

A New Approach to Engineering for Safety and Security

Prof. Nancy Leveson

Aeronautics and Astronautics
MIT



Bottom Line Up Front (BLUF)

- Complexity is reaching a new level (tipping point)
  - Old safety approaches becoming less effective
  - New causes of losses appearing (especially related to use of software and autonomy)
- Traditional analysis approaches do not provide the information necessary to prevent losses in these systems
- Need a paradigm change to a "systems approach"

Change focus

~~Increase component reliability (prevent failures)~~

Enforce safe system behavior (constraints on system behavior)

BLUF (2)

- Allows creating new analysis and engineering approaches
  - More powerful and inclusive
  - Orders of magnitude less expensive
  - Work on extremely complex systems (top-down system engineering)
  - Help to design safety, security, and other properties in from the beginning
- New paradigm works much better than old techniques:
  - Empirical evaluations and controlled studies show it finds more causal scenarios (the “unknown unknowns”)
  - Can be used before a detailed design exists to design safe and secure systems from the beginning

General Definition of “Safety”

- Accident = Loss: Any undesired and unplanned event that results in a loss
  - Loss of human life or injury,
  - Property damage,
  - Environmental pollution or damage,
  - Mission loss,
  - Negative business impact (damage to reputation, etc.), product launch delay, legal entanglements
- Includes inadvertent and intentional (so includes security)
- System goals vs. constraints
- Applies to any emergent “system” property (e.g., serviceability)

- Why is something new needed?
- Analytic Reduction vs. Systems Theory
- STAMP causality model
- STPA (System-Theoretic Process Analysis)
- New Research Directions (UAM and FVL)
  - Conceptual architectures
  - Risk assessment improvement
  - Human-Automation teaming/coordination
  - Leading indicators of increasing risk

Why is something new needed?

The first step in solving any problem is understanding it.

> *"It's never what we don't know that stops us.*
> *It's what we do know that just ain't so."*

Our current tools are all 50-65 years old but our technology is very different today

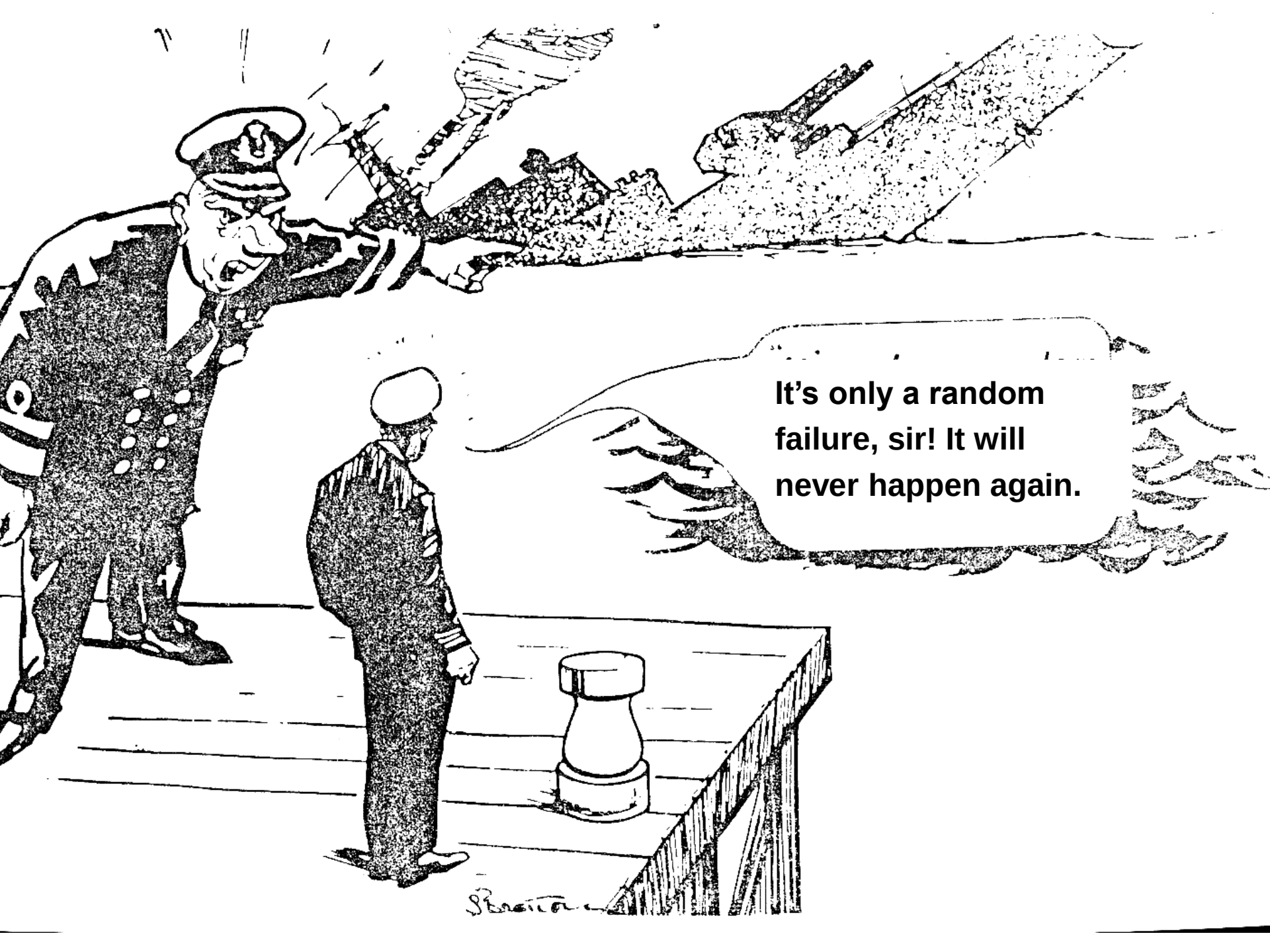
1940 1950 1960 1970 1980 1990 2000 2010 2020

FMEA

FTA HAZOP ETA Bow Tie

- Introduction of computer control
- Exponential increases in complexity
- New technology
- Changes in human roles

Assumes accidents caused by component/human failures


It’s only a random failure, sir! It will never happen again.

What Failed Here?

- Navy aircraft were ferrying missiles from one location to another.
- One pilot executed a planned test by aiming at aircraft in front and firing a dummy missile.
- Nobody involved knew that the software was designed to substitute a different missile if the one that was commanded to be fired was not in a good position.
- In this case, there was an antenna between the dummy missile and the target so the software decided to fire a live missile located in a different (better) position instead.

Warsaw A320 Accident

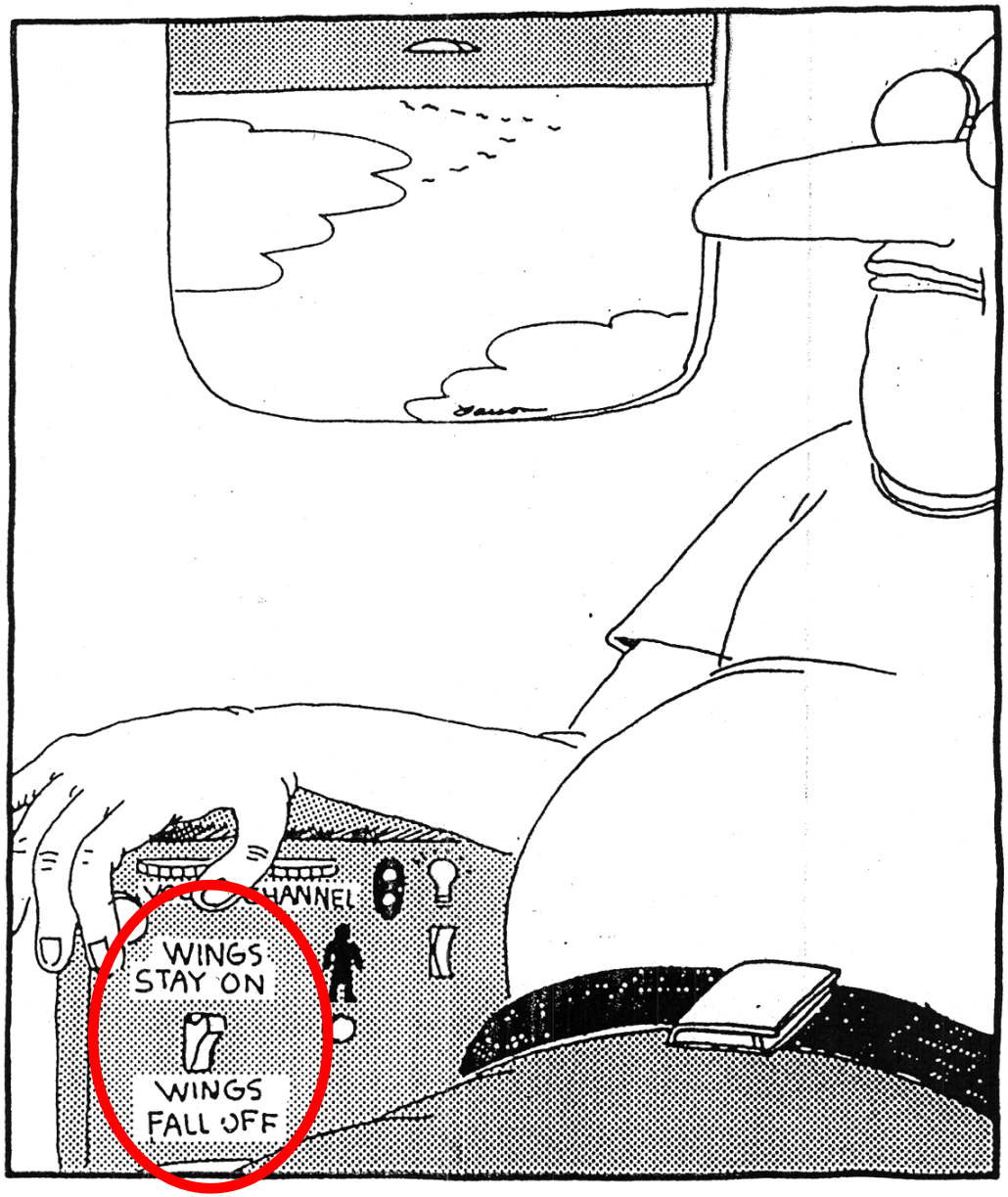
- Software protects against activating thrust reversers when airborne
- Hydroplaning and other factors made the software think the plane had not landed
- Pilots could not activate the thrust reversers and ran off end of runway into a small hill.

Traditional Approach to Safety

- Traditionally view safety as a failure problem
  - Chain of directly related failure events leads to loss
  - Try to prevent component failures or establish barriers between events
- Limitations
  - Systems are becoming more complex
    - Accidents often result from interactions among components
    - Cannot anticipate all potential interactions; cannot exhaustively test
  - Omits or oversimplifies important factors
    - Human error
    - New technology (including software)
    - Culture and management
    - Evolution and adaptation

Accidents are <u>not</u> just the result of random failure

What about The role of humans in Systems?



Fumbling for his recline button Ted unwittingly instigates a disaster

Change in the Way We Conceive of Human Error

Traditional Approach:

- Operators/pilots responsible for most accidents
- So fire, train them not to make mistakes, or add more automation (which marginalizes the pilot and causes more and different errors)

Systems Approach:

- Human behavior always affected by the context in which it occurs
  - We are designing systems in which human error inevitable
  - The role of operators is also changing
- **Human error is a symptom of a system that needs to be redesigned**.

Another Accident Involving Thrust Reversers

- Tu-204, Moscow, 2012
- Red Wings Airlines Flight 9268
- The soft 1.12g touchdown made runway contact a little later than usual.
- With the crosswind, this meant weight-on-wheels switches did not activate and the thrust-reverse system would not deploy.

Another Accident Involving Thrust Reversers

- Pilots believe the thrust reversers are deploying like they always do. With the limited runway space, they quickly engage high engine power to stop quicker. Instead this accelerates the Tu-204 forwards, eventually colliding with a highway embankment.

Another Accident Involving Thrust Reversers

- Pilots believe the thrust reversers are deploying like they always do. With the limited runway space, they quickly engage high engine power to stop quicker. Instead this accelerates the Tu-204 forwards, eventually colliding with a highway embankment.

In complex systems, human and technical considerations cannot be isolated

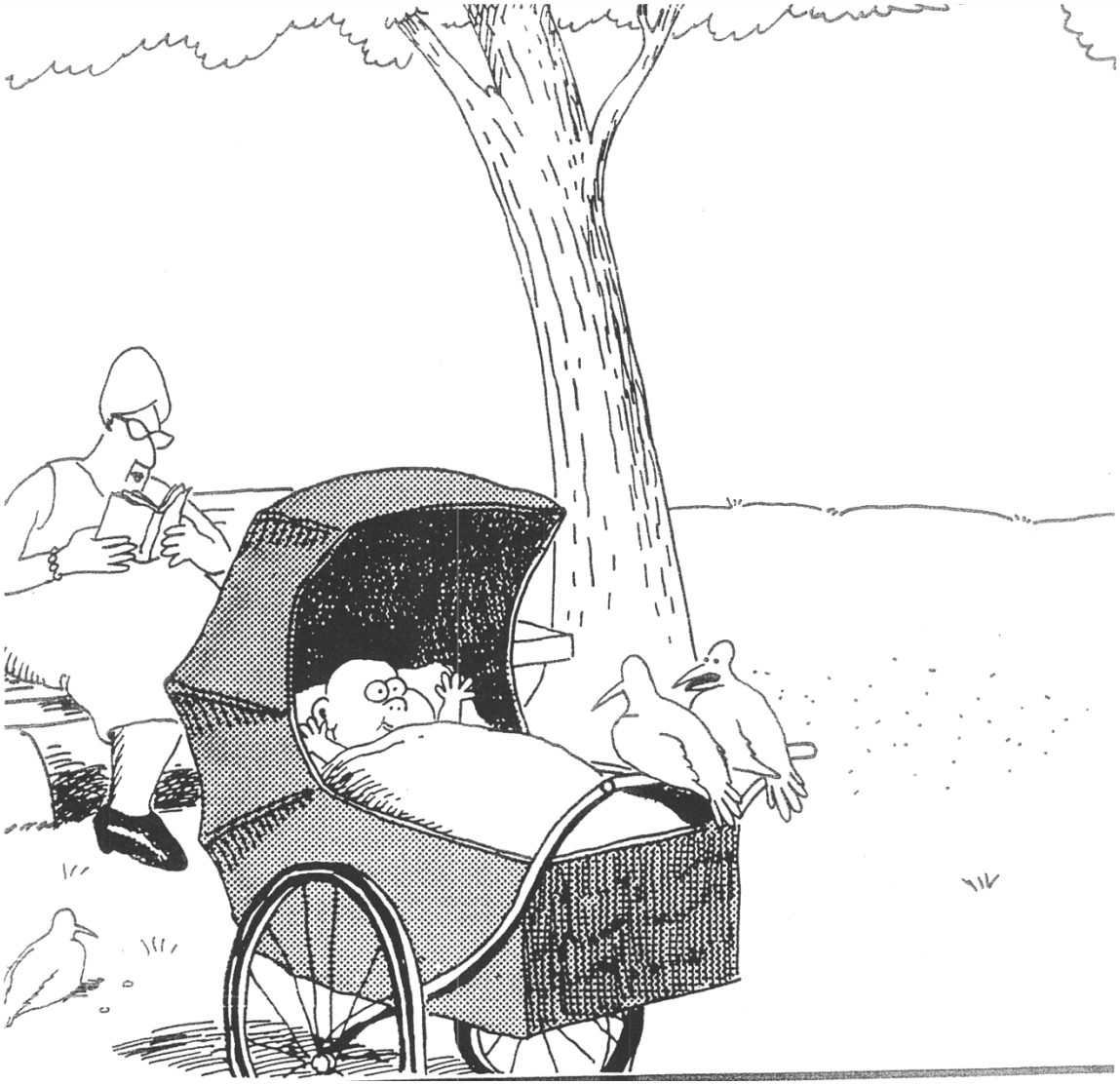
Human factors concentrates on the “screen out”

Hardware/software engineering concentrates on the “screen in”

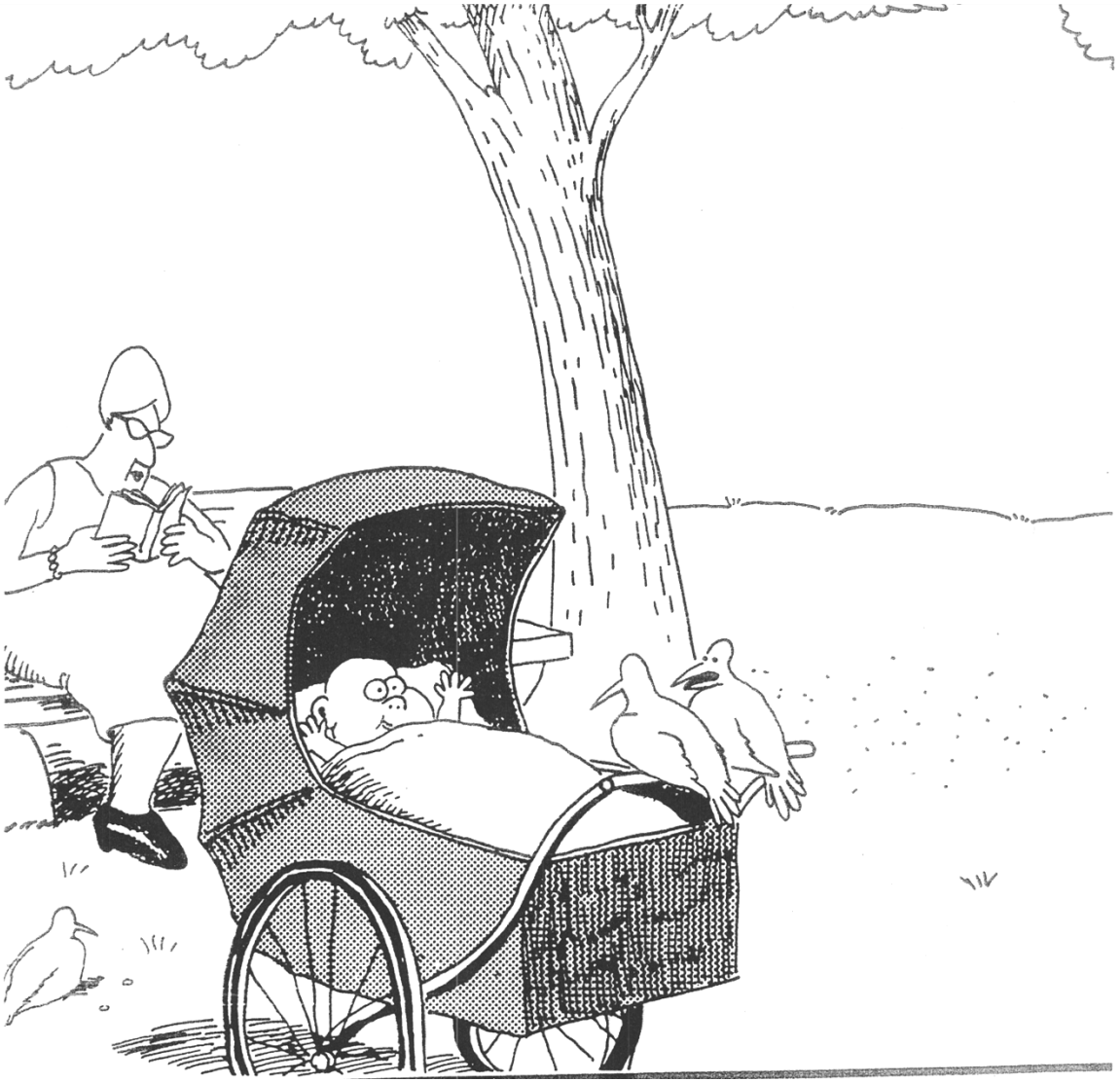
Not enough attention on integrated system as a whole

Increasing number of accidents involving mode confusion, situation awareness errors, inconsistent behavior, etc.

HMI: Human Machine ~~Interface~~? ~~Interaction~~? **Integration**



It’s still hungry … and I’ve been stuffing worms into it all day.



It’s still hungry … and I’ve been stuffing worms into it all day.

We Need New Tools for the New Problems

Analytic Reduction vs. Systems Theory

The Problem is Complexity

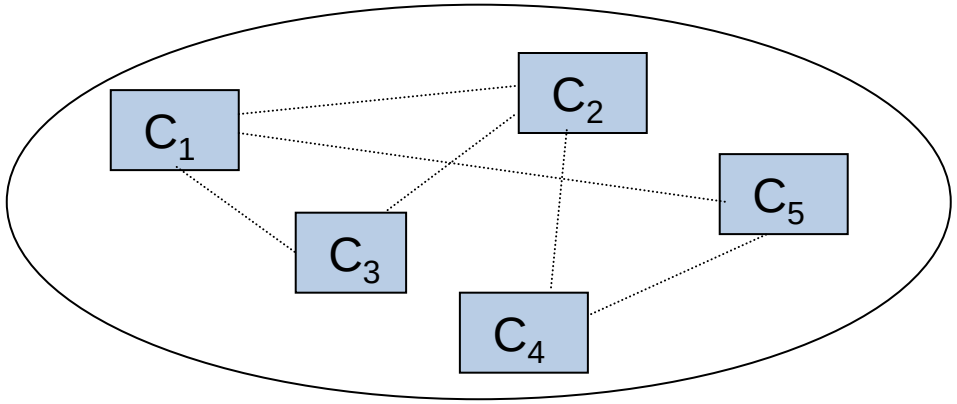
Ways to Cope with Complexity

- Analytic Decomposition
- Statistics
- Systems Theory

Analytic Decomposition ("Divide and Conquer")

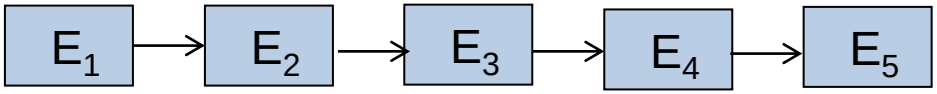
1. Divide system into separate parts

Physical/Functional: Separate into distinct components



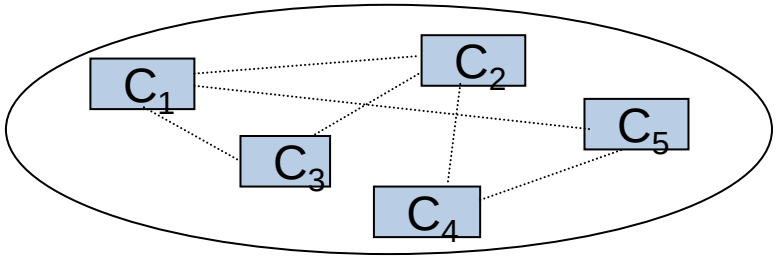
Components interact
In direct ways

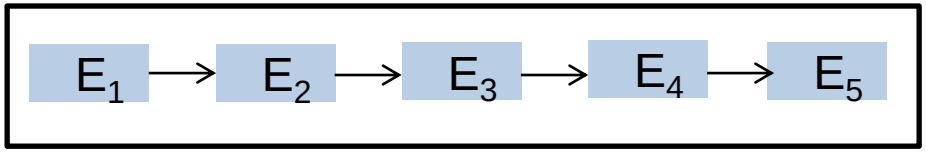
Behavior: Separate into events over time



Each event is the direct
result of the preceding event

Analytic Decomposition (2)





2. Analyze/examine pieces separately and combine results

- Assumes such separation does not distort phenomenon
  - ✓ Each component or subsystem operates independently
  - ✓ Components act the same when examined singly as when playing their part in the whole
  - ✓ Components/events not subject to feedback loops and non-linear interactions
  - ✓ Interactions can be examined pairwise

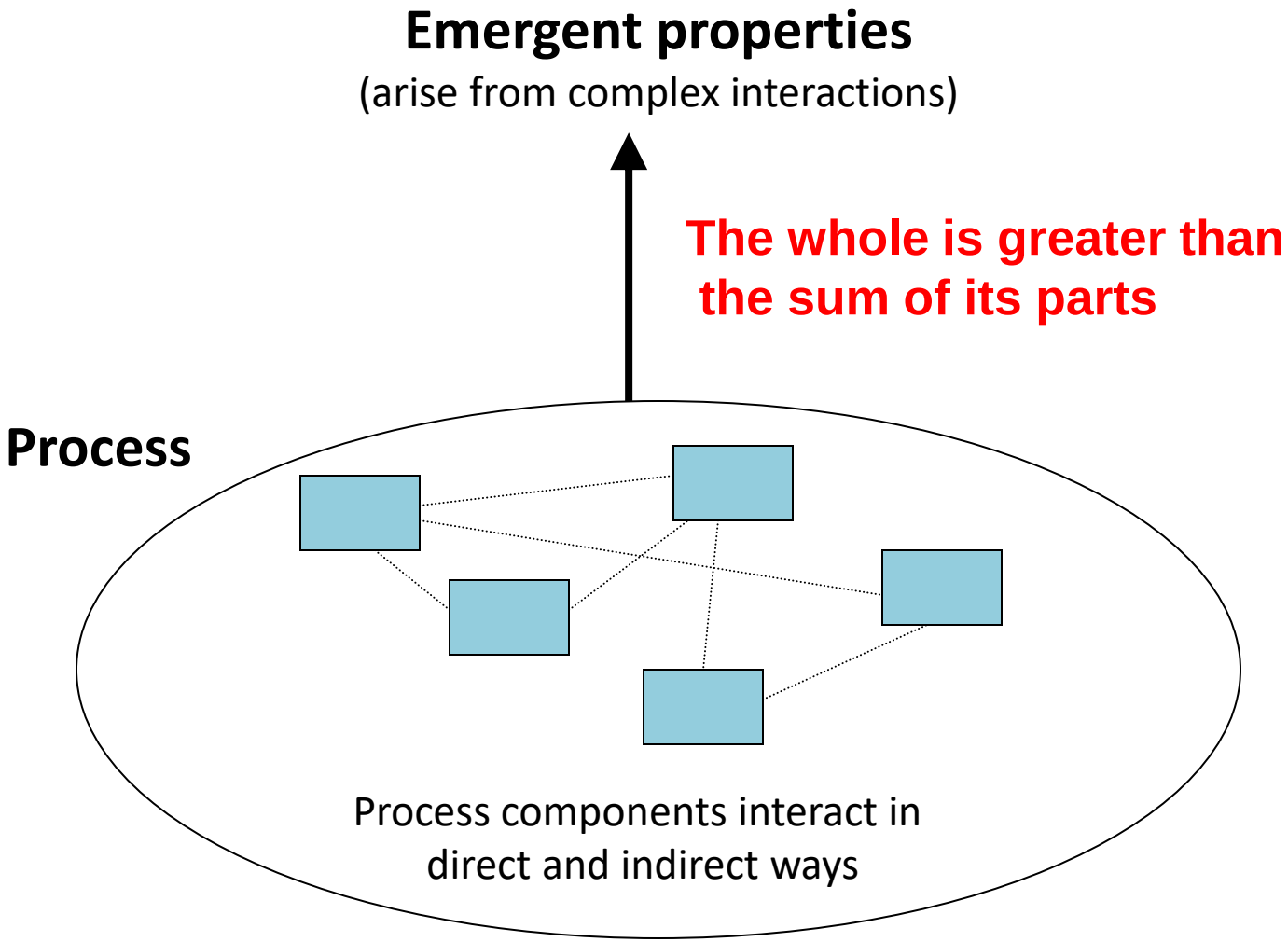
The Problem

- These assumptions are no longer true in our
  - Tightly coupled
  - Software intensive
  - Highly automated
  - Connected

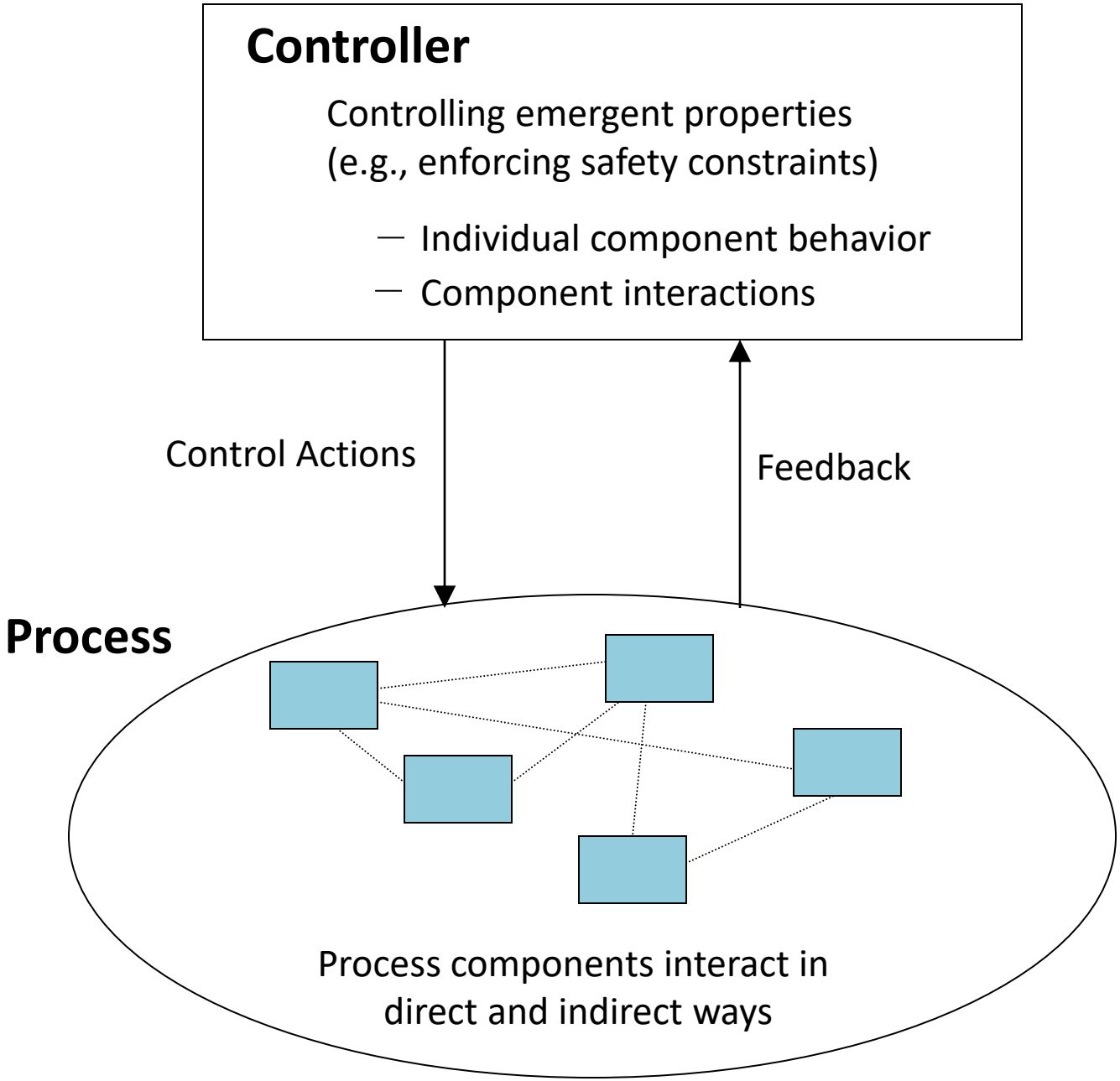
  engineered systems

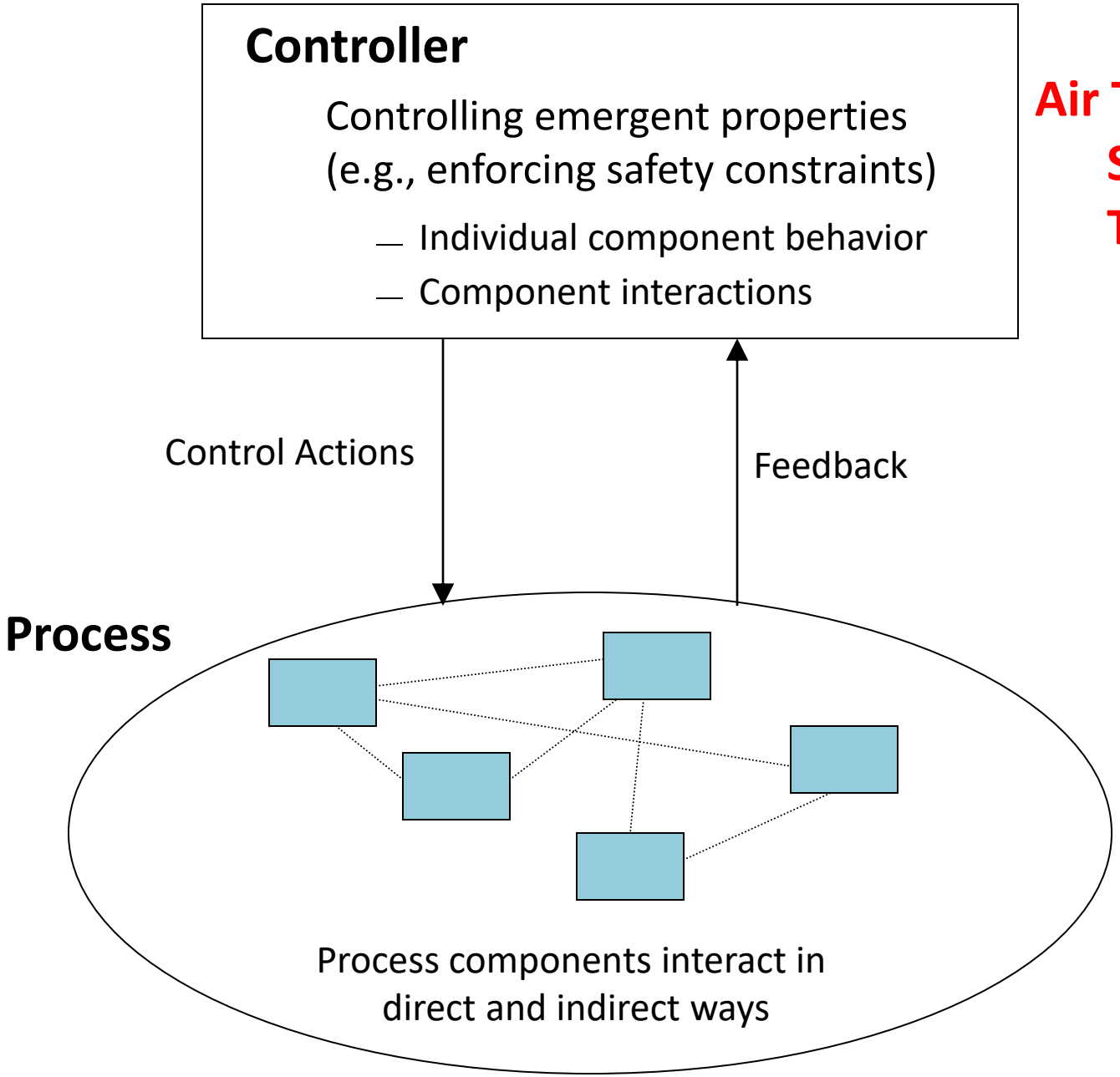
- Need a new theoretical basis
  - *System theory* can provide it

System Theory



Safety and security are emergent properties


Controller
Controlling emergent properties
(e.g., enforcing safety constraints)
— Individual component behavior
— Component interactions
Control Actions
Feedback
Process
Process components interact in
direct and indirect ways



Air Traffic Control:
Safety
Throughput

A Broad View of "Control"

Component failures and unsafe interactions may be "controlled" through design

(e.g., redundancy, interlocks, fail-safe design)

or through process

- Manufacturing processes and procedures
- Maintenance processes
- Operational processes

or through social controls

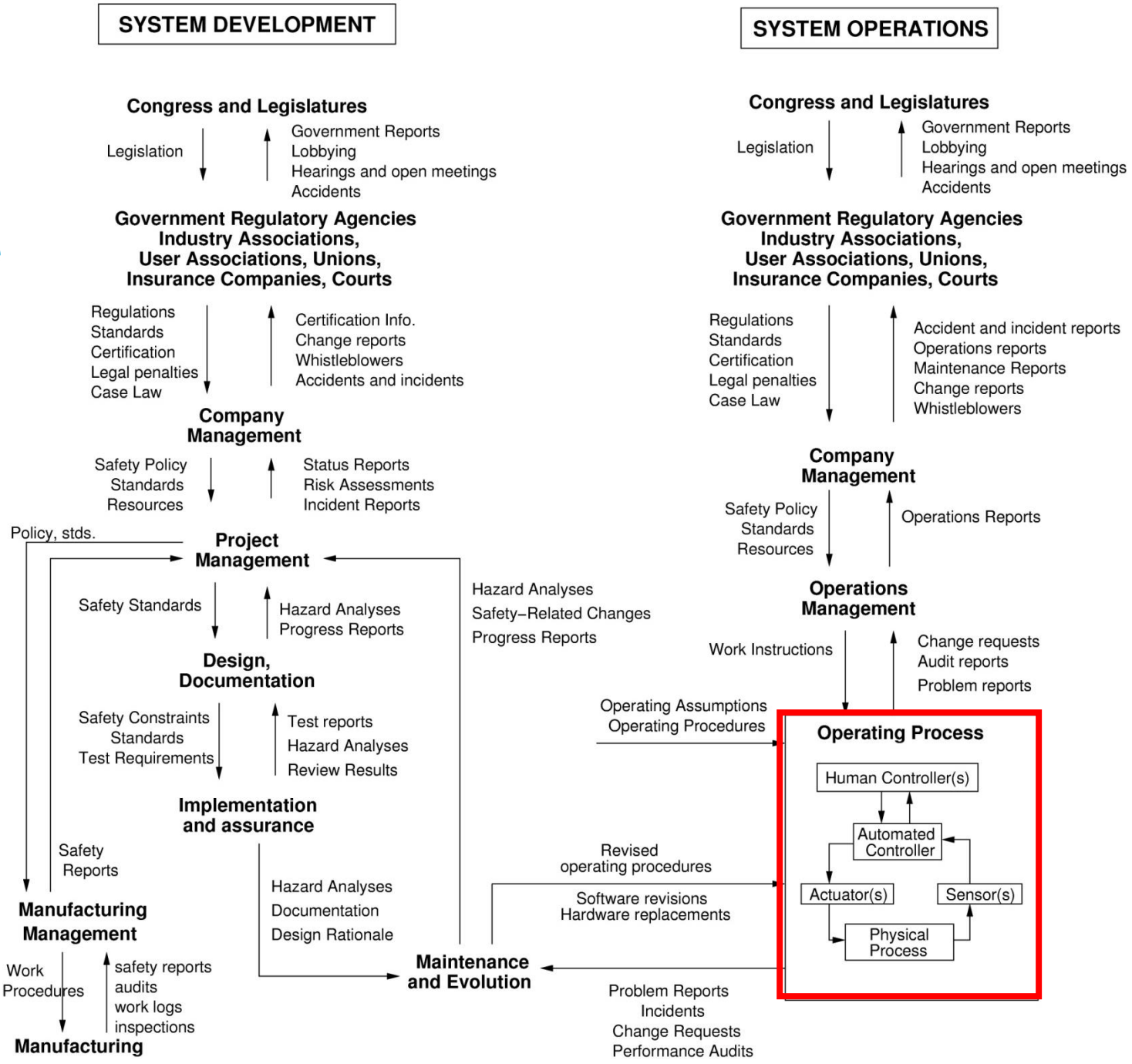
- Governmental or regulatory
- Culture
- Insurance
- Law and the courts
- Individual self-interest (incentive structure)

Controls/Controllers Enforce Safety Constraints

- Power must never be on when access door open
- Two aircraft/automobiles must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Public health system must prevent exposure of public to contaminated water, food products, and viruses
- Pressure in a offshore well must be controlled
- Toxic chemicals/radiation must not be released from plant
- Workers must not be exposed to workplace hazards
- Integrity of hull must be maintained on a submarine

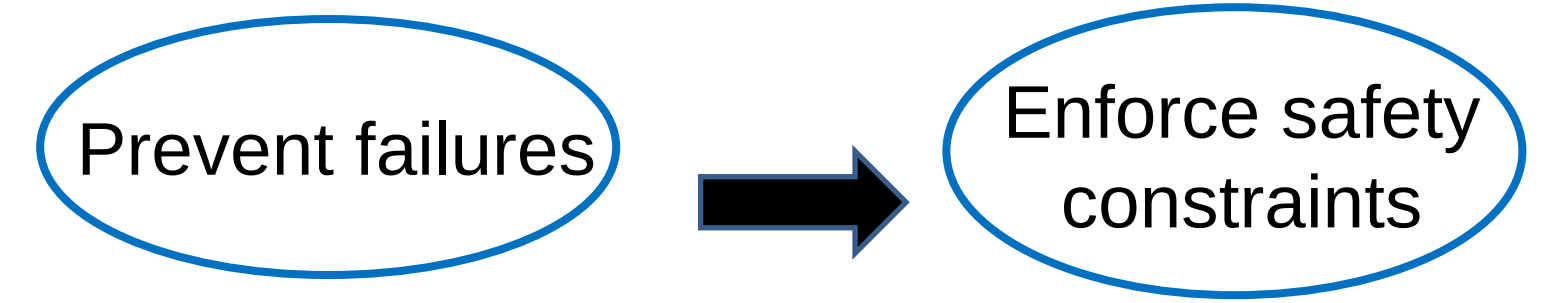
These are the High-Level Functional Hazard-Related Safety/Security Requirements to Address During Design

Example Safety Control Structure (SMS)



Here comes the paradigm change!

The paradigm change for effective safety and security engineering!



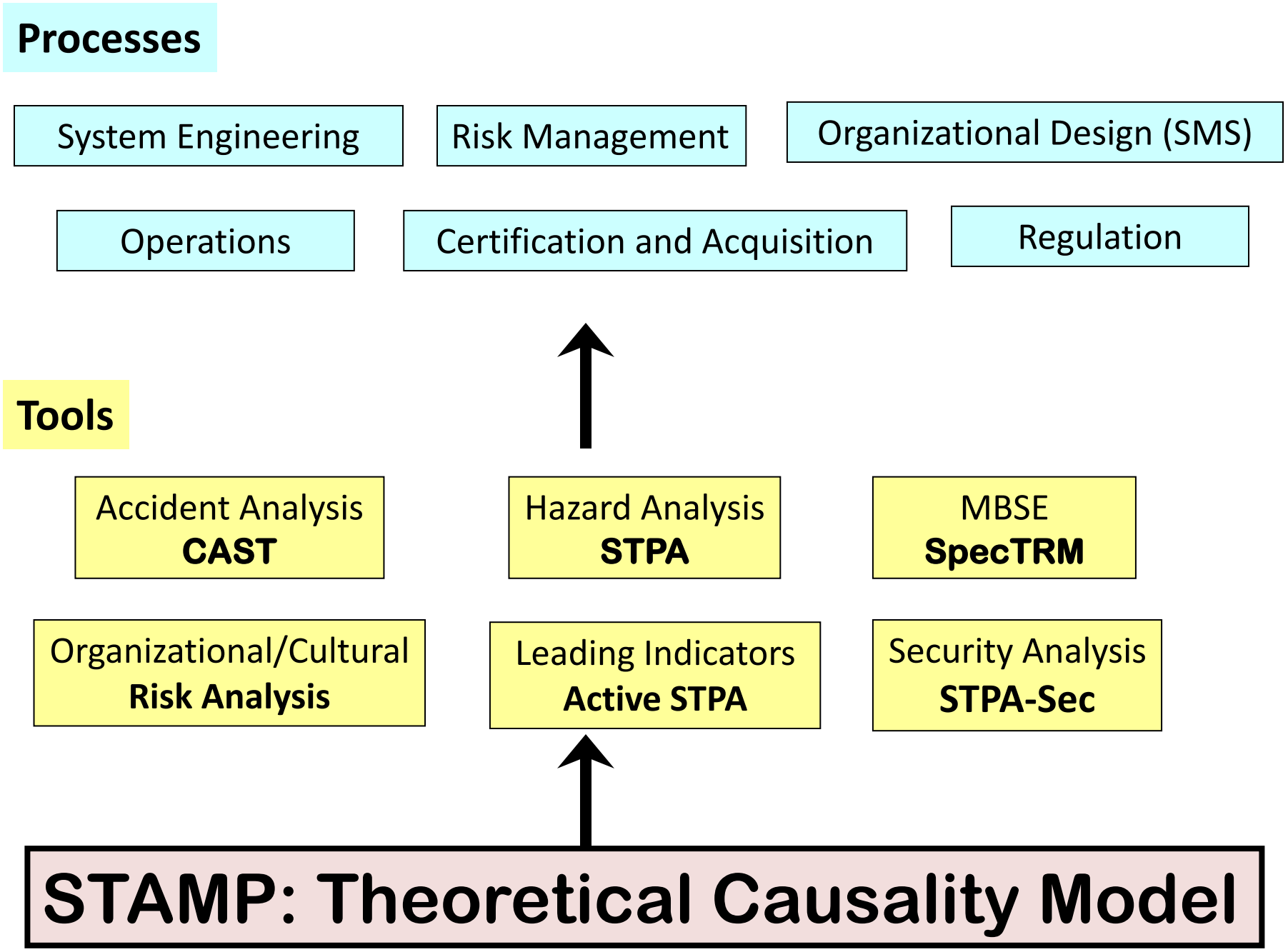
Treat Safety as a **Reliability** Problem

Treat Safety as a **Control** Problem

The STAMP Causality Model

STAMP
(System-Theoretic Accident Model and Processes)

- A new, more powerful accident/loss causality model
- Based on systems theory, not reliability theory
- Treats accidents/losses as a dynamic control problem (vs. a failure problem):
  - Control individual component failures
  - Control interactions among components
- Includes
  - Entire socio-technical system (not just technical part)
  - Component interaction accidents
  - Software and system design errors
  - Human errors
  - Plus all the old accidents


Processes
System Engineering
Risk Management
Organizational Design (SMS)
Operations
Certification and Acquisition
Regulation
Tools
Accident Analysis
CAST
Hazard Analysis
STPA
MBSE
SpecTRM
Organizational/Cultural
Risk Analysis
Leading Indicators
Active STPA
Security Analysis
STPA-Sec
STAMP: Theoretical Causality Model

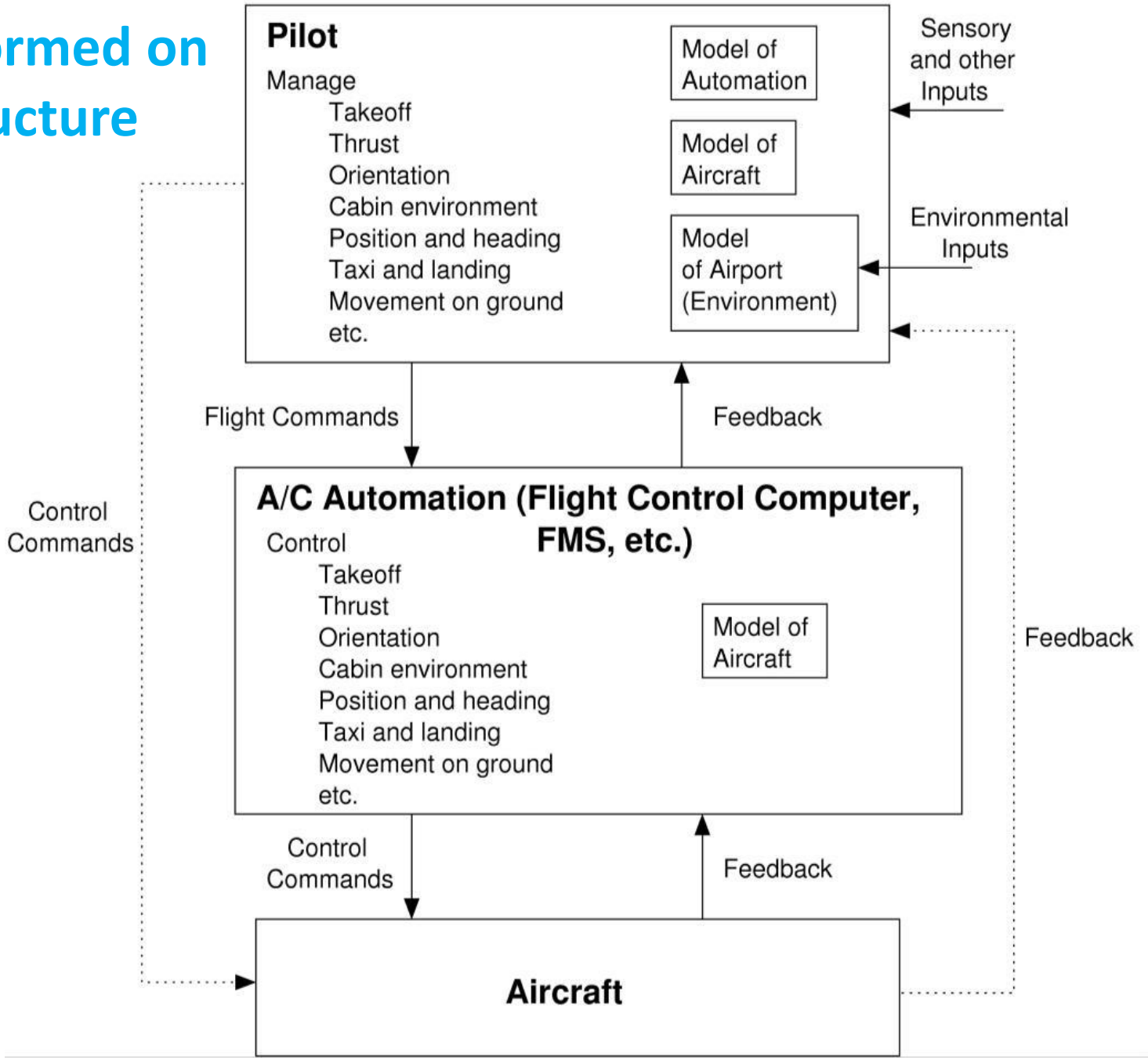
STPA

System-Theoretic Process Analysis

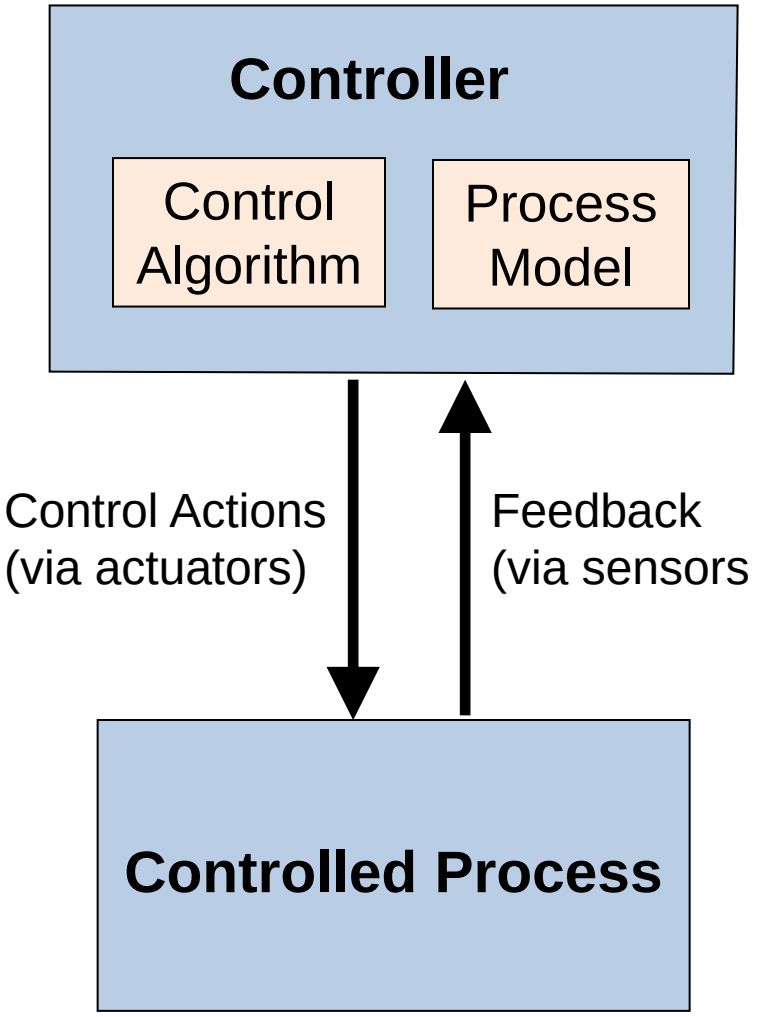
STPA: System-Theoretic Process Analysis

- A top-down, system engineering analysis technique
- Identifies safety and security (or any emergent property) constraints (system and component requirements)
- Identifies scenarios leading to violation of constraints (requirements); use results to design or redesign system to be safer
- Can be used on technical design and organizational design
- Supports a safety-driven design process where
  - Analysis influences and shapes early design decisions
  - Analysis iterated and refined as design evolves

STPA is performed on a control structure



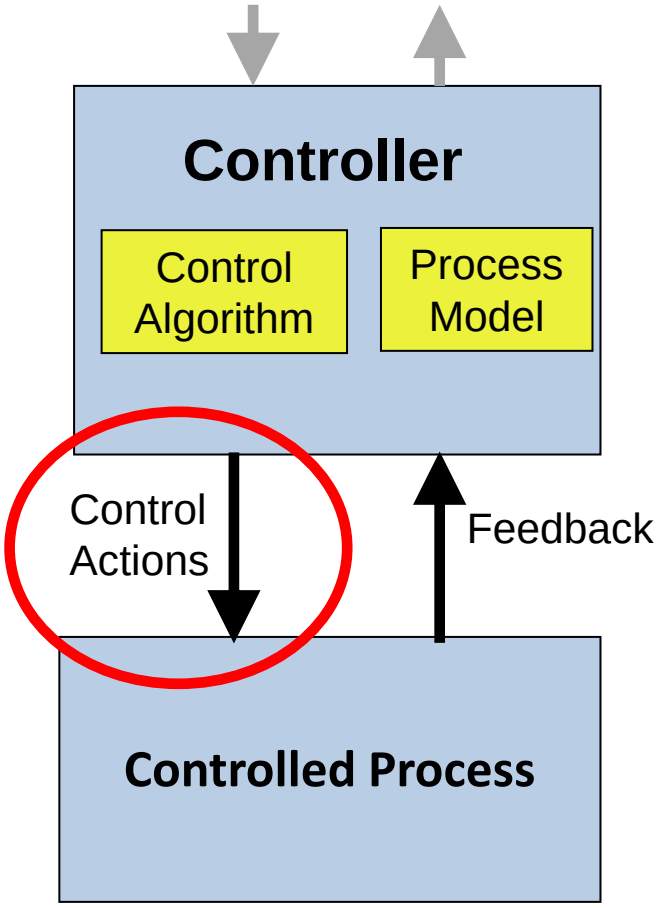
Safety as a Control Problem (vs. Failure Problem)



- Controllers use a **process model** to determine control actions
- Software/human related accidents usually occur when the process model is incorrect
- Captures software errors, human errors, flawed requirements …

Treat safety as a control problem, not a failure problem

Hazard and Accident Analysis with STPA



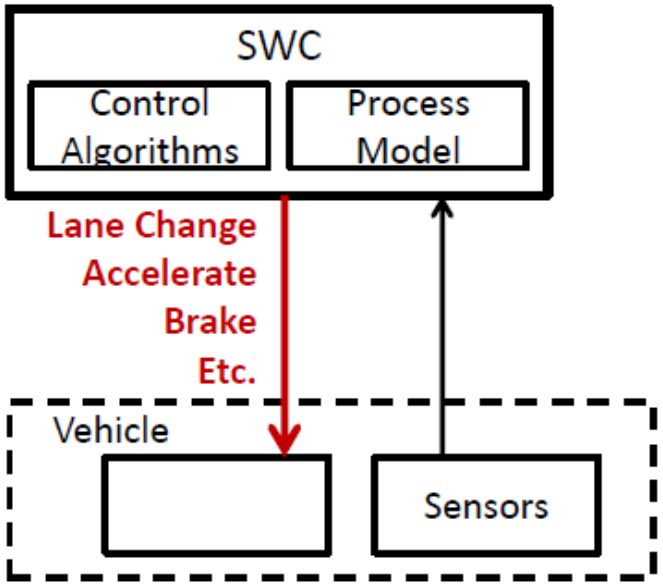
Four types of unsafe control actions

1) Control commands required for safety are not given
2) Unsafe commands are given
3) Potentially safe commands but given too early, too late, or in wrong order
4) Control action stops too soon or applied too long (continuous control)

Analysis and Design:

1. Identify potential unsafe control actions
2. Identify why they might be given (scenarios)
3. Eliminate scenarios through design or operations
4. If safe ones provided, then why not followed?

STPA: Unsafe Control Actions (UCA)



Example:

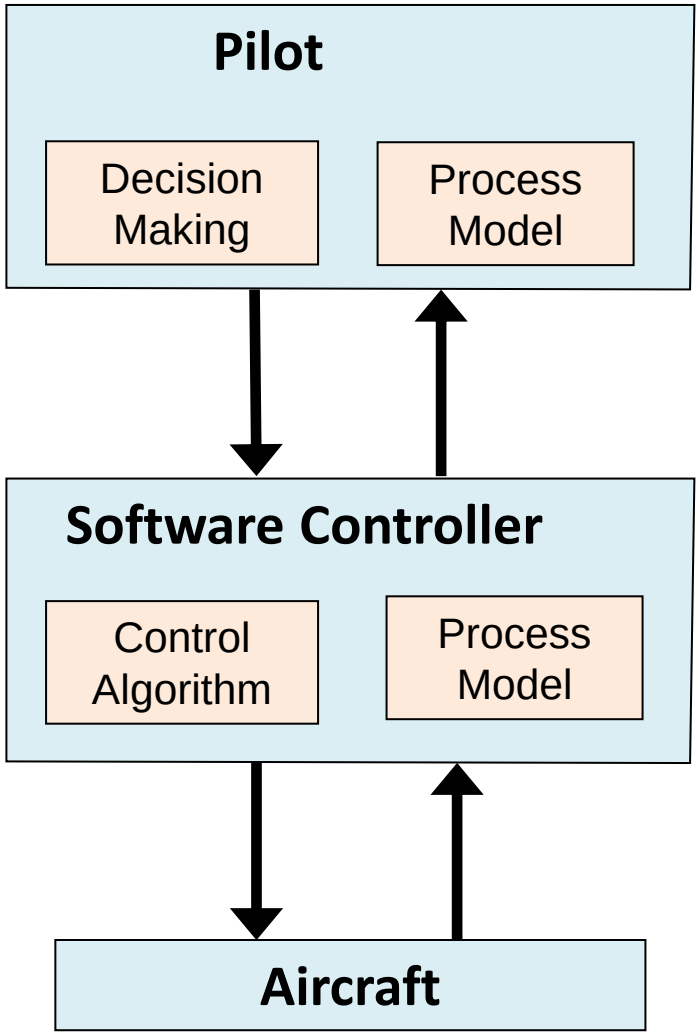
"SWC does not provide brake cmd when path is obstructed"

- SWC — Source Controller
- does not provide — Type
- brake cmd — Control Action
- path is obstructed — Context

	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped Too Soon / Applied too long
Brake Command	UCA-1: SWC does not provide Brake cmd when vehicle path is obstructed			

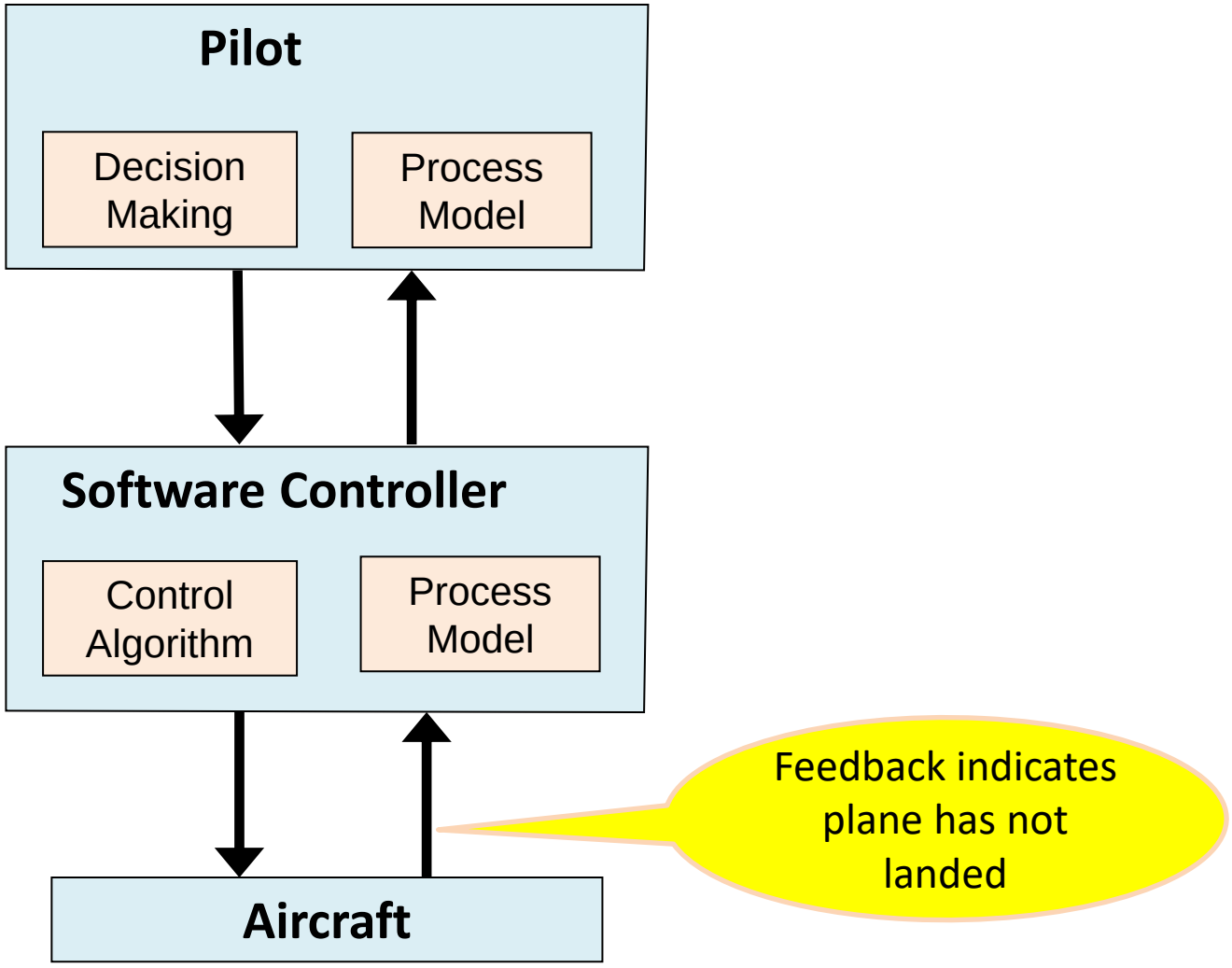
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



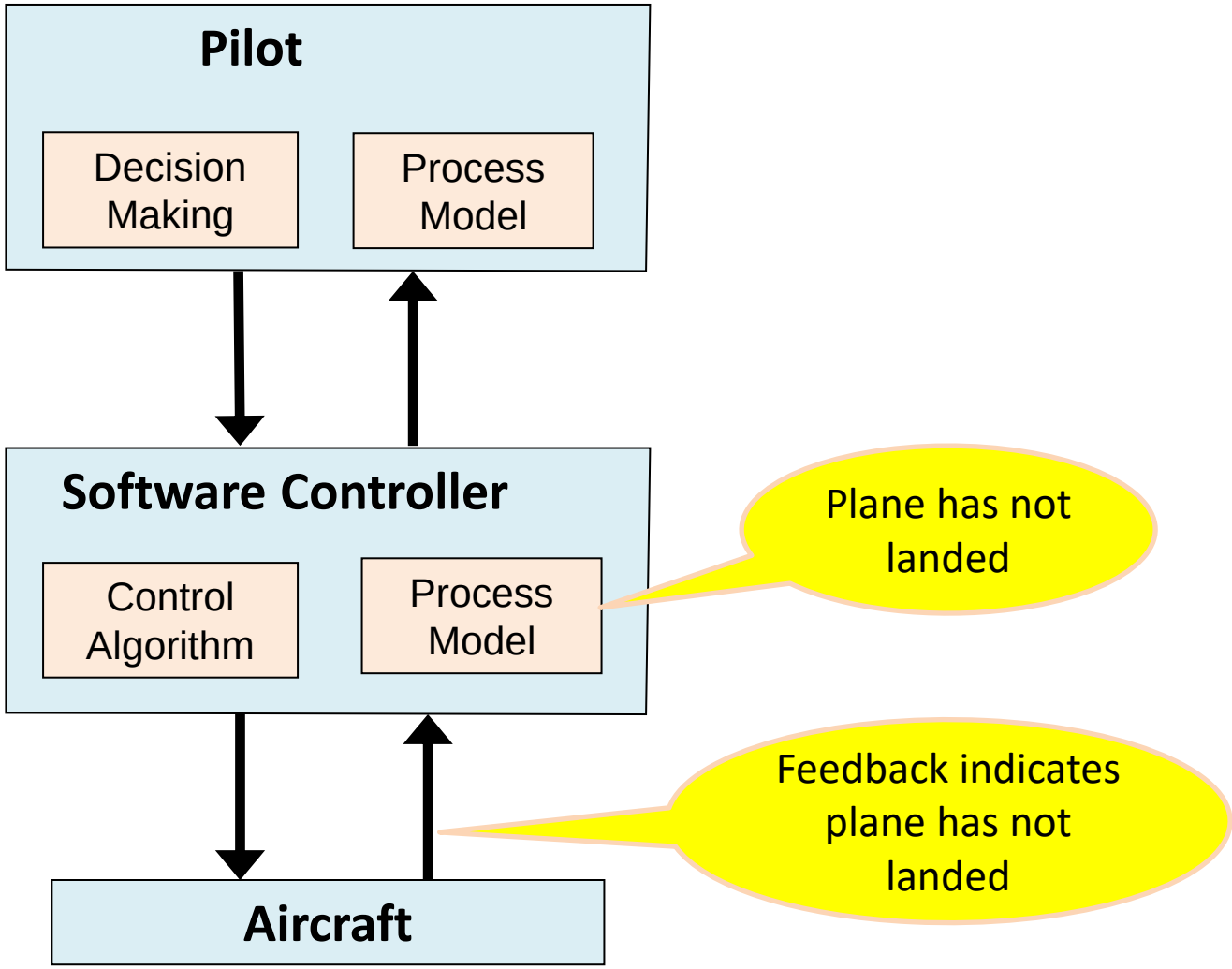
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



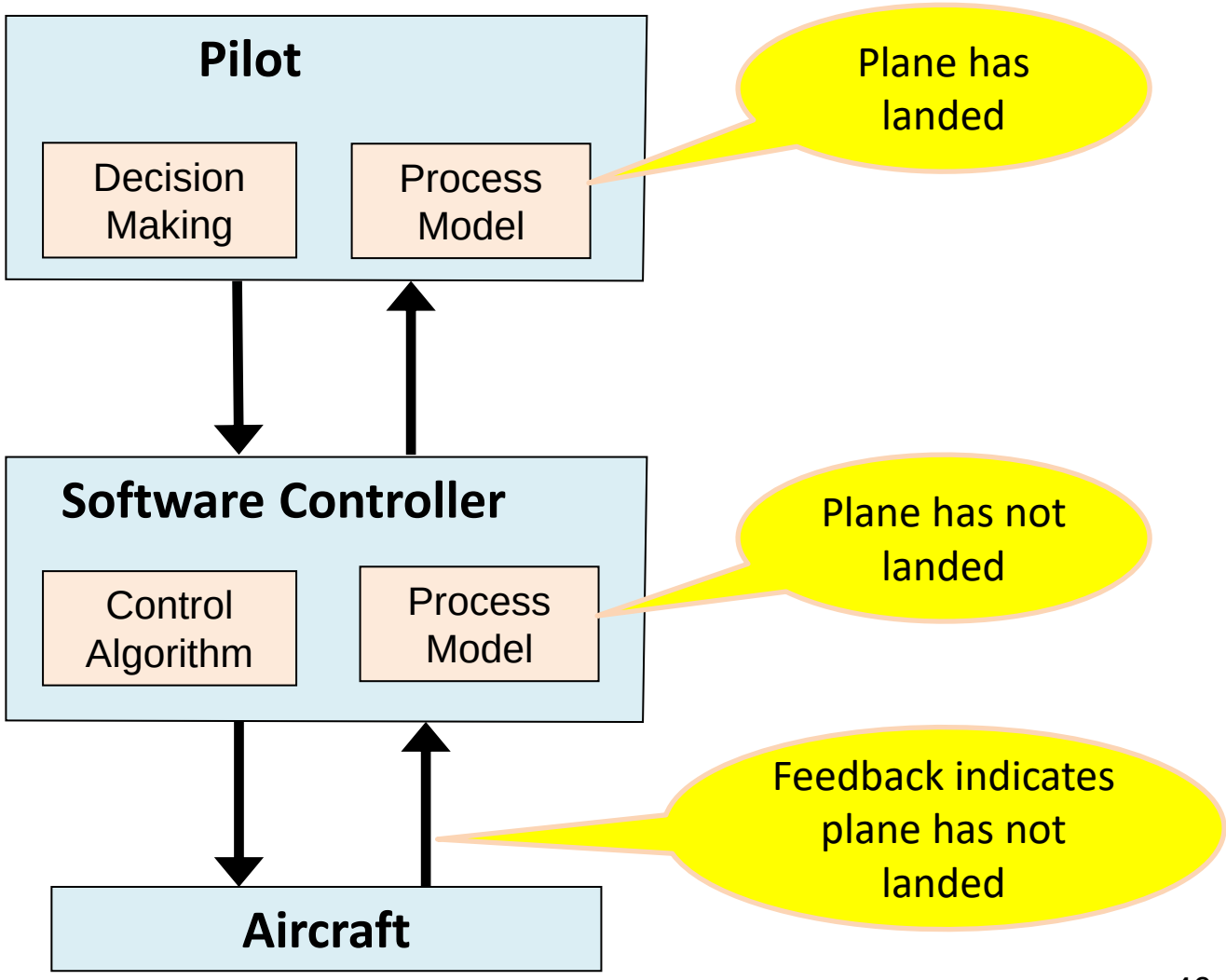
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



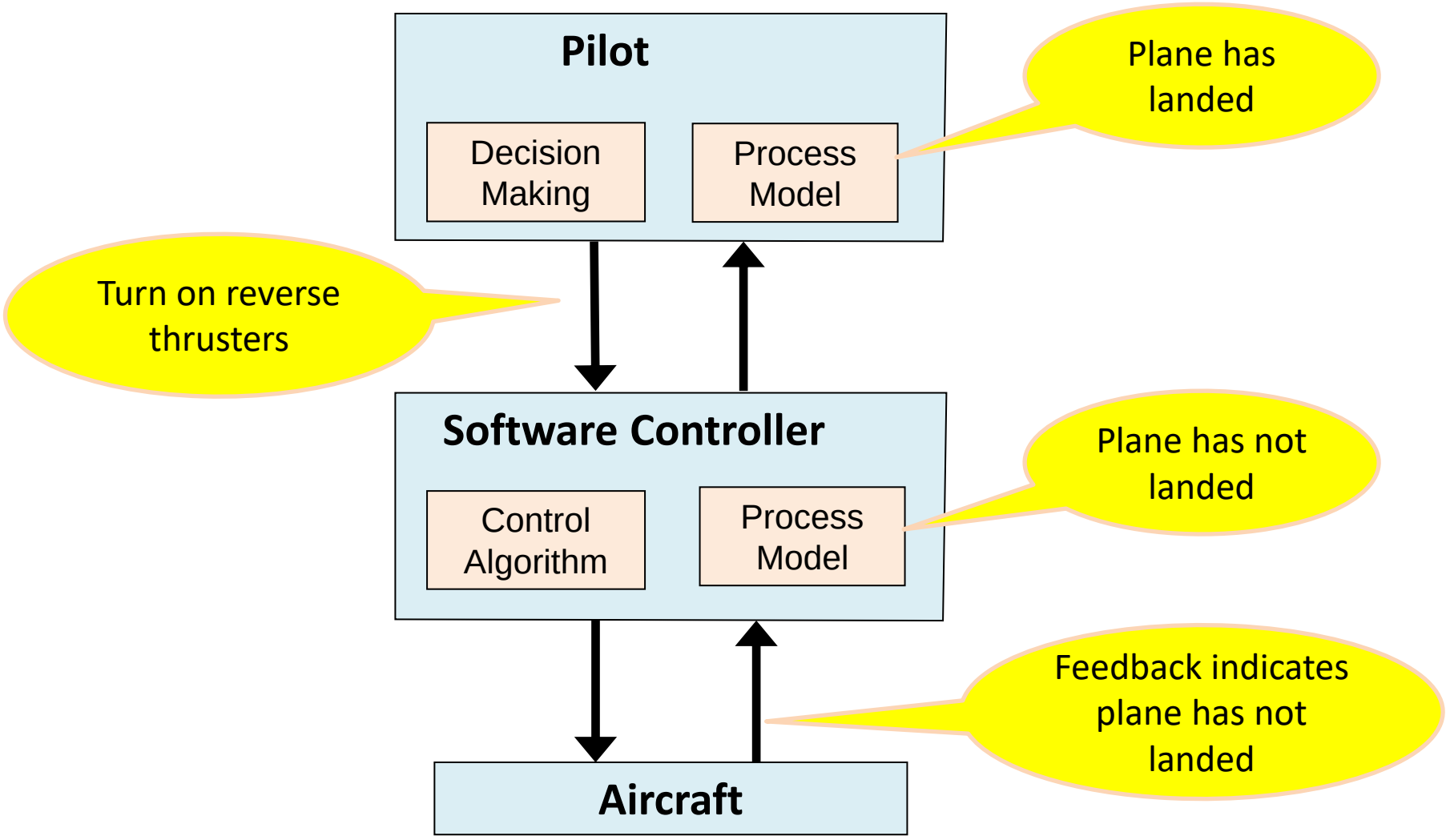
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



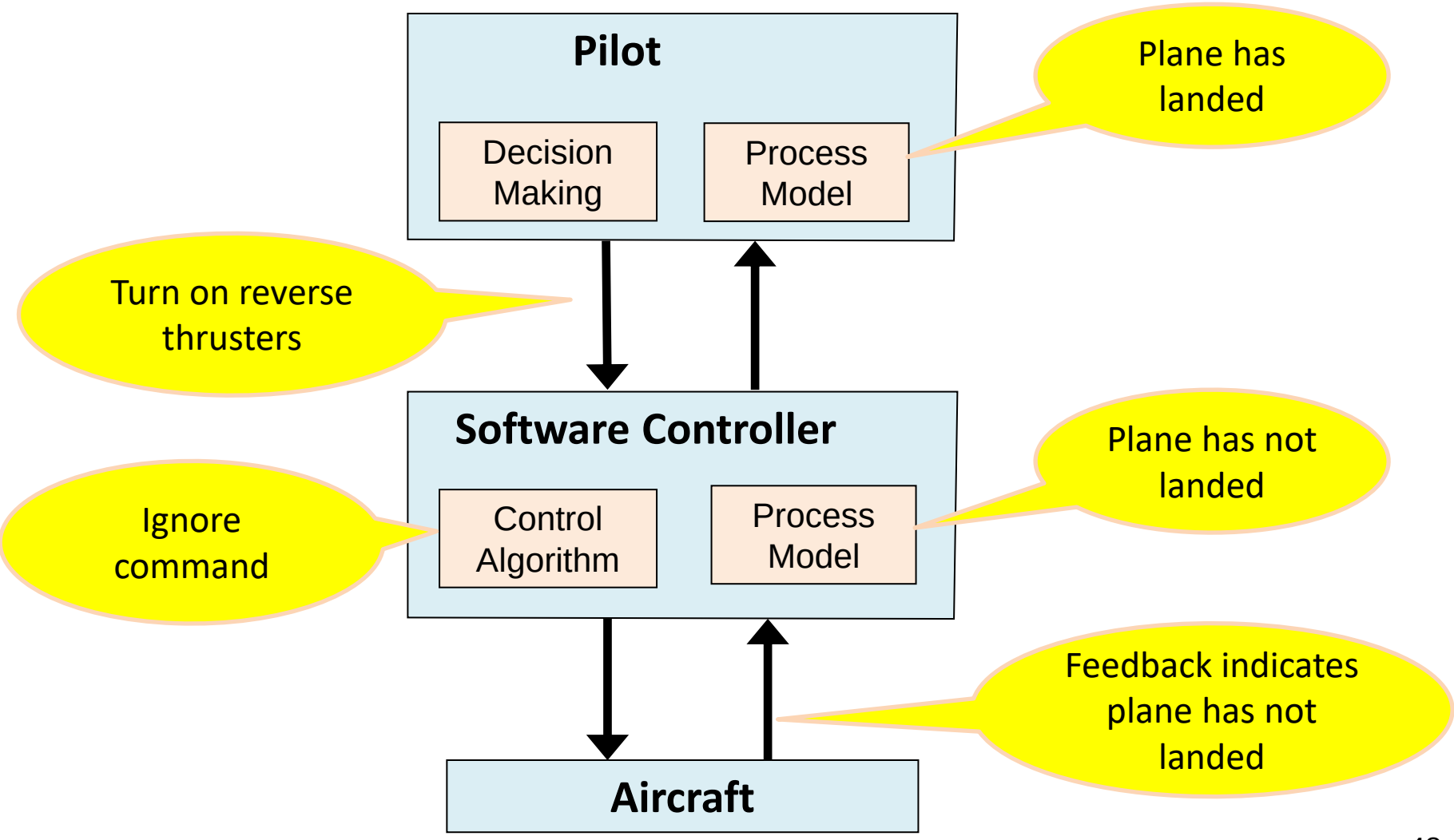
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



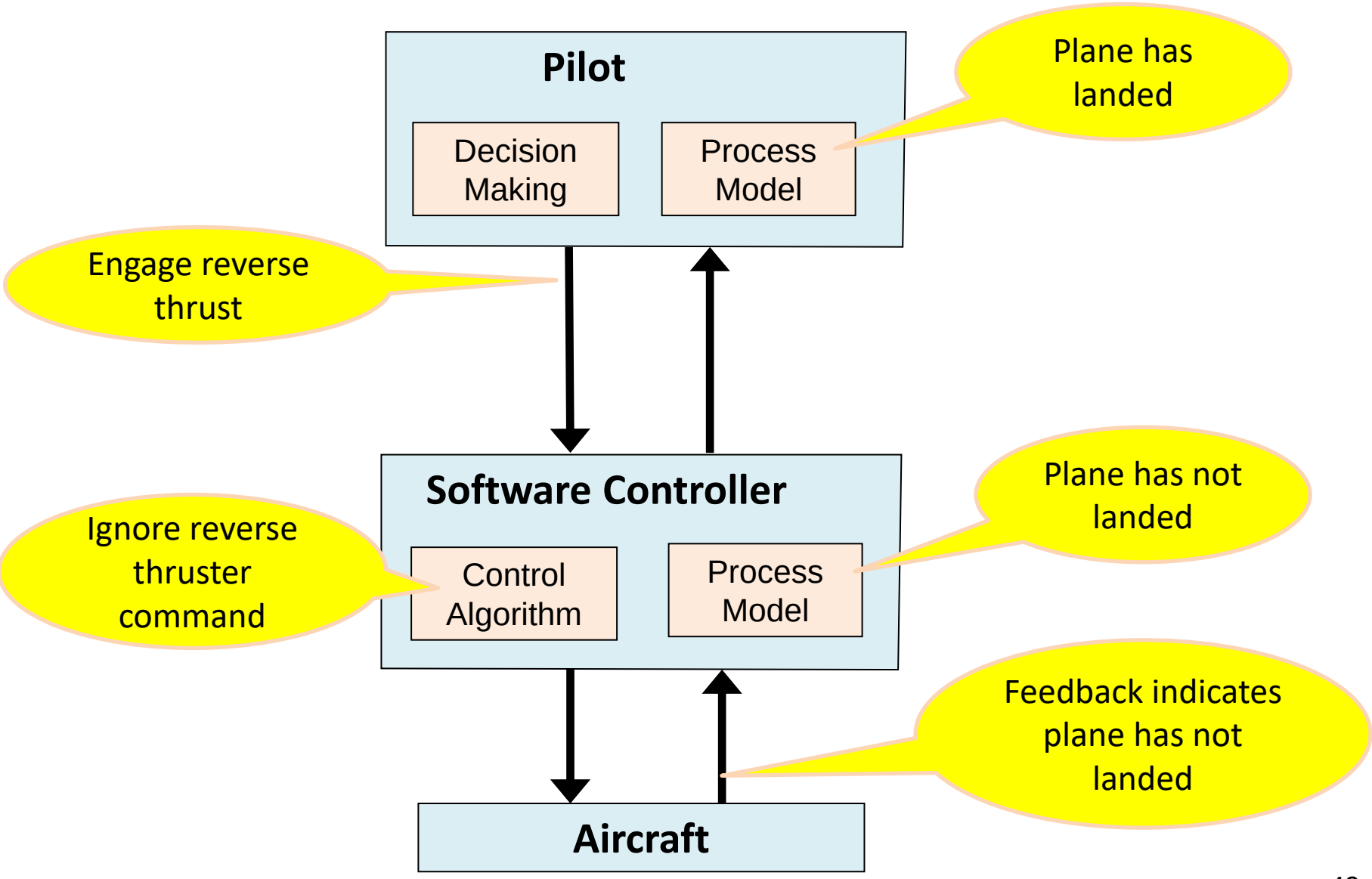
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



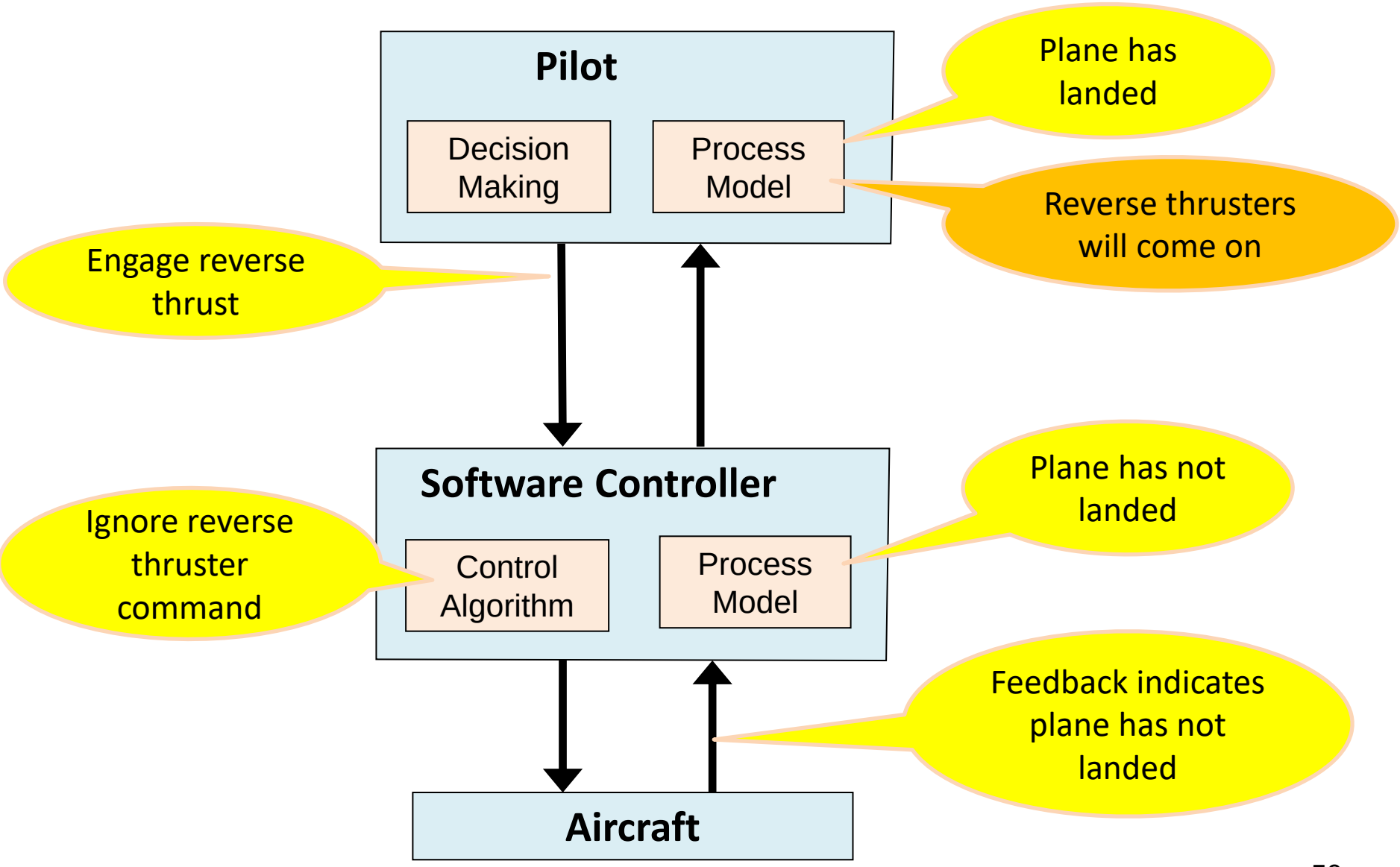
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



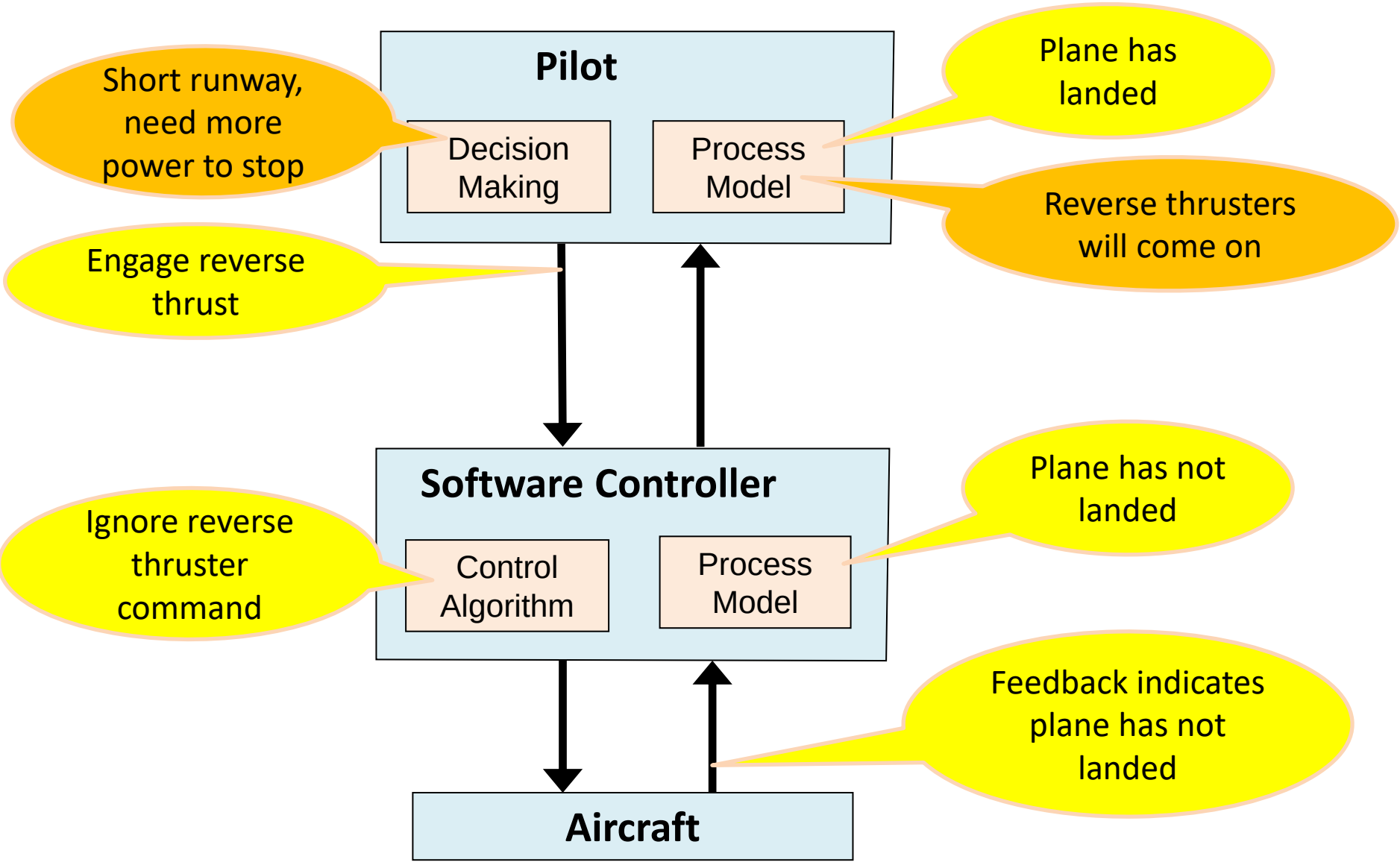
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



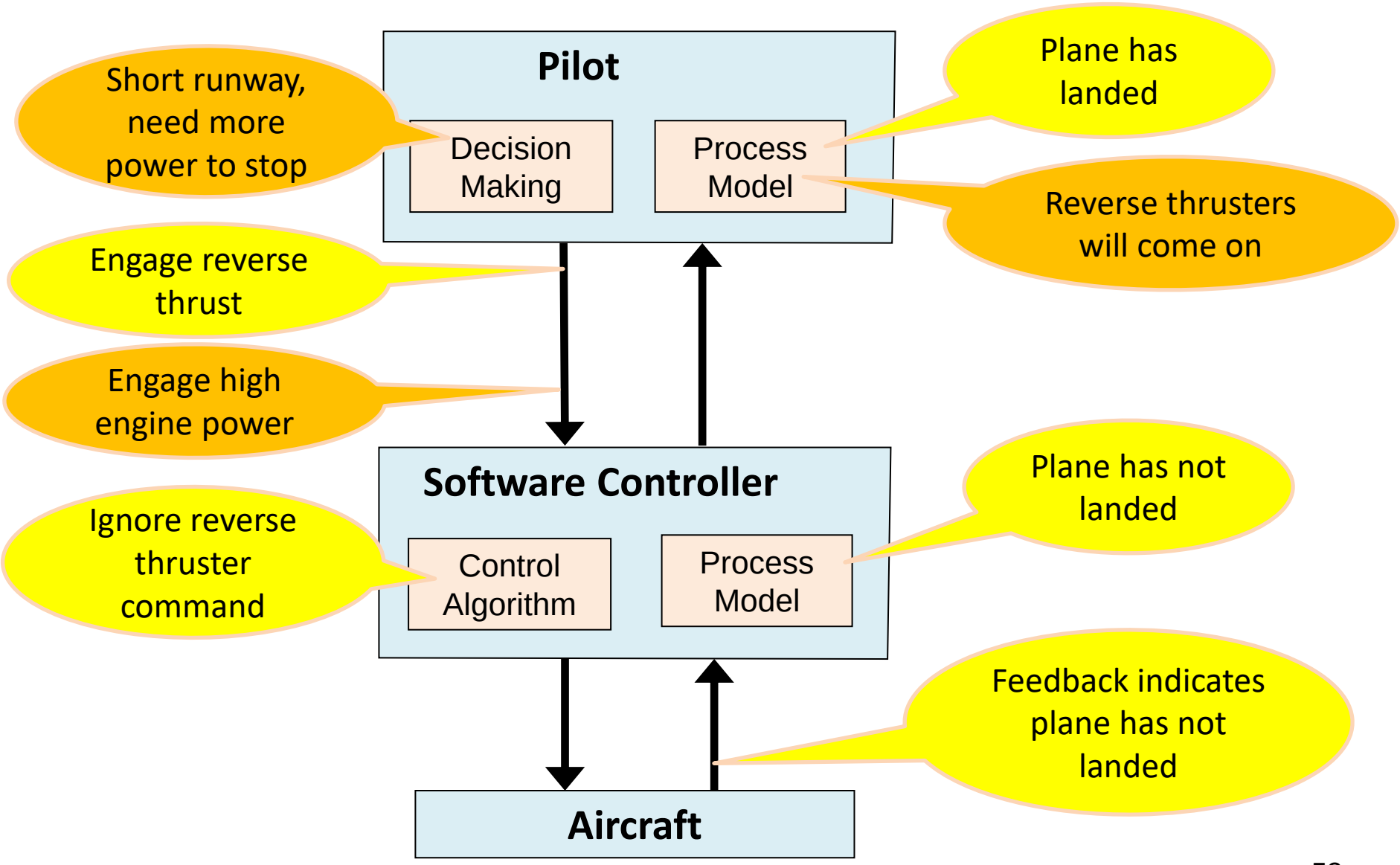
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



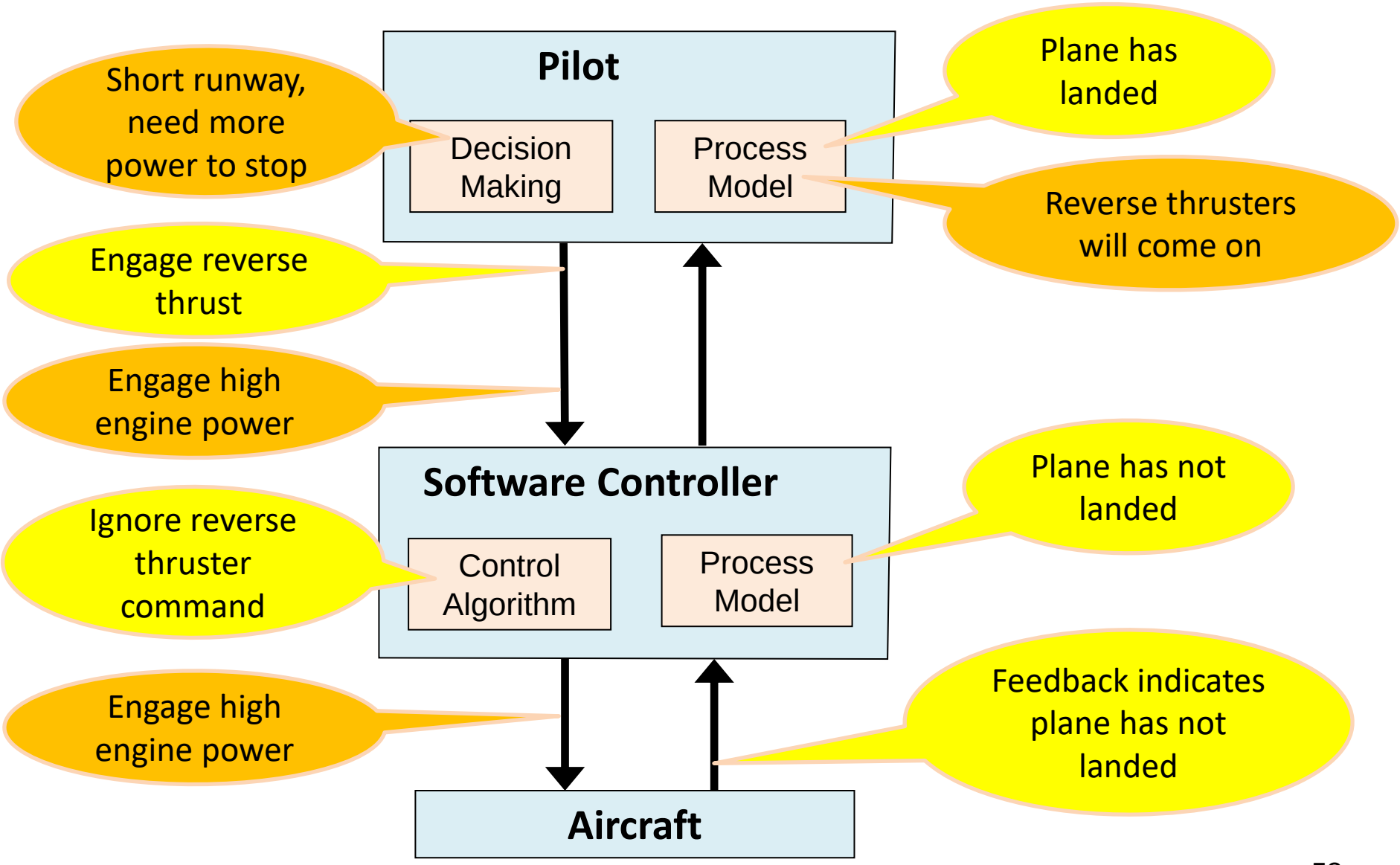
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing

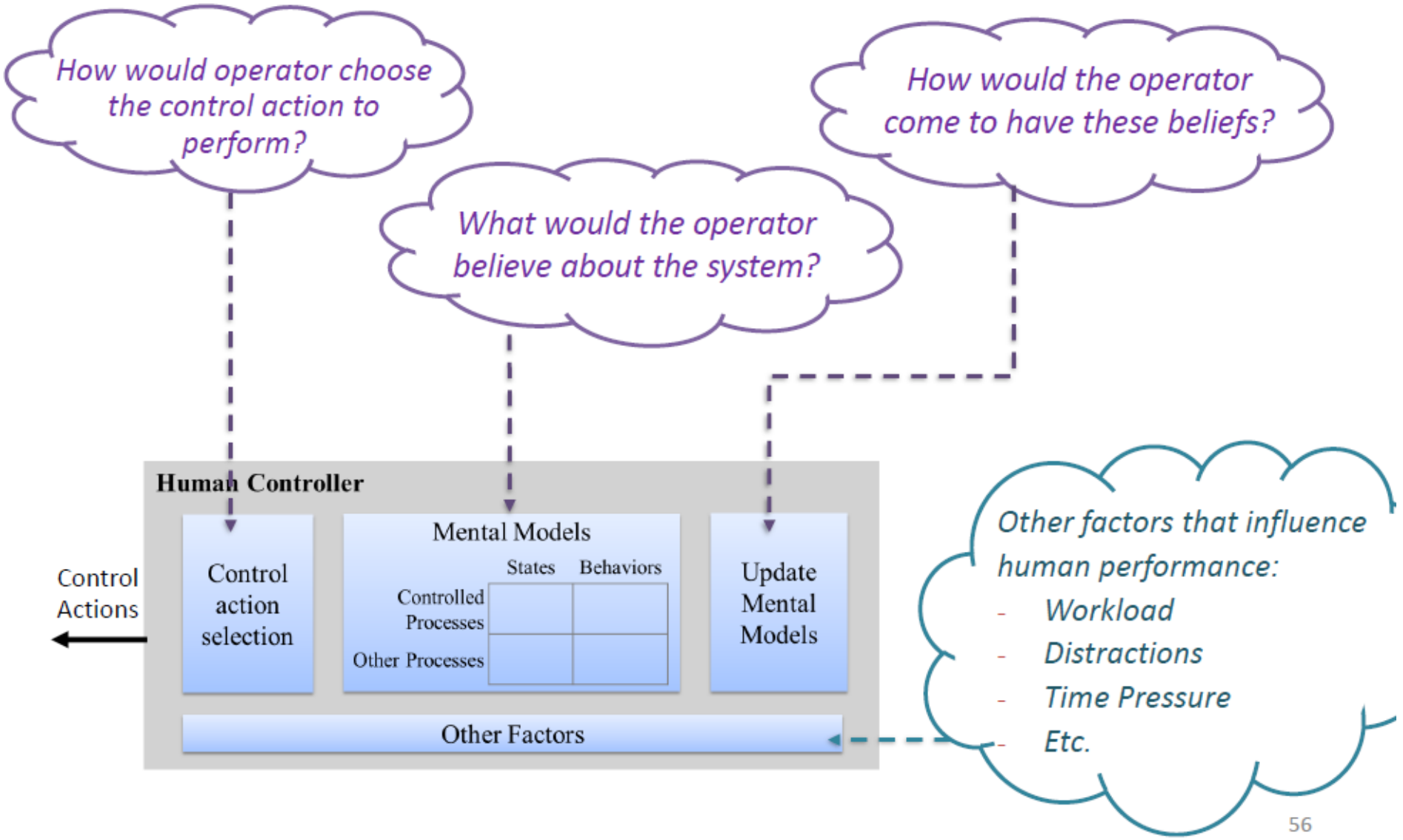


Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



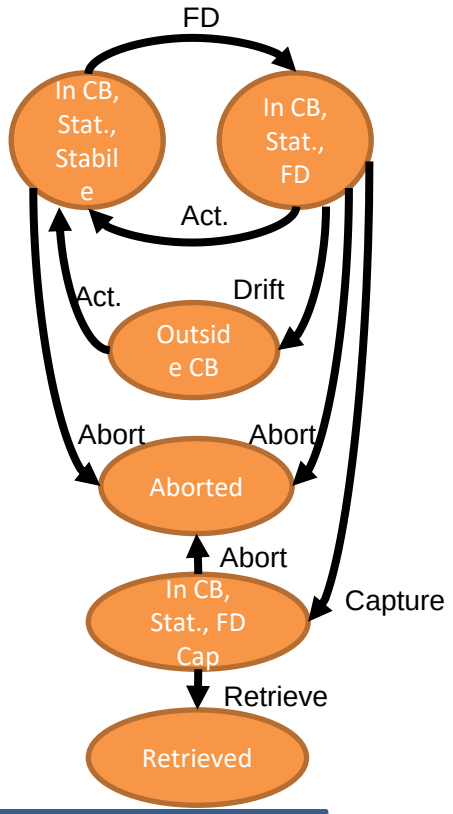
STPA Human Control Model



Model-Based System Engineering (MBSE) and Safety Analysis (MBSA) using STPA

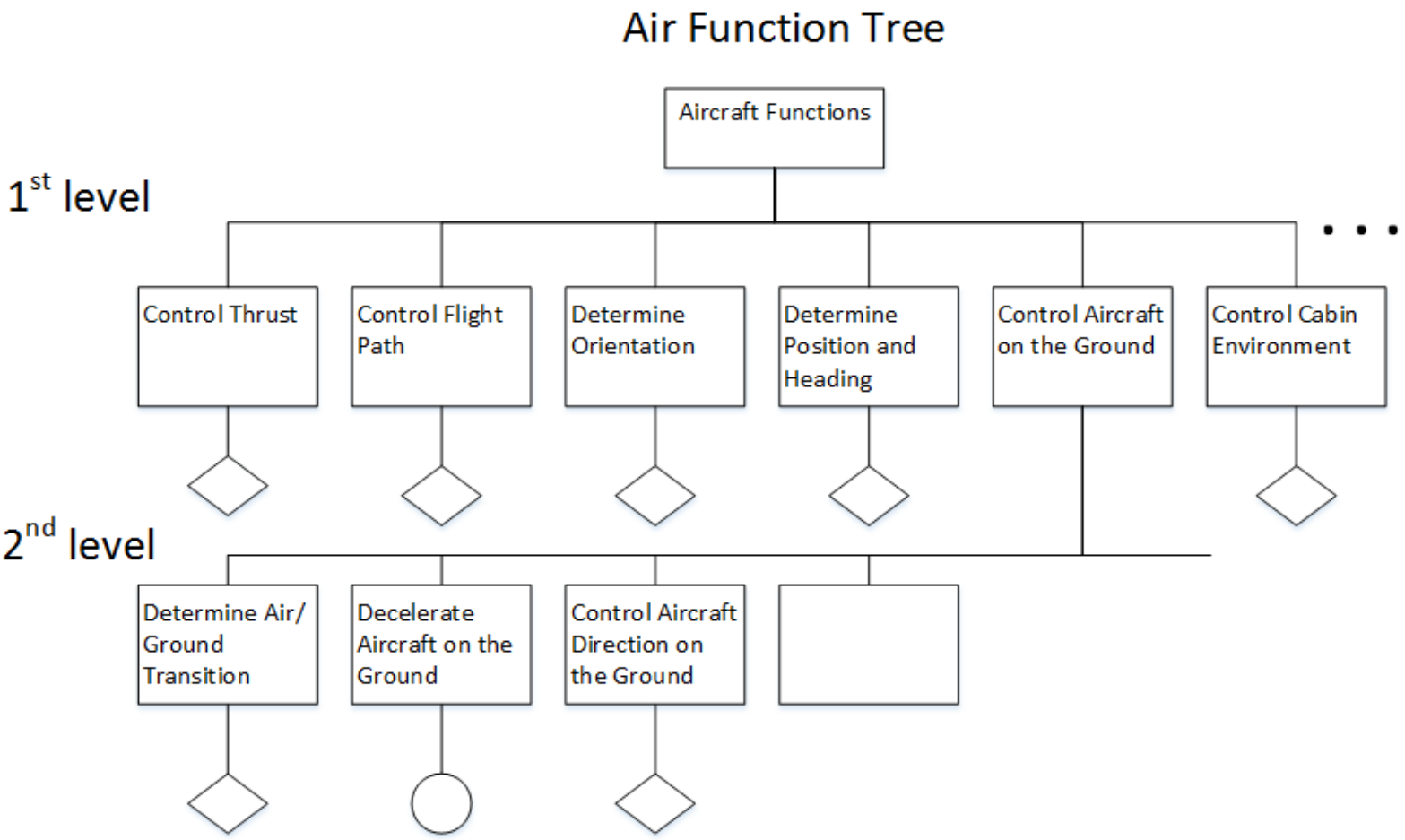
1. Describing hazardous, functional, and required behavior
   - HP(h ∈ H, ca ∈ CA, c ∈ Co)
     - True iff providing command ca in context c will cause hazard h
   - HNP(h ∈ H, ca ∈ CA, c ∈ Co)
     - True iff not providing command ca in context c will cause hazard h
   - FP(f ∈ F, ca ∈ CA, c ∈ Co)
     - True iff providing command ca in context c is necessary to achieve function f
   - R(ca ∈ CA, c ∈ Co)
     - True iff command CA is required to be provided in context c
2. Consistency checks
   - ∀h1 ∈ H,h2 ∈ H ¬ ∃ ca ∈ CA, c ∈ C : HP(h1, ca, c) ^ HNP(h2, ca, c)
     - For every potential context, it must be possible to avoid hazardous control actions/inactions. In other words, if it is hazardous to provide CA then it should be non-hazardous to not provide CA
   - ∀ h ∈ H, f ∈ F ¬ ∃ ca ∈ CA, c ∈ C : HP(h, ca, c) ^ F(f, ca, c)
     - For every potential context, if it is necessary to provide a command to fulfill a function then it must not be hazardous to provide the command in that context
3. Requirements generation (SpecTRM-RL tables)
   - Compute R(ca ∈ CA, c ∈ C) to satisfy the following:
   - ∀h,ca,c: h ∈ H ^ ca ∈ CA ^ c ∈ C → [HP(h, ca, c) → ¬ R(ca, c)]
   - ∀h,ca,c: h ∈ H ^ ca ∈ CA ^ c ∈ C → [R(ca, c) → HNP(h, ca, c)]
   - ∀f,ca,c: f ∈ F ^ ca ∈ CA ^ c ∈ C → [FP(f, ca, c) → R(ca, c)]

Generated requirements / initial model for HTV / ISS crew interaction



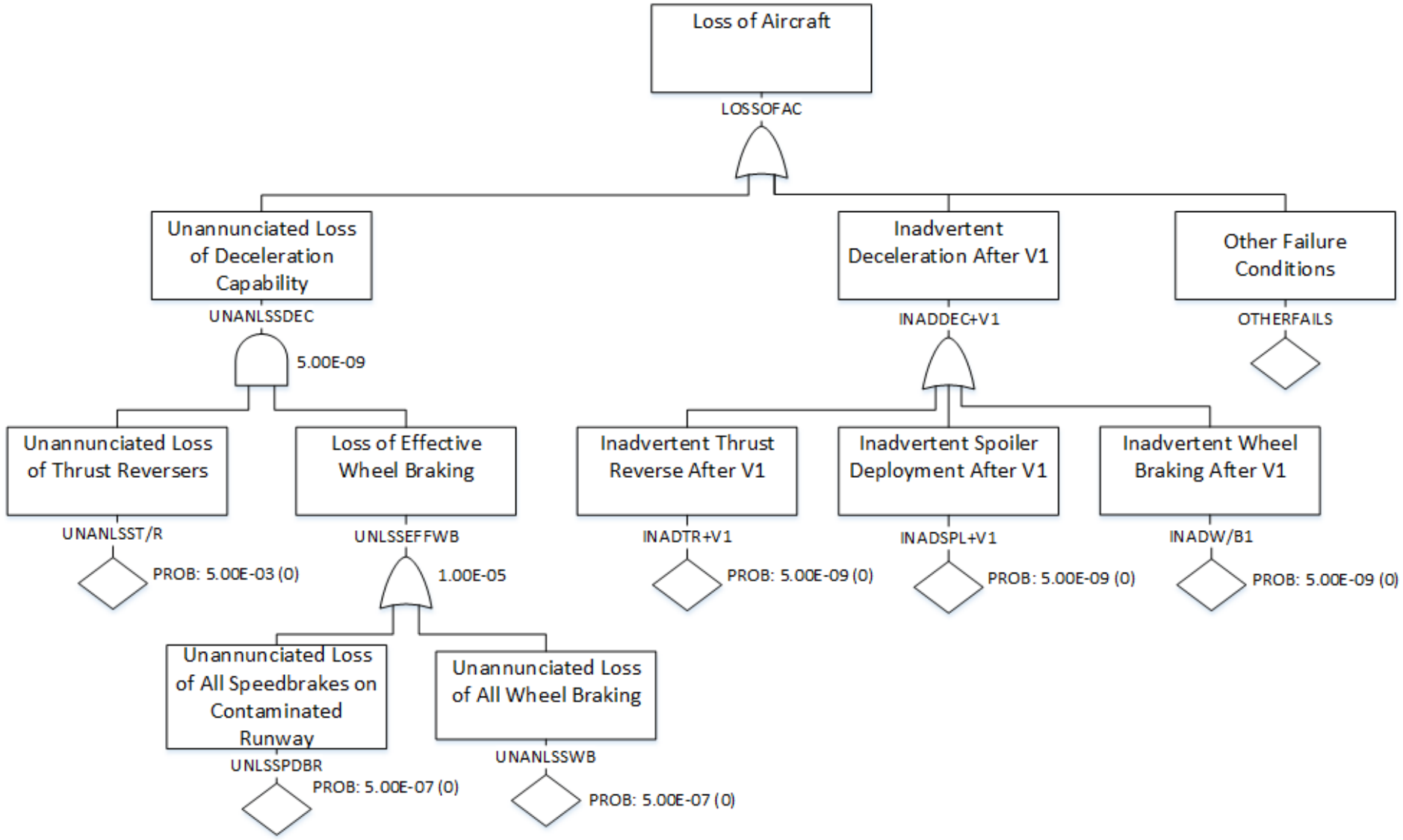
STPA used to automatically generate suitable models (executable)

Typical Decomposition Approach (SAE ARP 4761)



- ARP 4761A adding interactions among “failures” of functions but that is not the problem. Still bottom up.

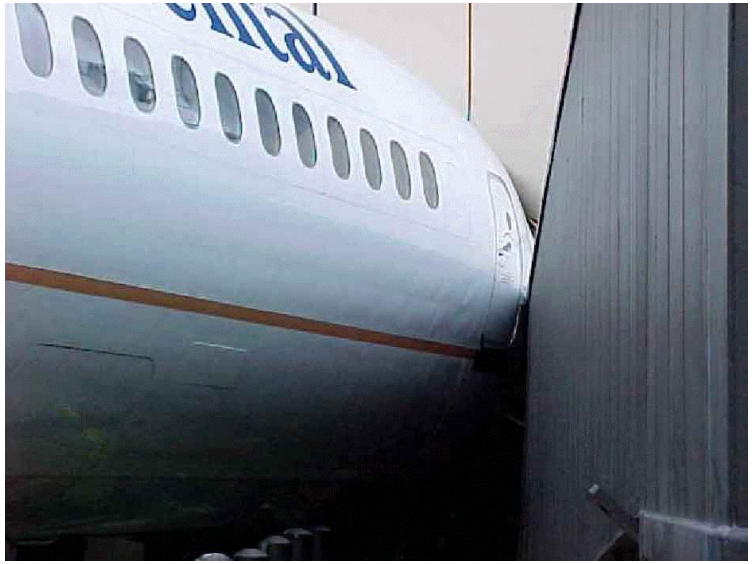
Combine individual component analyses bottom up

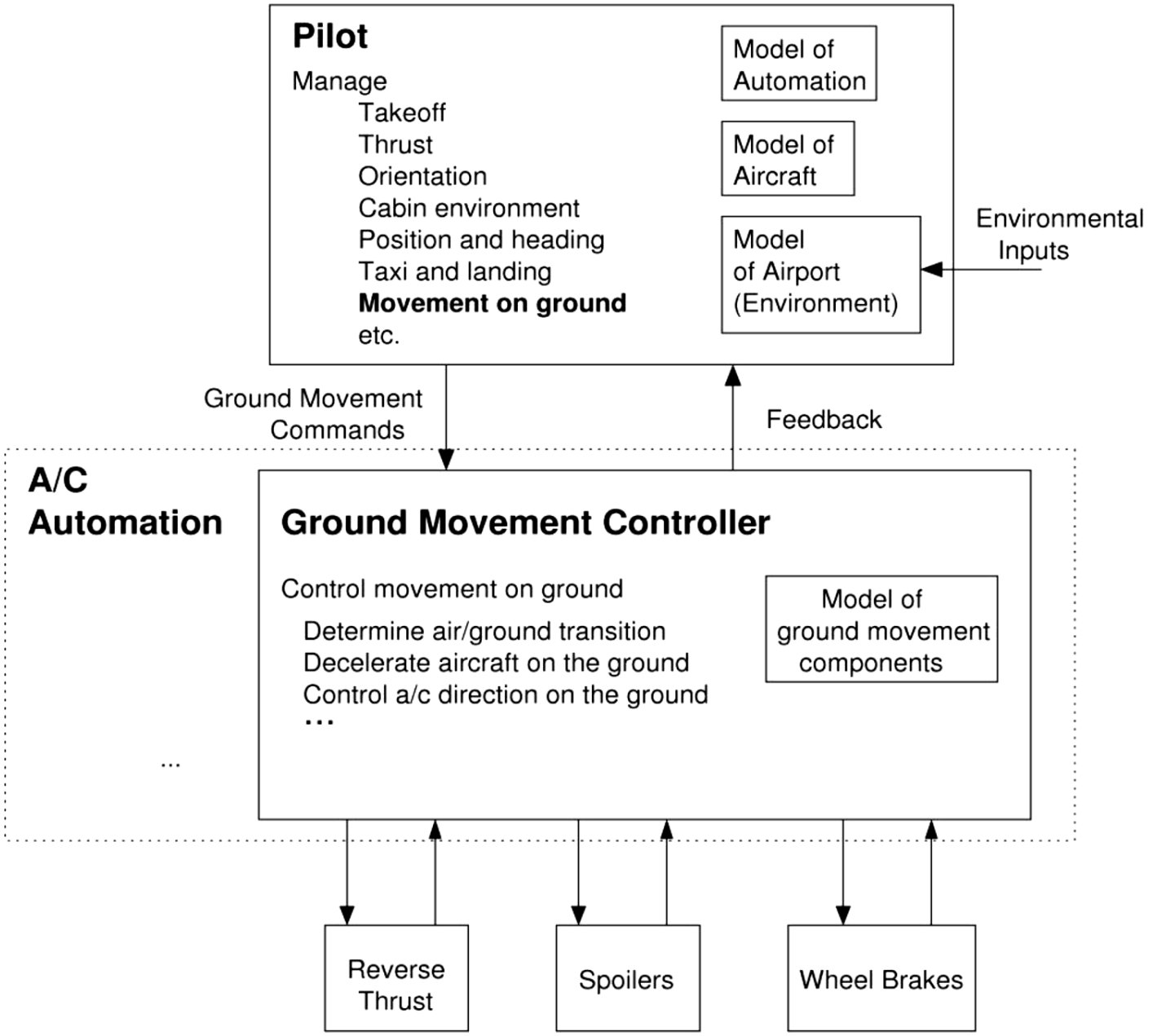


From SAE ARP 4761

Function	Failure Condition (Hazard Description)	Phase	Effect of Failure Condition on Aircraft/Crew	Classification
Decelerate Aircraft on the Ground	Loss of Deceleration Capability	Landing/ RTO/ Taxi		
	…	…	…	
	c. Unannunciated loss of deceleration capability	Taxi	Crew is unable to stop the aircraft on the taxi way or gate resulting In low speed contact with terminal, aircraft, or vehicles	Major
	d. Annunciated loss of deceleration capability	Taxi	Crew steers the aircraft clear of any obstacles and calls for a tug or portable stairs	No Safety Effect

Continental Airlines introduces the improved disembarkation method




Pilot
Manage
Takeoff
Thrust
Orientation
Cabin environment
Position and heading
Taxi and landing
Movement on ground
etc.
Model of Automation
Model of Aircraft
Model of Airport (Environment)
Environmental Inputs
Ground Movement Commands
Feedback
A/C Automation
Ground Movement Controller
Control movement on ground
Determine air/ground transition
Decelerate aircraft on the ground
Control a/c direction on the ground
...
Model of ground movement components
...
Reverse Thrust
Spoilers
Wheel Brakes

Examples of Requirements/Constraints Generated on the Interaction Between Deceleration Components

- **SC-BS-1**: Spoilers must deploy when the wheel brakes are activated manually or automatically above TBD speed.
- **SC-BS-2**: Wheel brakes must activate upon retraction of landing gear.
- **SC-BS-3**: Activation of ground spoilers must activate armed automatic braking (autobrake) system.
- **SC-BS-4**: Automatic braking system must not activate wheel brakes with forward thrust applied.
- **SC-BS-5**: Automatic spoiler system must retract the spoilers when forward thrust is applied.

New Research Directions (UAM and FVL)

Conceptual Architectures

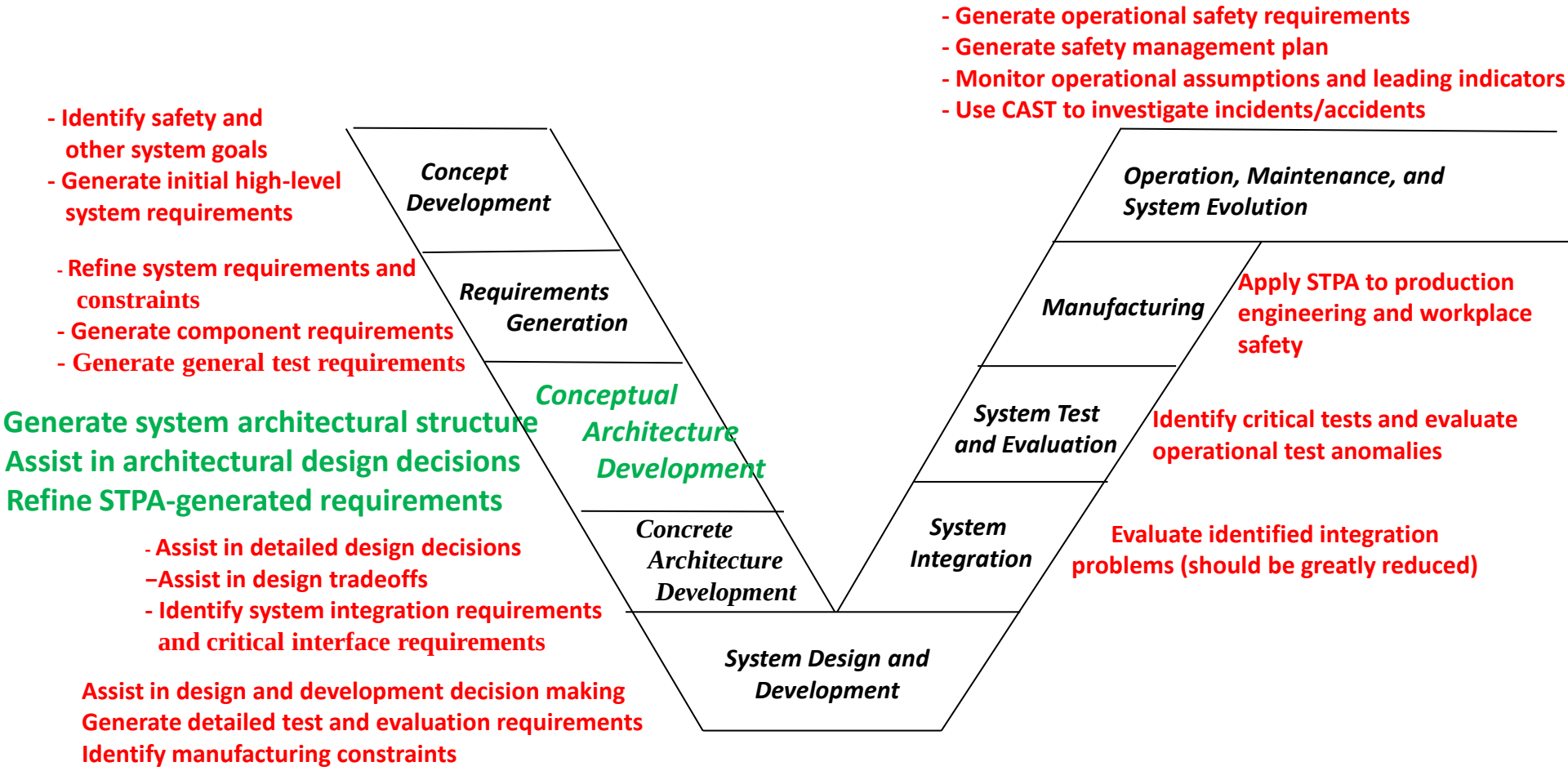
The problem:

- Architectures often created before
  - Know all requirements and constraints
  - Independent of specific system requirements or constraints
- Often just use standard architectures
  - Not necessarily reflective of system type being developed
  - Reflect some goals/constraints but not others
- Often dive into details prematurely
  - Decompose into standard functional components
  - Specify logical and physical details of connections between components (network structures, interface specs)
  - Design physical interactions before know what connections are important or needed
  - Usually little or no tracing to detailed requirements and desired system-level properties

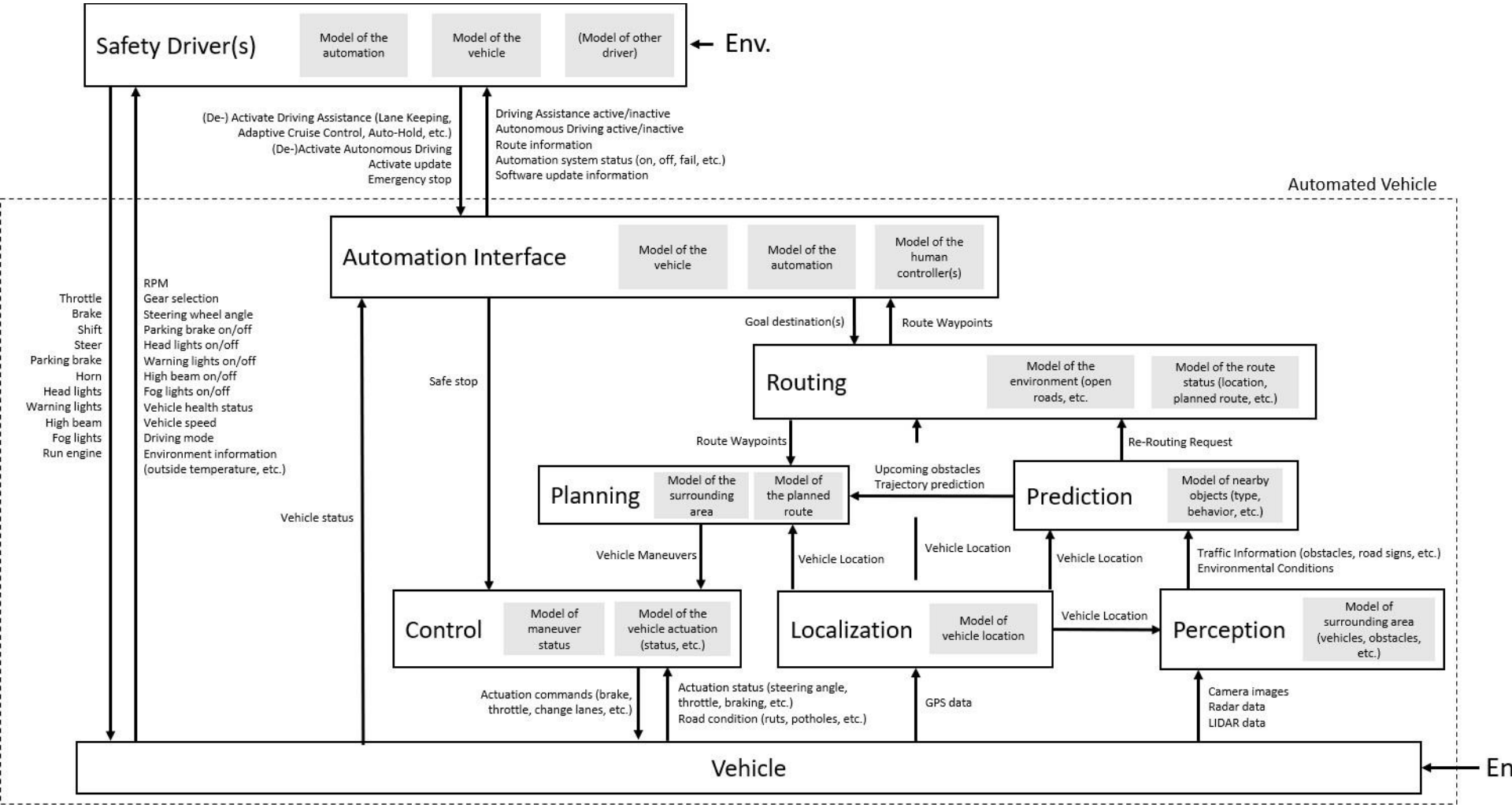
The Problem in Architectural Development (2)

- Results in systems where unique requirements only vaguely tied to architecture
- Safety engineering efforts reduced to producing a lot of paper with no real impact on actual system design
- Security efforts delayed until little impact on system design and system cannot be protected against adversaries
- Makes it harder and costlier to ensure safety/security/etc. are satisfied by implementation if haven't designed these qualities into the system from beginning and may even be infeasible
- Maintenance and upgrades may be enormously expensive

New step in V-model: Conceptual Architecture Development

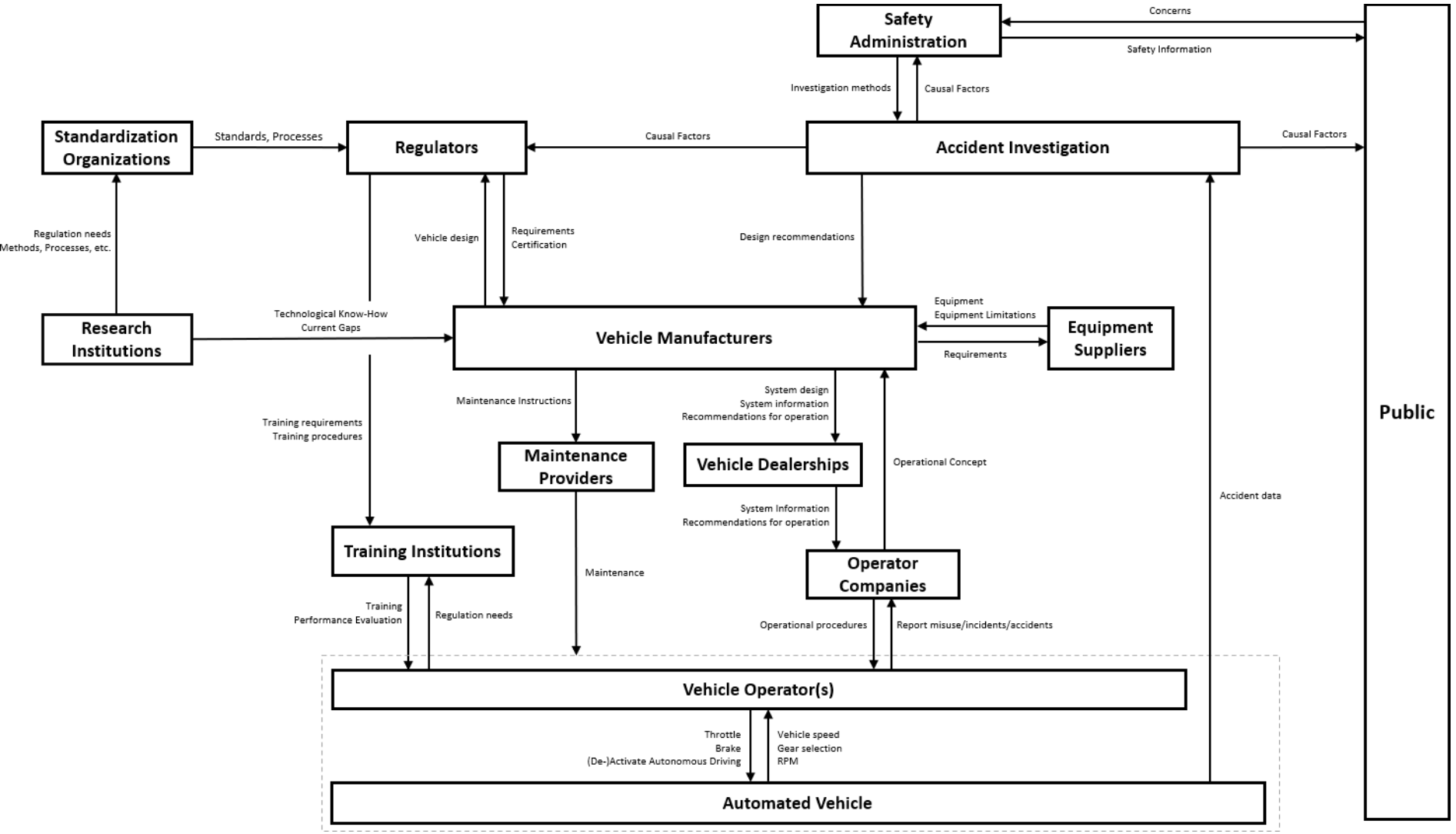


- **Bridges gap between "shall statements" and detailed architecture development**
- **Provides concrete tracing between requirements and design**

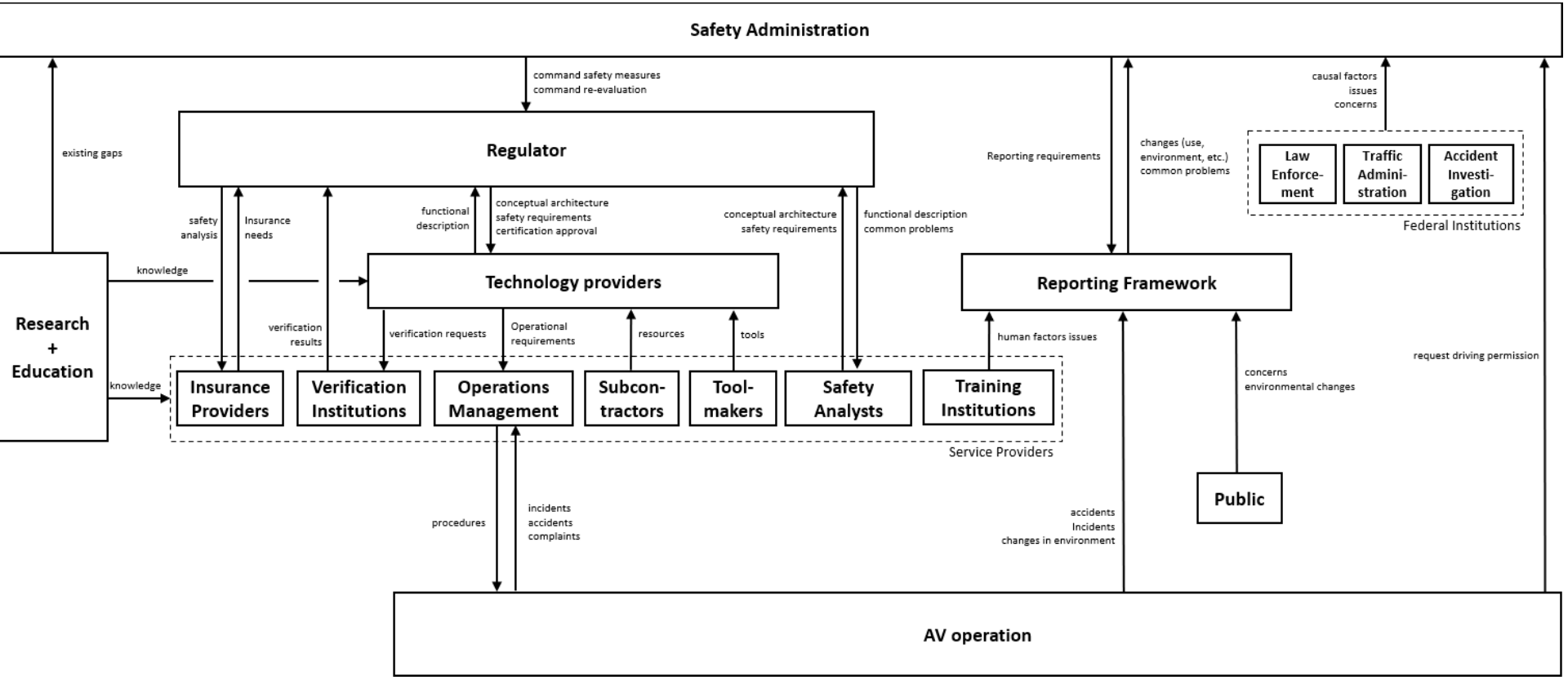


Conceptual architecture model for AVs

Michael Schmid

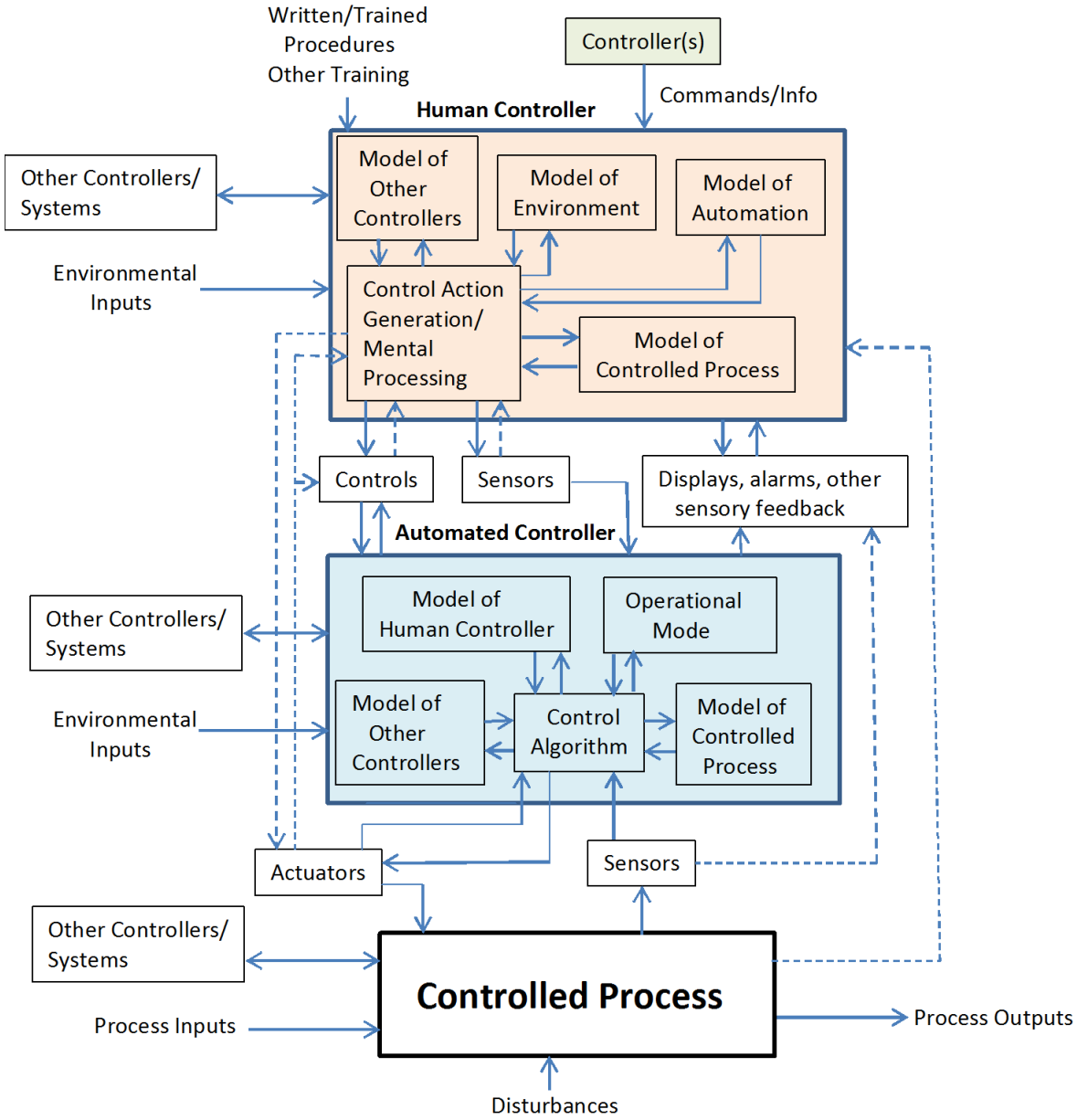


Socio-technical control structure for AV safety



Control Structure for AV Certification

General Format of Conceptual Architecture (Control Structure)



Start at high level of abstraction and refine using results of STPA and other analyses

Note potential for human-centered design

Human-Automation Teaming (Col. Kip Johnson)

- Johnson (2017) defined a theoretical basis for coordination between humans and automation.
  - Coordination involves mutually beneficial interactions (vs. master-slave)
  - Interdependency may arise from organizational, temporal, reciprocal, and shared resource conditions.
  - To manage interdependencies, need dynamic strategies that complicate analysis.
- Case Studies: Patriot Friendly Fire

  UAS in Commercial Airspace

Want to:

- Create extensions to STPA to assure safety and security of advanced coordination strategies and demonstrate their effectiveness.

Improving Standard Risk Assessment

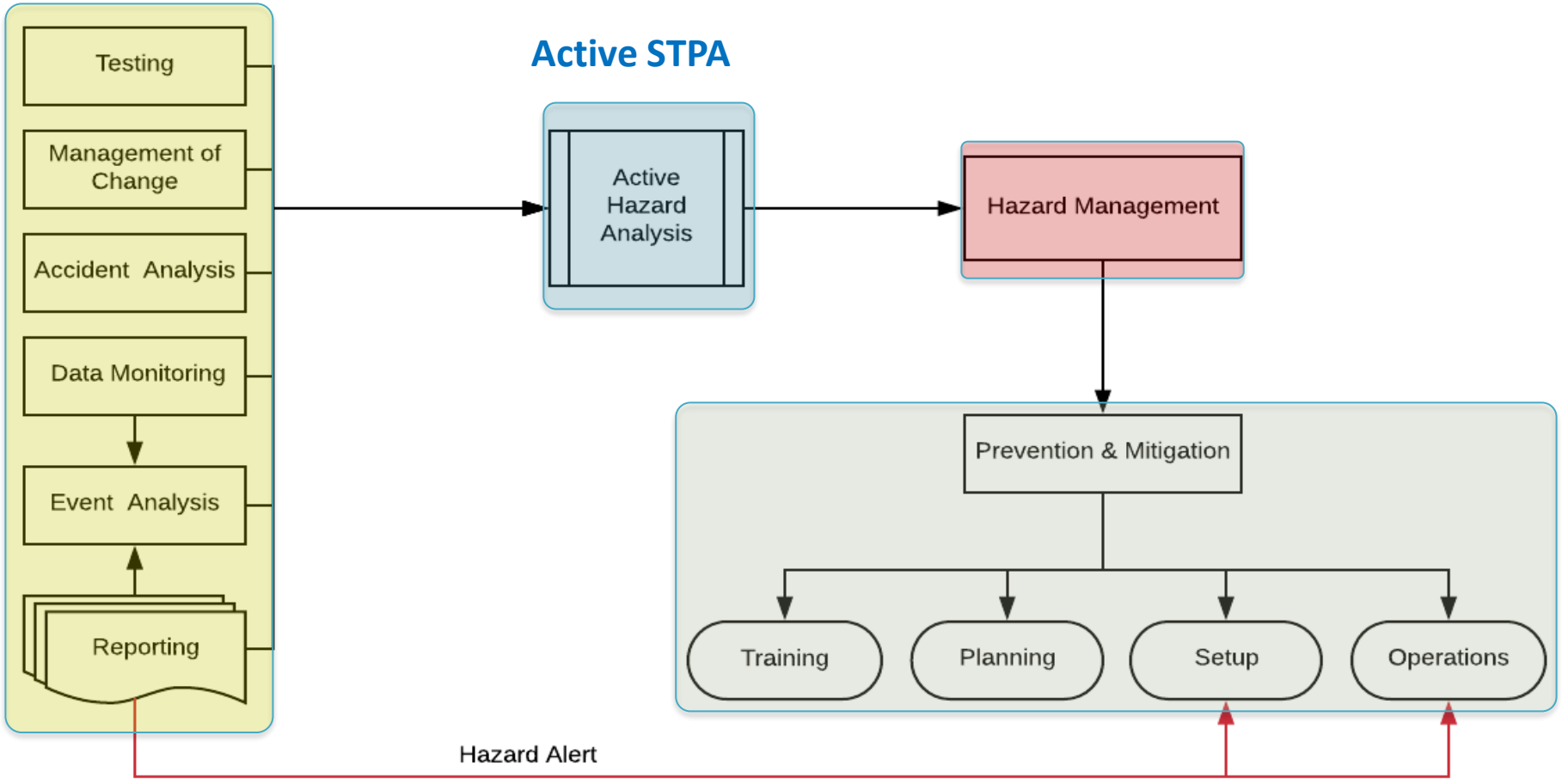
- Goal: Use STAMP to get better estimates of “likelihood”

RISK ASSESSMENT MATRIX				
SEVERITY / PROBABILITY	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
Frequent (A)	High	High	Serious	Medium
Probable (B)	High	High	Serious	Medium
Occasional (C)	High	Serious	Medium	Low
Remote (D)	Serious	Medium	Medium	Low
Improbable (E)	Medium	Medium	Medium	Low
Eliminated (F)	Eliminated			

Risk Management During Operations

- Design of Safety Management System
- Leading Indicators of Increasing Risk
  - *“Assumption-based leading indicators”*
  - Create new tools to use during operations

Major Diogo Castilho



Validated on a major international airline

Is STPA Practical?

- STPA has been or is being used in a large variety of industries
  - Automobiles (>80% use)
  - Aircraft and Spacecraft (extensive use and growing)
  - Defense systems (UAVs, AF GBSD, Army FVL, etc.)
  - Ships/Marine
  - Air Traffic Control
  - Medical Devices and Hospital Safety
  - Chemical plants
  - Oil and Gas
  - Nuclear and Electric Power
  - Robotic Manufacturing / Workplace Safety
- 2,316 registrants (73 countries) for STAMP Workshop this year
- New international standards (autos, aircraft, defense) created or in development, or STPA already satisfies (MIL-STD-882)

Evaluations and Estimates of ROI

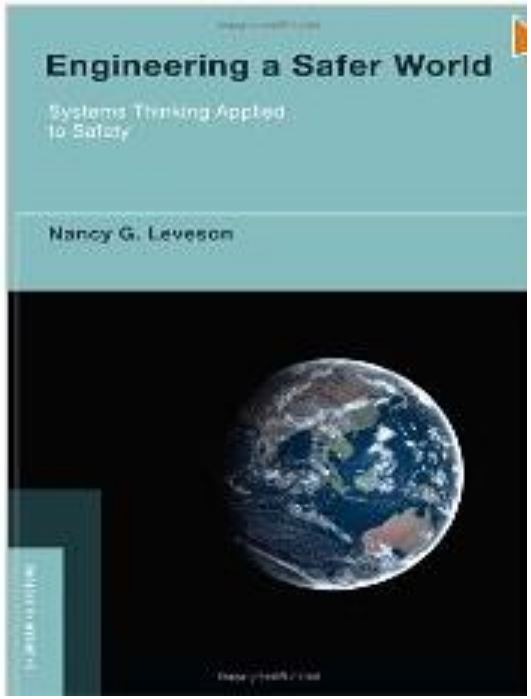
- Hundreds of evaluations and comparison with traditional approaches used now
  - Controlled scientific and empirical (in industry)
  - All show STPA is better (identifies more critical requirements or design flaws)
  - All (that measured) show STPA requires orders of magnitude fewer resources than traditional techniques
- ROI estimates only beginning but one large defense industry contractor claims they are seeing 15-20% savings on whole contract cost when using STPA

Summary: A Systems Approach to Safety and Security

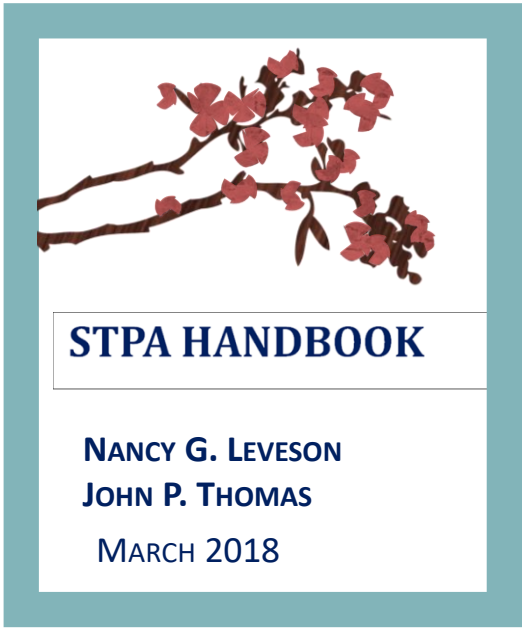
- Emphasizes building in safety/security rather than measuring it or assuring it (can start early in concept formation stage)
- Looks at system as a whole, not just components (a top-down holistic approach)
- Takes a larger view of causes than just failures
  - Accidents today are not just caused by component failures
  - Includes software and requirements flaws, cognitively complex human decision making, design flaws, etc.
  - Treats safety/security as a control problem, not a failure problem
- Goal is to use modeling and analysis to design and operate the system to be safe/secure, not to predict the likelihood of a loss or provide after the fact assurance.

More Information

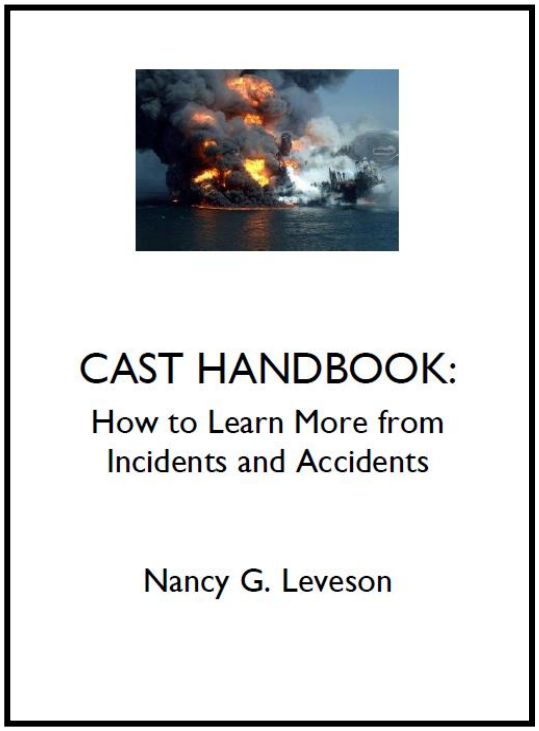
- http://psas.scripts.mit.edu (papers, presentations from conferences, tutorial slides, examples, etc.)



Free download:
http://mitpress.mit.edu/books/engineering-safer-world



http://psas.scripts.mit.edu
(**80,000+** downloads in 30 mos. Japanese, Chinese, and Korean versions)



Free download:
http://sunnyday.mit.edu/CAST-Handbook.pdf

BACKUP

Ballistic Missile Defense System (MDA)

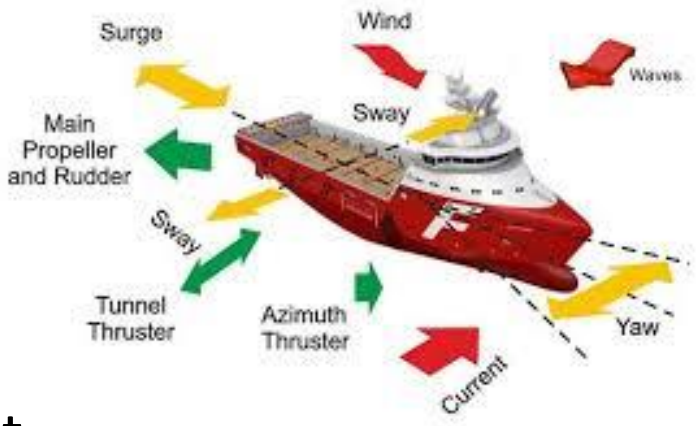
- Hazard was inadvertent launch
- Analyzed right before deployment and field testing (so done late)
  - 2 people, 5 months (unfamiliar with system)
  - Found so many paths to inadvertent launch that deployment delayed six months
- One of first uses of STPA on a real defense system (2005)

Sea-based sensors on the Aegis platform, upgraded early warning radars (UEWR), the Cobra Dane Upgrade (CDU), Ground-based Midcourse Defense (GMD) Fire Control and Communications (GFC/C), a Command and Control Battle Management and Communications (C2BMC) Element, and Ground-based interceptors (GBI). Future block upgrades were originally planned to introduce additional Elements into the BMDS, including Airborne Laser (ABL) and Terminal High Altitude Area Defense (THAAD).

Example Hazard Scenarios Found

- Missing software and human operator requirements, for example:
  - Operator could input a legal (but unanticipated) instruction at same time that radars detect a potential (but not dangerous) threat
  - Could lead to software issuing an instruction to enable firing an interceptor at a non-threat
- Timing conditions that could lead to incorrectly launching an interceptor
- Situations in which simulator data could be taken as real data

Navy Escort Vessels (Lt. Blake Abrecht)



- Dynamic positioning system
- Ran into each other twice during test
- Performed a CAST analysis (on two incidents) and STPA on system as a whole
- STPA found scenarios not found by MIL-STD-882 analysis (fault trees and FMECA)
- Did not implement our findings: “We’ve used PRA for 40 years and it works just fine”
- Put into operation and within 2 months ran into a submarine
- Scenario was one we had found

UH-60MU (Blackhawk)

- Analyzed Warning, Caution, and Advisory (WCA) system

- STPA results were compared with an independently conducted hazard analysis of the UH-60MU using traditional safety processes described in SAE ARP 4761 and MIL-STD-882E.
  - STPA found the same hazard causes as the traditional techniques and
  - Also identified things not found using traditional methods, including design flaws, unsafe human behavior, and component integration and interactions flaws

UH-60MU SAR Hazard Classification

UH-60MU SAR marginal hazards

- Loss of altitude indication in DVE
- Loss of heading indication in DVE
- Loss of airspeed indication in DVE
- Loss of aircraft health information
- Loss of external communications
- Loss of internal communications

UH-60MU SAR identified various hazards as marginal that could lead to a catastrophic accident

STPA Unsafe Control Action

The Flight Crew does not provide collective control input necessary for level flight, resulting in controlled flight into terrain

Scenario 1: The Flight Crew has a flawed process model and believes they are providing sufficient control input to maintain level flight. This flawed process model could result from:

a) The altitude indicator and attitude indicator are malfunctioning during IFR flight and the pilots are unable to maintain level flight

b) The Flight Crew believes the aircraft is trimmed in level flight when it is not

c) The Flight Crew has excessive workload due to other tasks and cannot control the aircraft

d) The Flight Crew has degraded visual conditions and cannot perceive slow rates of descent that result in a continuous descent

e) The Flight Crew does not perceive rising terrain and trims the aircraft for level flight that results in controlled flight into terrain

UH-60MU SAR Failure based Hazards

UH-60MU SAR residual hazard

- **APU Chaffing can lead to failure of the UH-60MU APU and can affect blade deice operations when the loss of a main generator occurs**

STPA Unsafe Control Action

UCA: The Flight Crew does not switch APU generator power ON when either GEN 1 or GEN 2 are not supplying power to the helicopter and the Blade Deice System is required.

Scenario 1: The Flight Crew does not know that APU generator power is needed to run the Blade Deice System. This flawed process model could result from:

a) The ICE DETECTED, MR DEICE FAULT/FAIL, or TR DEICE FAIL cautions are not given to the Flight Crew when insufficient power is available for the Blade Deice System

b) The Flight Crew does not know that two generators are not providing power to the Blade Deice System

c) The Flight Crew acknowledged the GEN1 or GEN 2 Fail cautions prior to needing the Blade Deice system and failed to start the APU GEN when the additional power was required for the Blade Deice System

STPA identifies non-failure scenarios that can lead to a hazardous system state that are not identified by traditional hazard analysis techniques

EPRI Evaluation

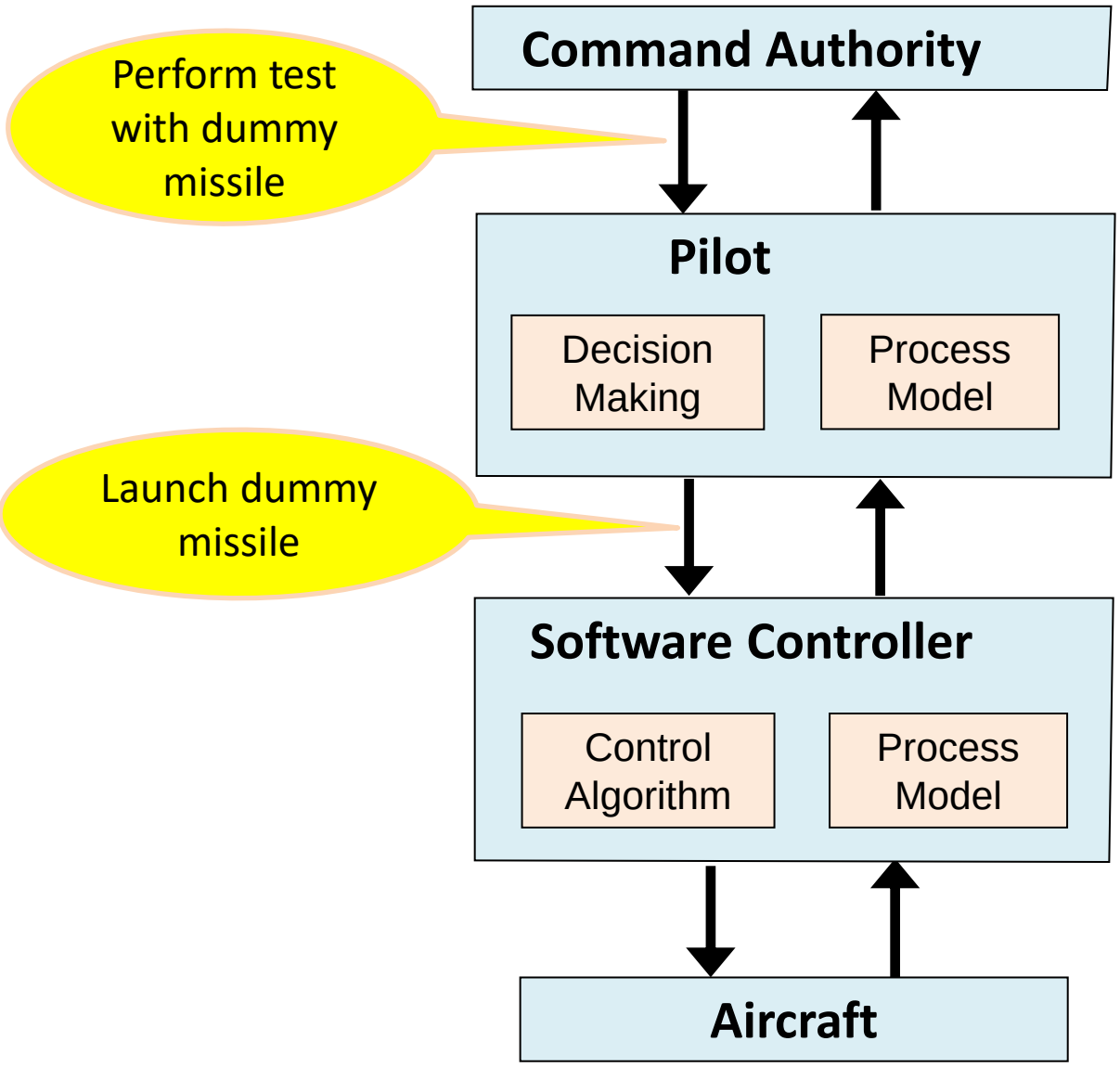
- Same design of a nuclear power plant safety system provided to everyone
- Independent and expert teams did: FTA, ETA, FMEA, HAZOP, etc. and we did STPA (two students, two weeks)
- After submitting final analyses, teams were told that there had been a very serious event in plant with that design
- Only STPA found the scenario that had occurred

New EPRI Study

- Learnability (how much time before can find serious problems)
- Found serious design errors in 2-day beginner class

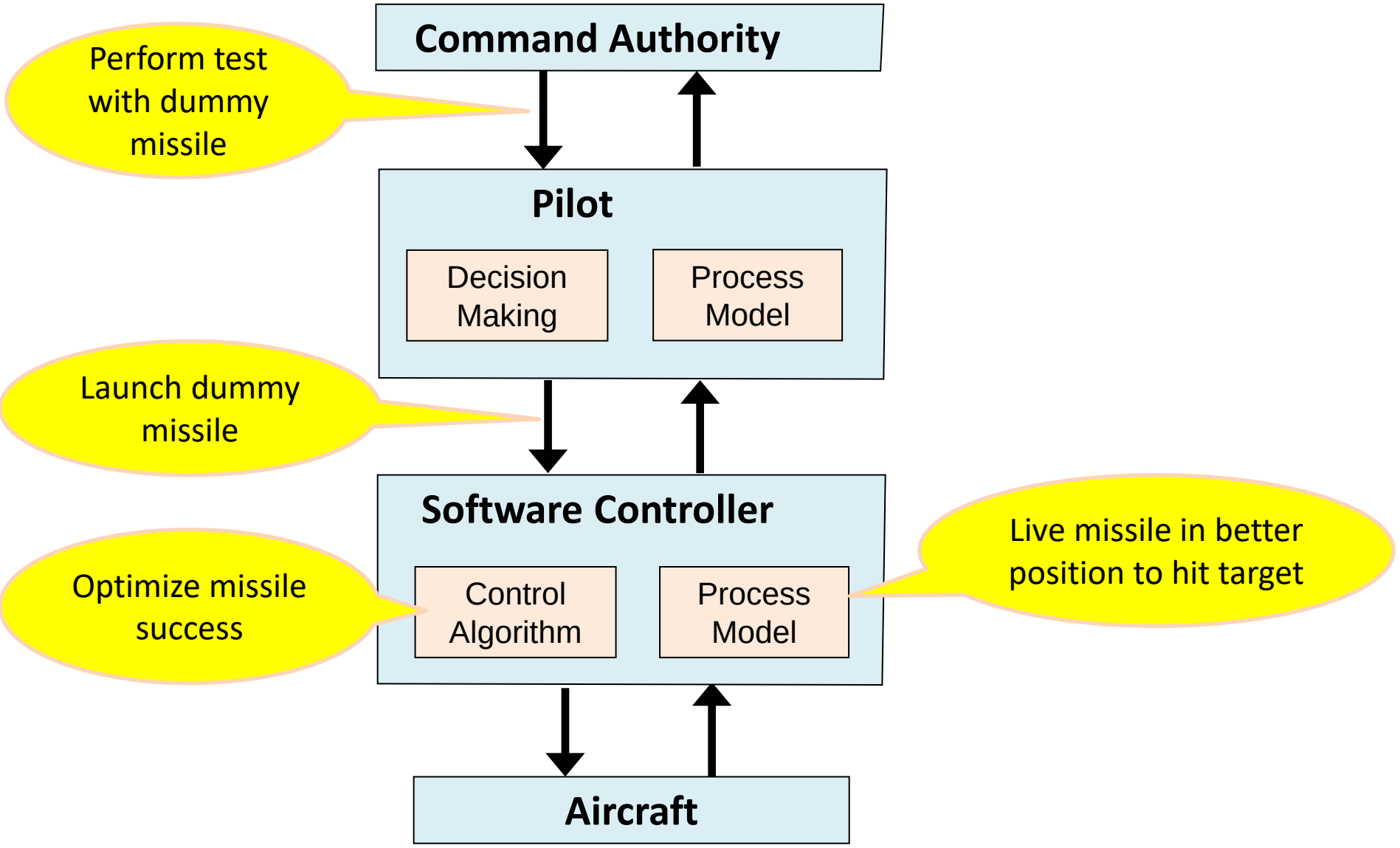
Missile Release Mishap

Hazard: Friendly Fire



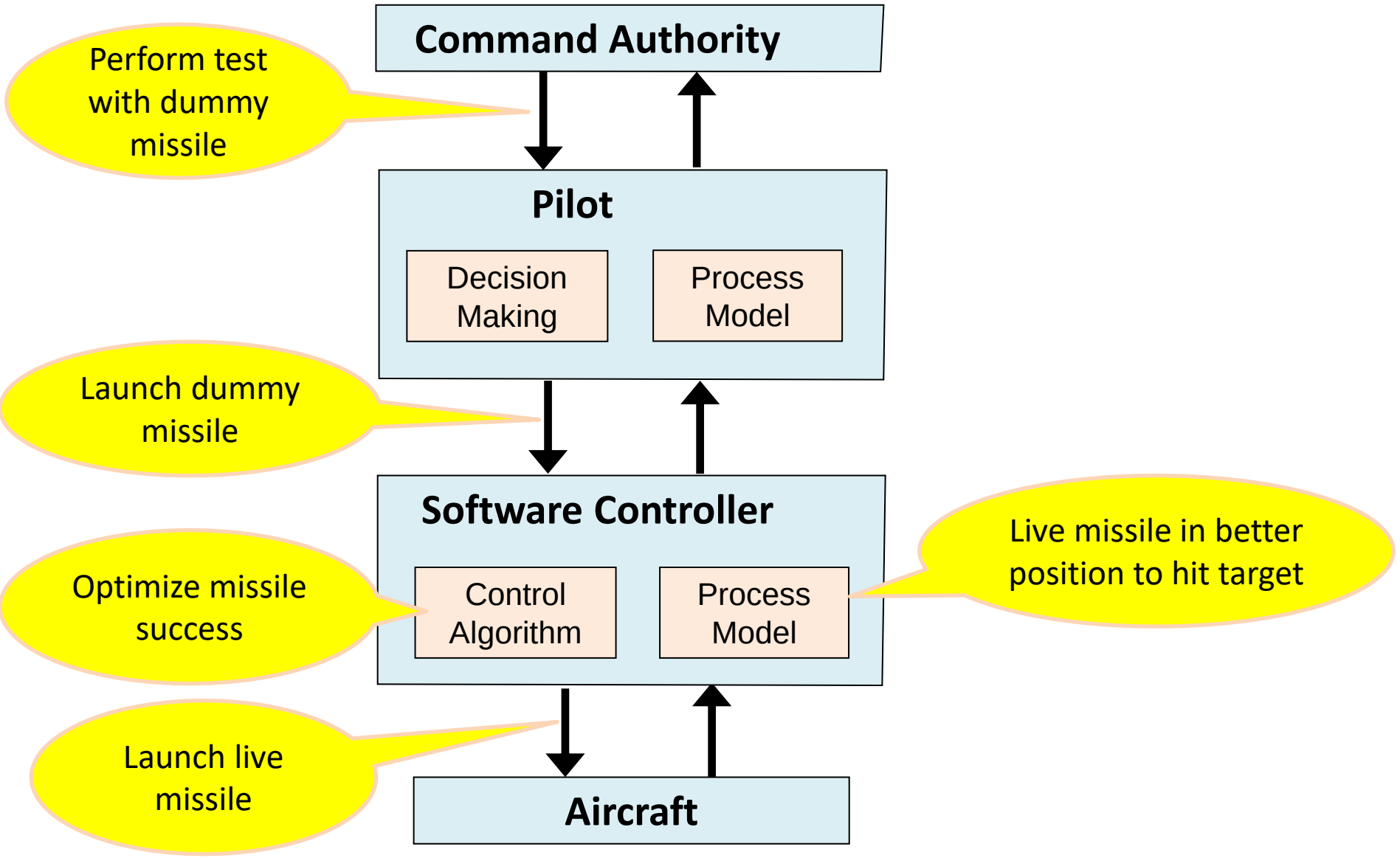
Missile Release Mishap

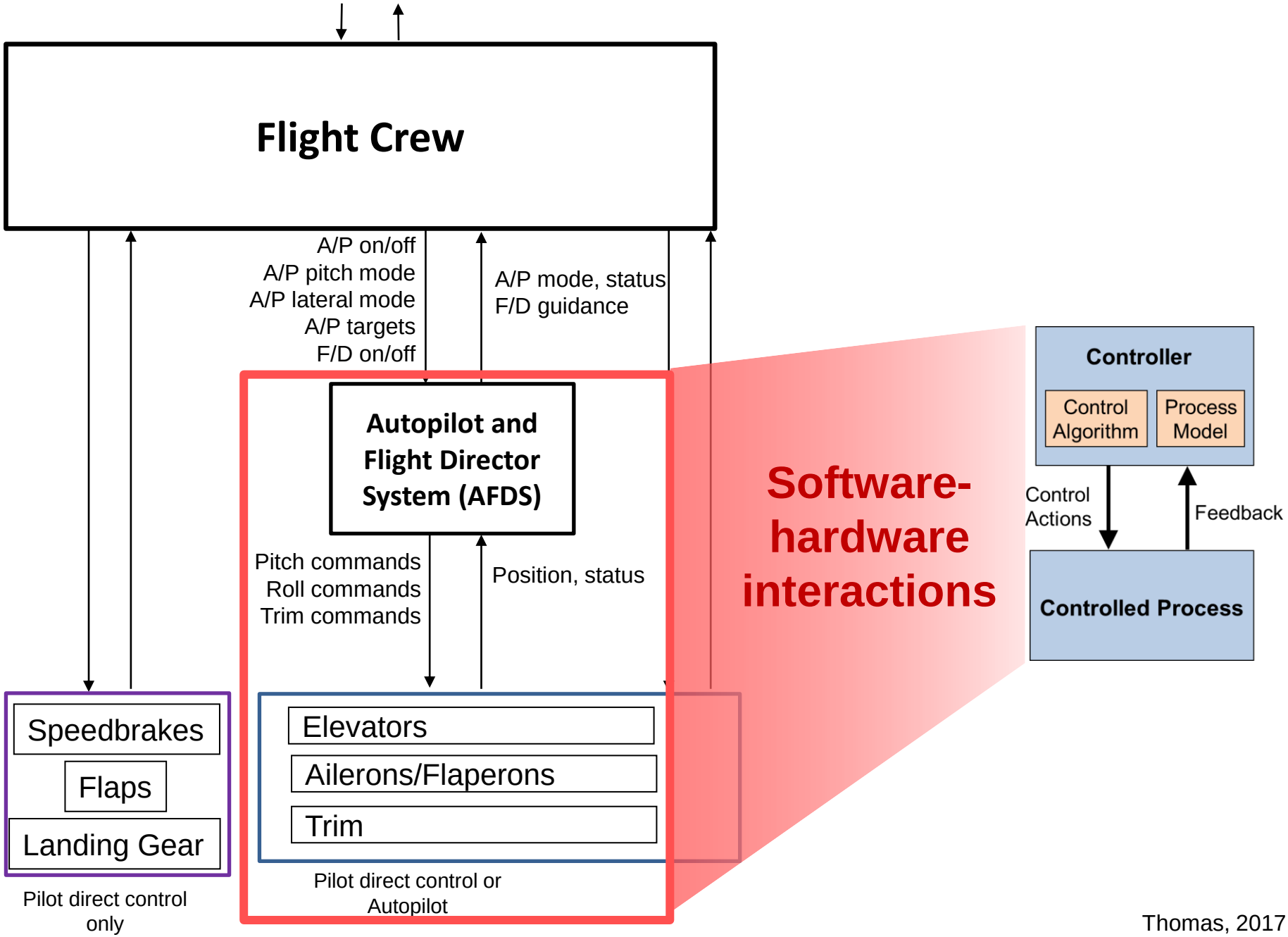
Hazard: Friendly Fire



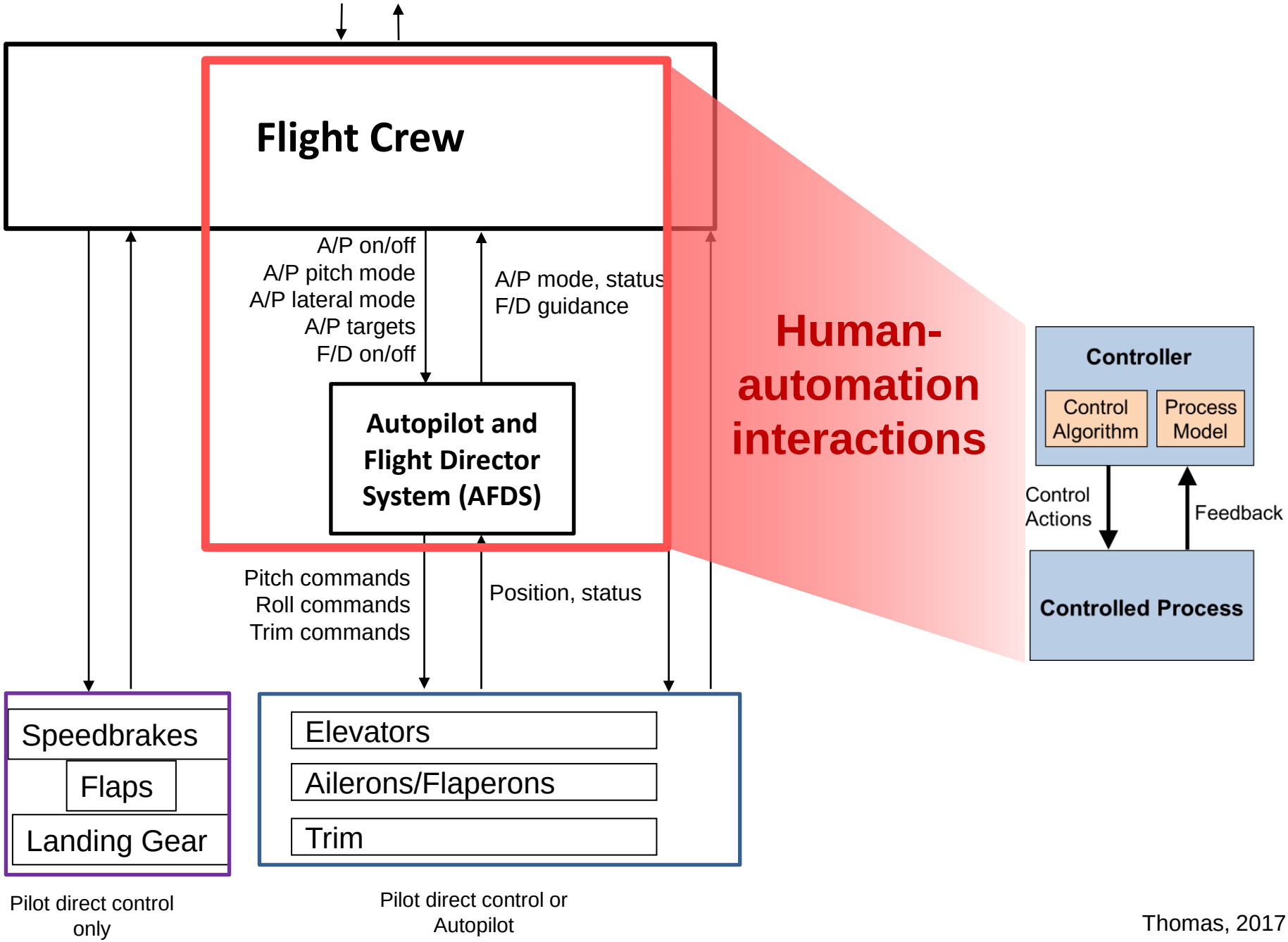
Missile Release Mishap

Hazard: Friendly Fire

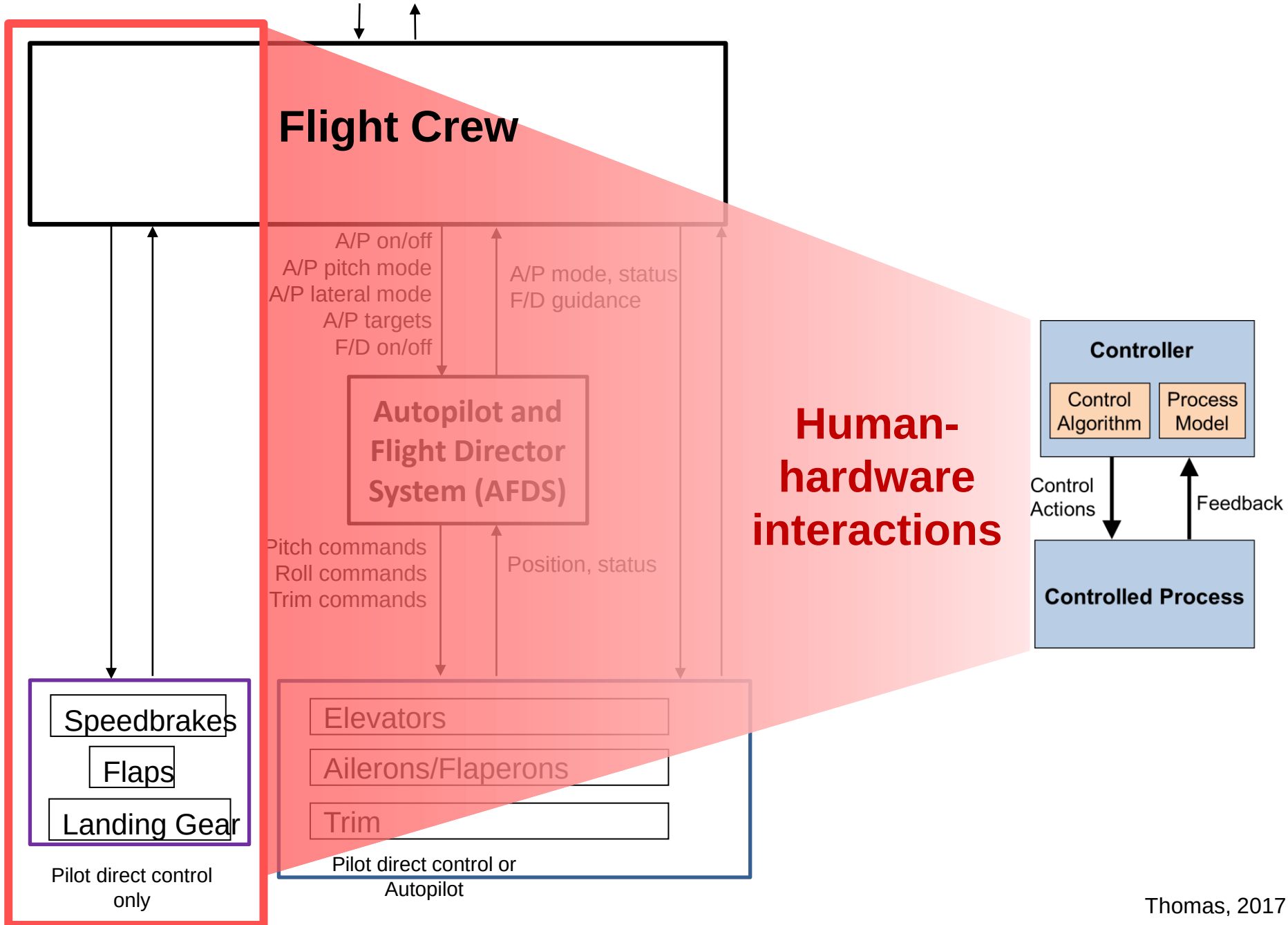



Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Speedbrakes
Flaps
Landing Gear
Elevators
Ailerons/Flaperons
Trim
Pilot direct control only
Pilot direct control or Autopilot
Software-hardware interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process

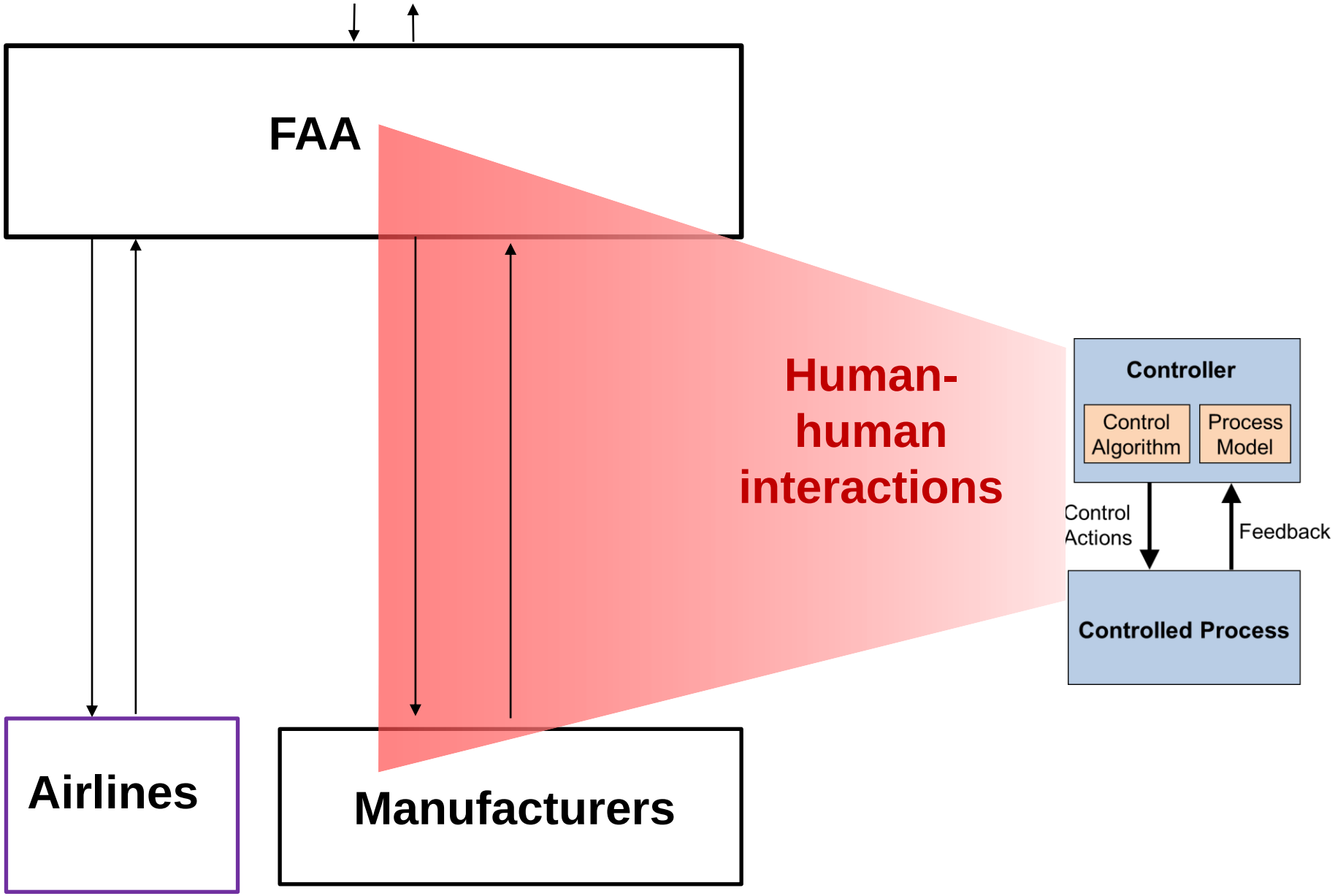
Thomas, 2017


Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Human-automation interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process
Speedbrakes
Flaps
Landing Gear
Elevators
Ailerons/Flaperons
Trim
Pilot direct control only
Pilot direct control or Autopilot

Thomas, 2017


Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Speedbrakes
Flaps
Landing Gear
Pilot direct control only
Elevators
Ailerons/Flaperons
Trim
Pilot direct control or Autopilot
Human-hardware interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process

Thomas, 2017


FAA
Human-human interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process
Airlines
Manufacturers

Thomas, 2017

Integrated Approach to Safety and Security (Col. Bill Young)

Current	STPA
Information security	Mission Assurance
Keep threats out	Ensure critical functions maintained if they get in

- New paradigm for safety will work for security too
  - May have to add new causes, but rest of process is the same
- A top-down, system engineering approach to designing safety and security into systems
- Integrated analysis:
  - Start with STPA for safety
  - Add extra scenarios for security (intentional actions)

Example: Stuxnet

- Loss: Damage to reactor (in this case centrifuges)
- Hazard/Vulnerability: Centrifuges are damaged by spinning too fast
- Constraint to be Enforced: Centrifuges must never spin above maximum speed
- Hazardous control action: Issuing *increase speed* command when already spinning at maximum speed
- One potential causal scenario:
  - Incorrect process model: thinks spinning at less than maximum speed
    - Could be inadvertent or deliberate
- Potential controls:
  - Mechanical limiters (physical interlock), Analog RPM gauge

Focus on preventing hazardous state
(not keeping intruders out)