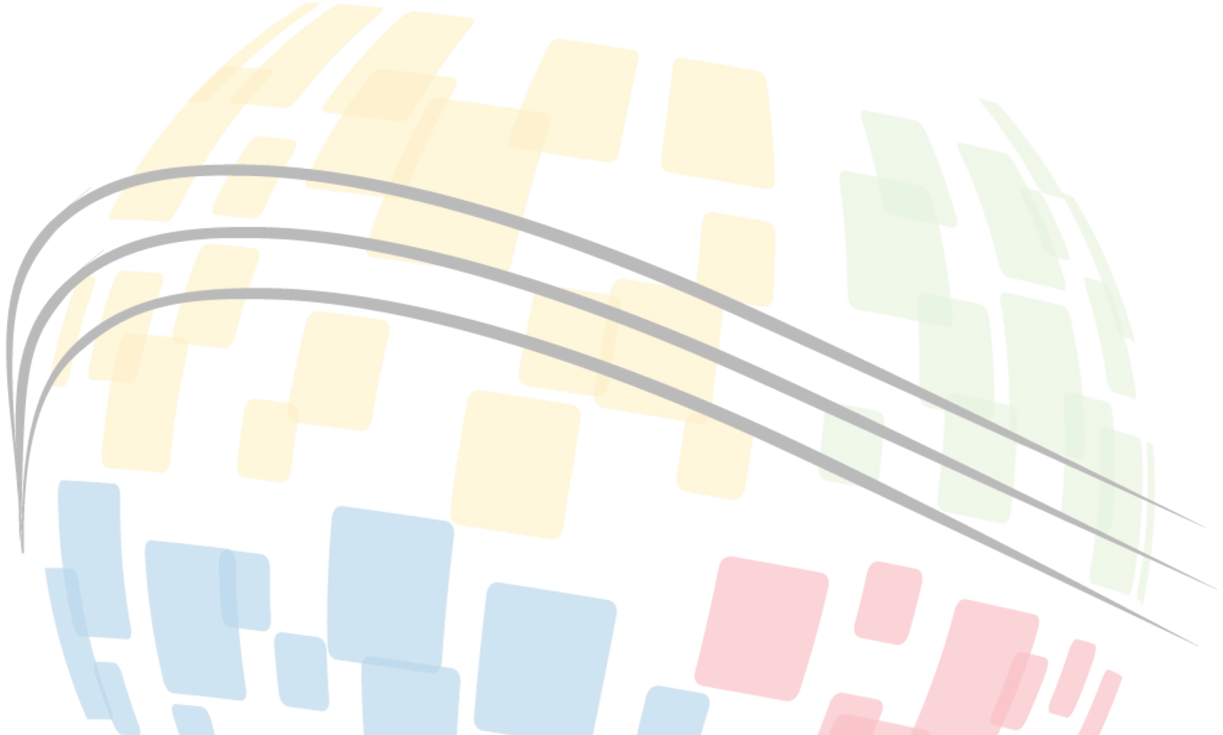
Beyond current modeling, analysis, design, and assurance methods

- **Defined collaborative control interactions using Systems Theory**
  - Taxonomy for structure of interactions between controllers
  - Defined 9 collaborative control dynamics
  - Analyzed 101 aerospace system interactions using framework
- **Foundation for extended hazard analysis and safety-guided design framework**

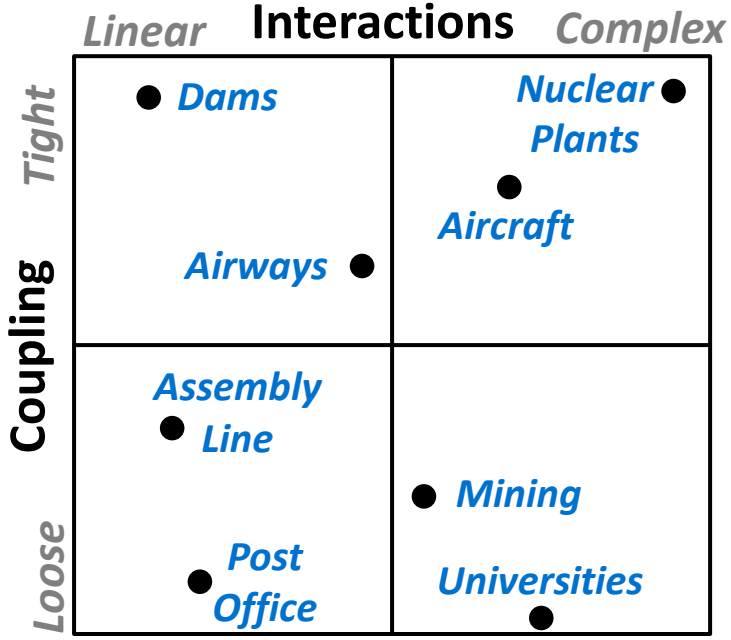
kopeikin@mit.edu

for more: check out PhD dissertation, early Fall 2023

Backup



Prior Taxonomies of System Interactions

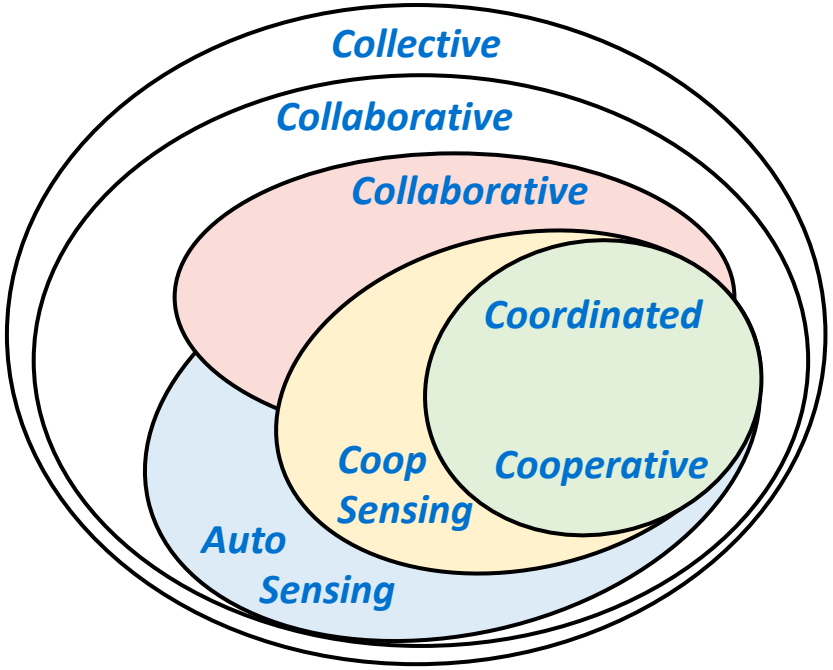


Safety *[Perrow '84]*

Socio-Technical Interactions for FRAM

- *Nature of Agents*
- *Output Nature*
- *Levelling*
- *Waiting Time*
- *Distance*
- *Degree of Coupling*
- *Visibility*
- *Hazards*
- *Parallel Replication*

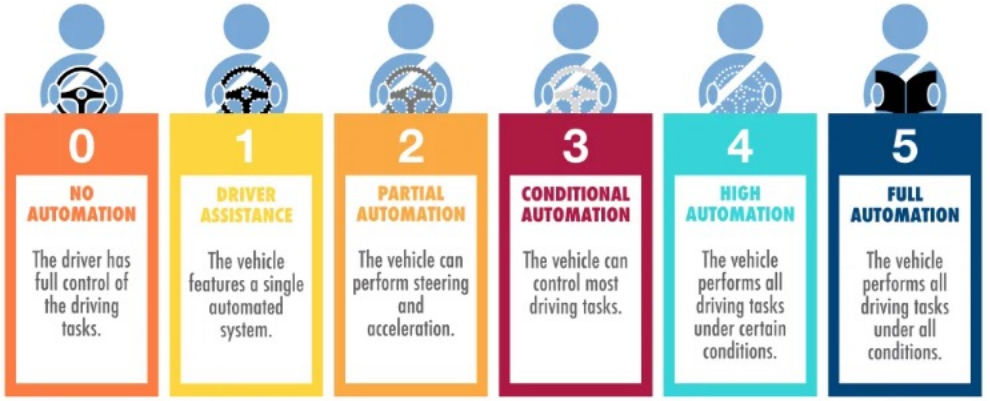
Safety *[Saurin & Patriarca '20]*



Distributed Control *[Murphey & Pardalos '02]*

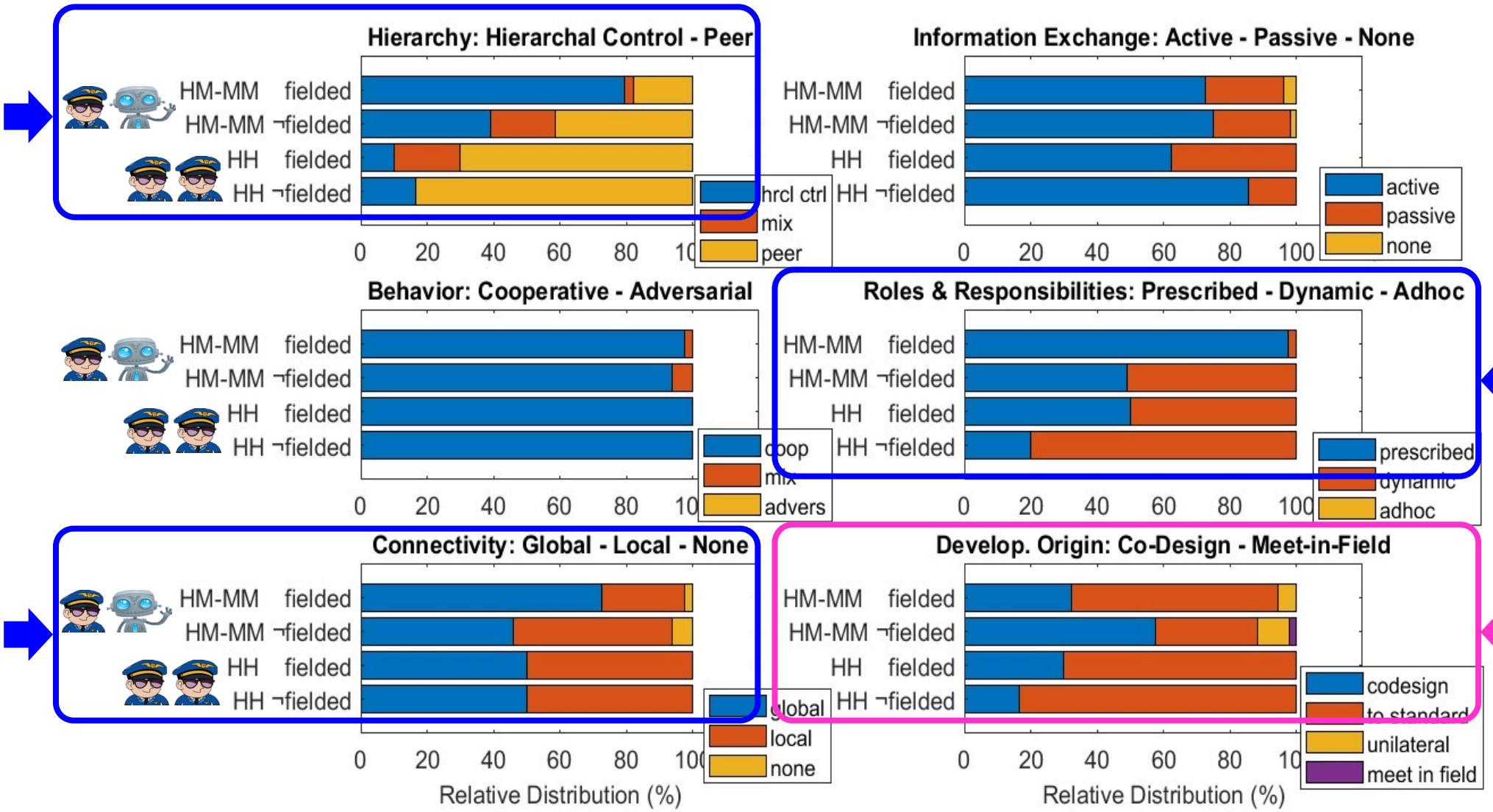
		Topology of Inter-Agent Relationships	
		Centralized (between Distinguished and Subordinate agents)	**Decentralized** (among Peer agents)
Information Flow	**Direct** (messages between agents)	***Construction*** (Build-Time) ***Command*** (Run-time)	***Conversation***
	Indirect (non-message interaction)	***Constraint***	***Stigmergy***[2] (generic) ***Competition*** (limited resources)

Multi-Agent *[Parunak et al. '04]*



Levels-of Automation *[Sheridan & Verplank '78]*

Comparing the Structure of Interactions



HM-MM structure in unfielded systems closer to HH interactions than in fielded HM-MM

Fielded HM-MM structure closer to HH than unfielded HM-MM

Not a quantitative analysis representative of all systems

HM-MM: Human-Machine & Machine-Machine, HH: Human-Human, ¬fielded: unfielded

Safety Assurance of Collaborative Systems

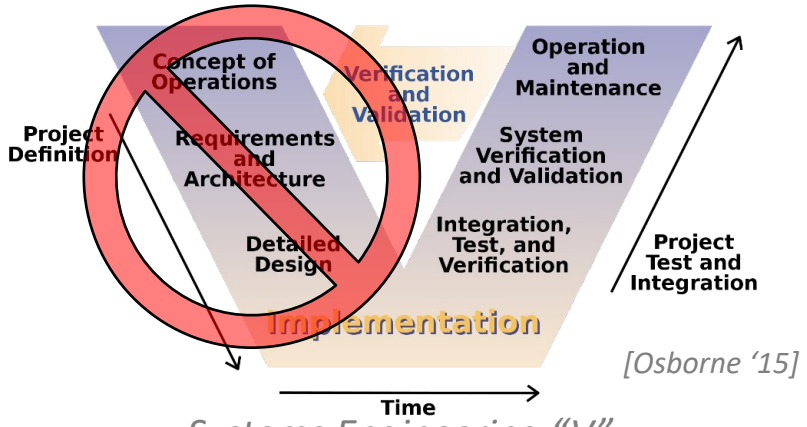
Activities for confidence system hazards eliminated / controlled *[Leveson '21]*

Hazard Analysis | **V&V*** | **Certification**

2 key problems with current practices *[Leveson '11; '21]*

Applied too late

- Emphasized on right-side of "V"
- Prevents designing safety in early

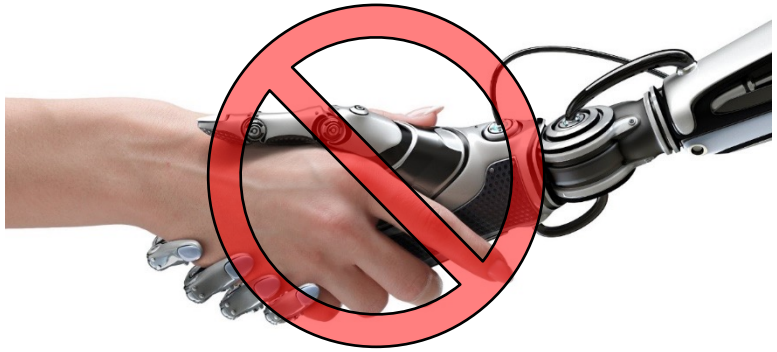


[Osborne '15]

Systems Engineering "V"

Inadequate

- HW, SW, Humans analyzed separately
- Unsuitable for collaborative control



www.dailymail.co.uk/sciencetech/

**V&V: Verification & Validation*