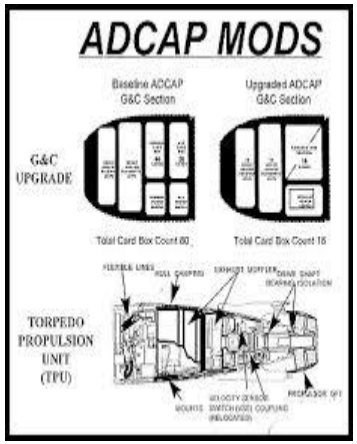
Safety and Cybersecurity in Human-Cyber-Physical Systems

Nancy Leveson
MIT


ADCAP MODS
Baseline ADCAP G&C Section
Upgraded ADCAP G&C Section
G&C UPGRADE
Total Card Box Count 60
Total Card Box Count 16
TORPEDO PROPULSION UNIT (TPU)

Bottom Line Up Front (BLUF)

- Complexity is reaching a new level (tipping point)
  - Old system engineering approaches becoming less effective
  - New causes of mishaps 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

  Change focus

  ~~Increase component reliability (analytic decomposition)~~

  ⇓

  Enforce safe behavior (dynamic control using systems theory)

BLUF (2)

- Allows creation of new analysis and engineering approaches
  - More powerful and inclusive
  - Orders of magnitude less expensive
  - Work on extremely complex systems (top-down system engineering)
  - Design safety and security and other properties in from the beginning
  - Compliant with MIL-STD-882E and other military standards, commercial standards being developed (autos, aircraft, defense)

- 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 create safety, security, and other requirements
  - Theoretically works on any system property

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,
  - mission loss,
  - negative business impact (damage to reputation, etc.), product launch delay, legal entanglements
- Includes inadvertent and intentional
- System goals vs. constraints (limits on how can achieve the goals)
- Hazard/vulnerability: A system state or set of conditions that, together with worst-case environmental conditions, will lead to a loss

Understanding The Problem

“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

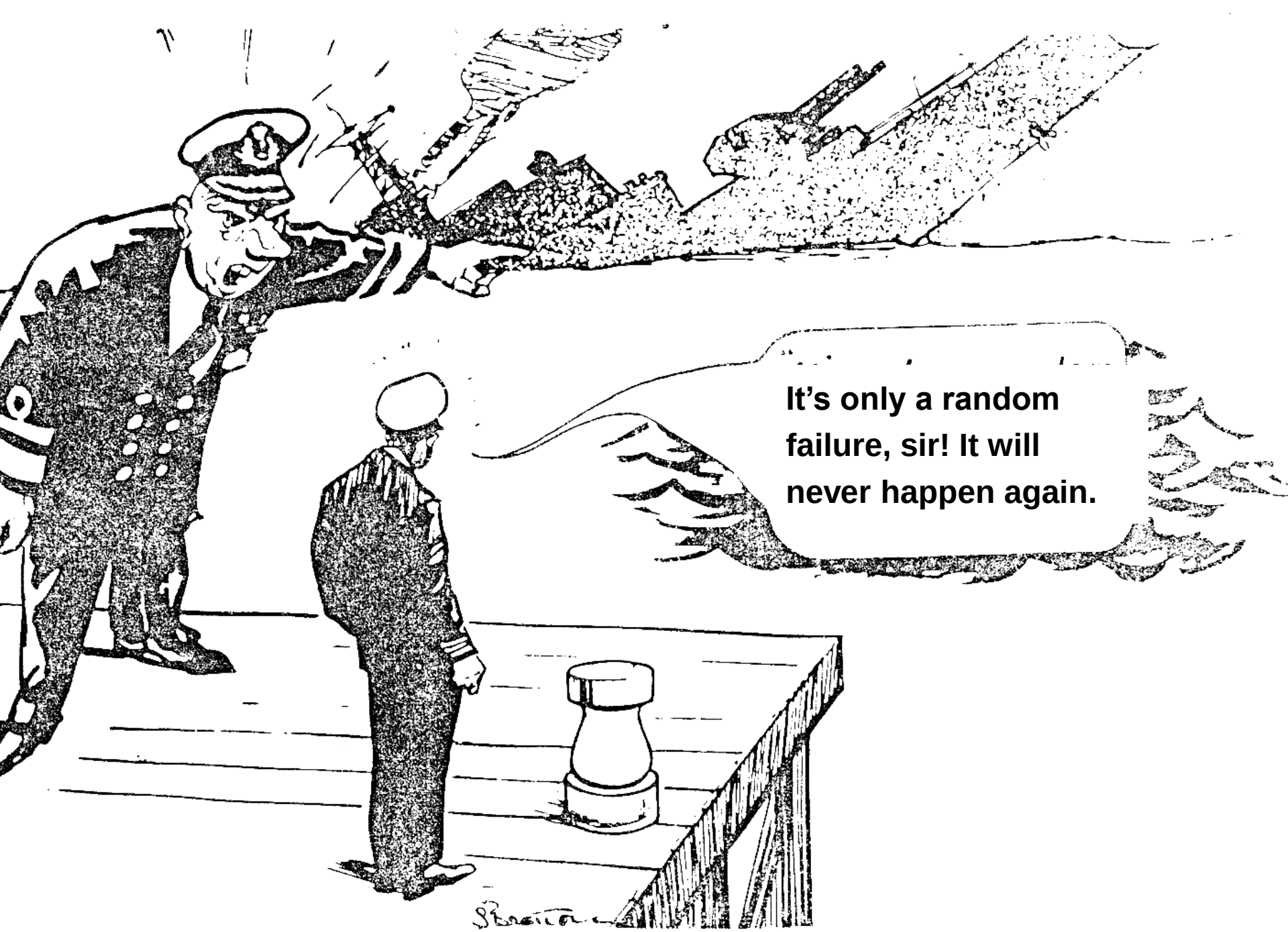
ETA

Bow Tie
(CCA)
FTA + ETA

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

Assumes accidents caused by component failures

© Copyright Nancy Leveson, June 2011


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.

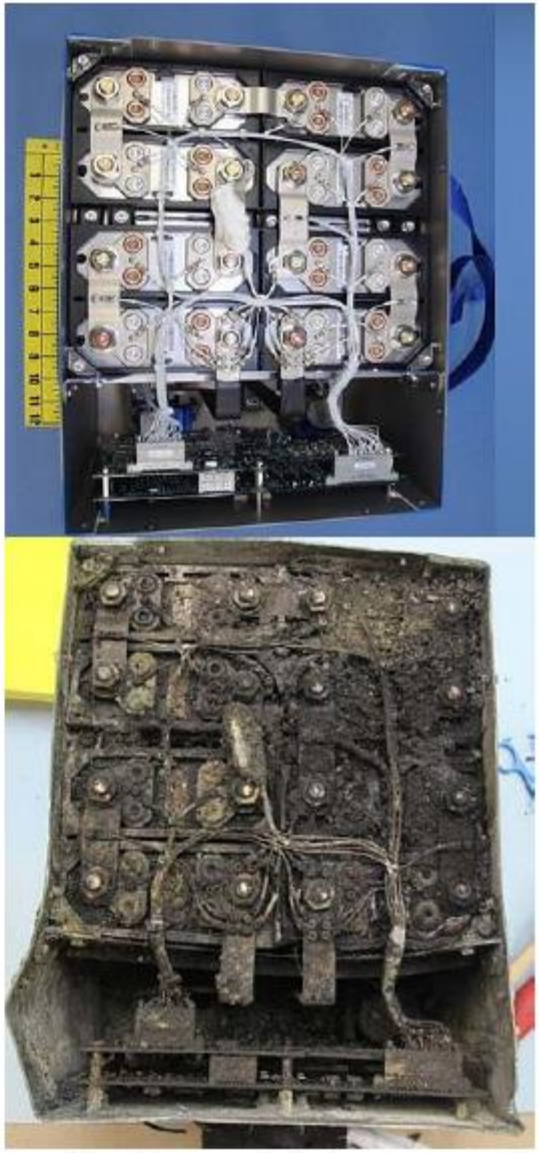
Warsaw A320 Accident

- 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.

- Boeing 787 LiCo Batteries

- Prediction/Certification:
  - No fires within 10^7 flight hours
  - Followed 4761 certification paradigm

- Actual experience:
  - Within 52,000 flight hours – 2 such events
  - 2.6×10^4 flight hours [NTSB 2013]



[http://upload.wikimedia.org/wikipedia/commons/9/95/Boeing_Dreamliner_battery_original_and_damaged.jpg]

Boeing 787 Lithium Battery Fires

- A module monitors for smoke in the battery bay, controls fans and ducts to exhaust smoke overboard.
- Power unit monitors for low battery voltage, shut down various electronics, including ventilation
- Smoke could not be redirected outside cabin

All software requirements were satisfied!
The requirements were unsafe

Washington State Ferry Problem

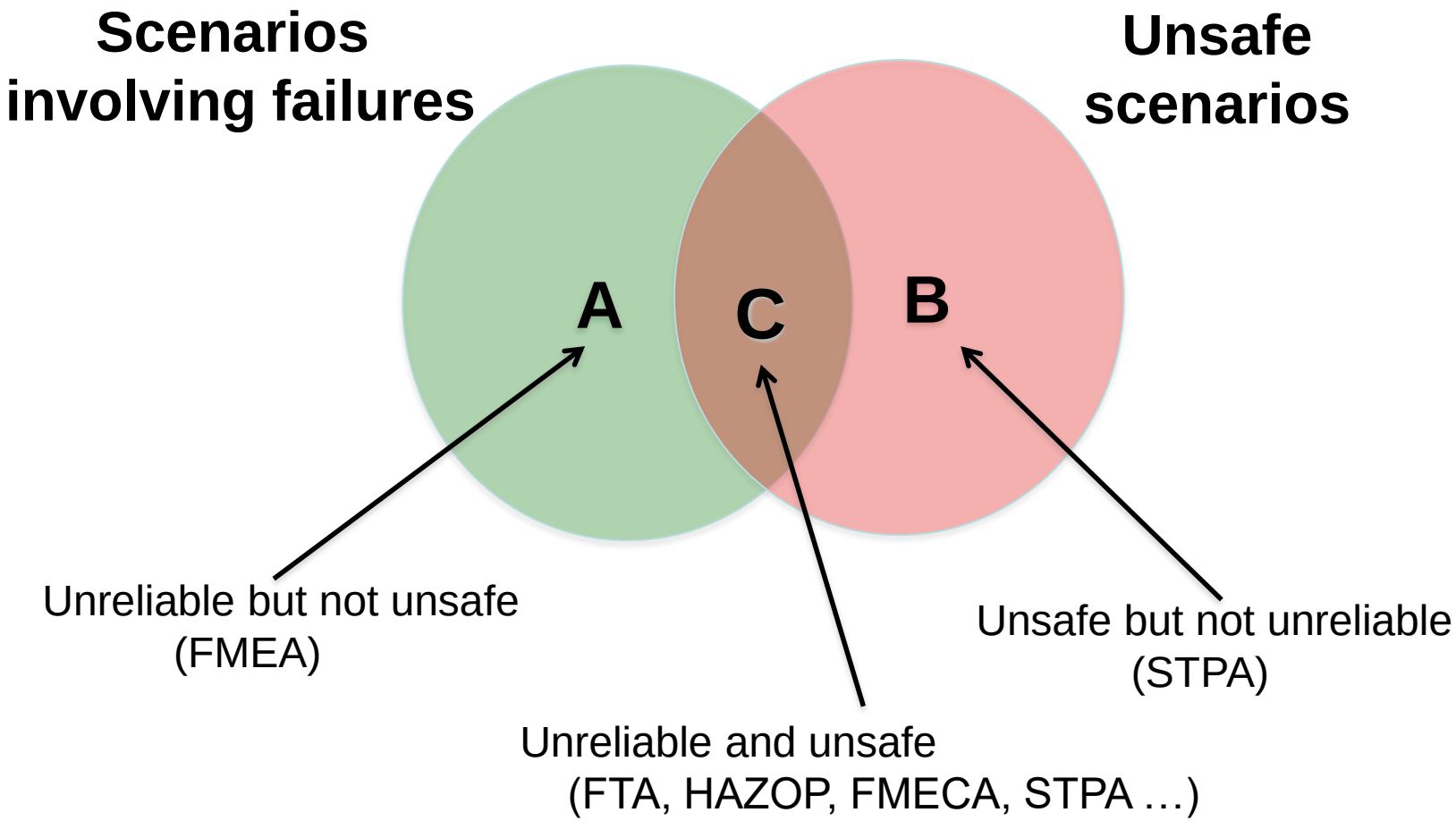
- Local rental car company installed a security device to prevent theft by disabling cars if car moved when engine was stopped
- When ferry moved and cars not running, disabled them.
- Rental cars could not be driven off ferries when got to port

Traditional Approach to Safety

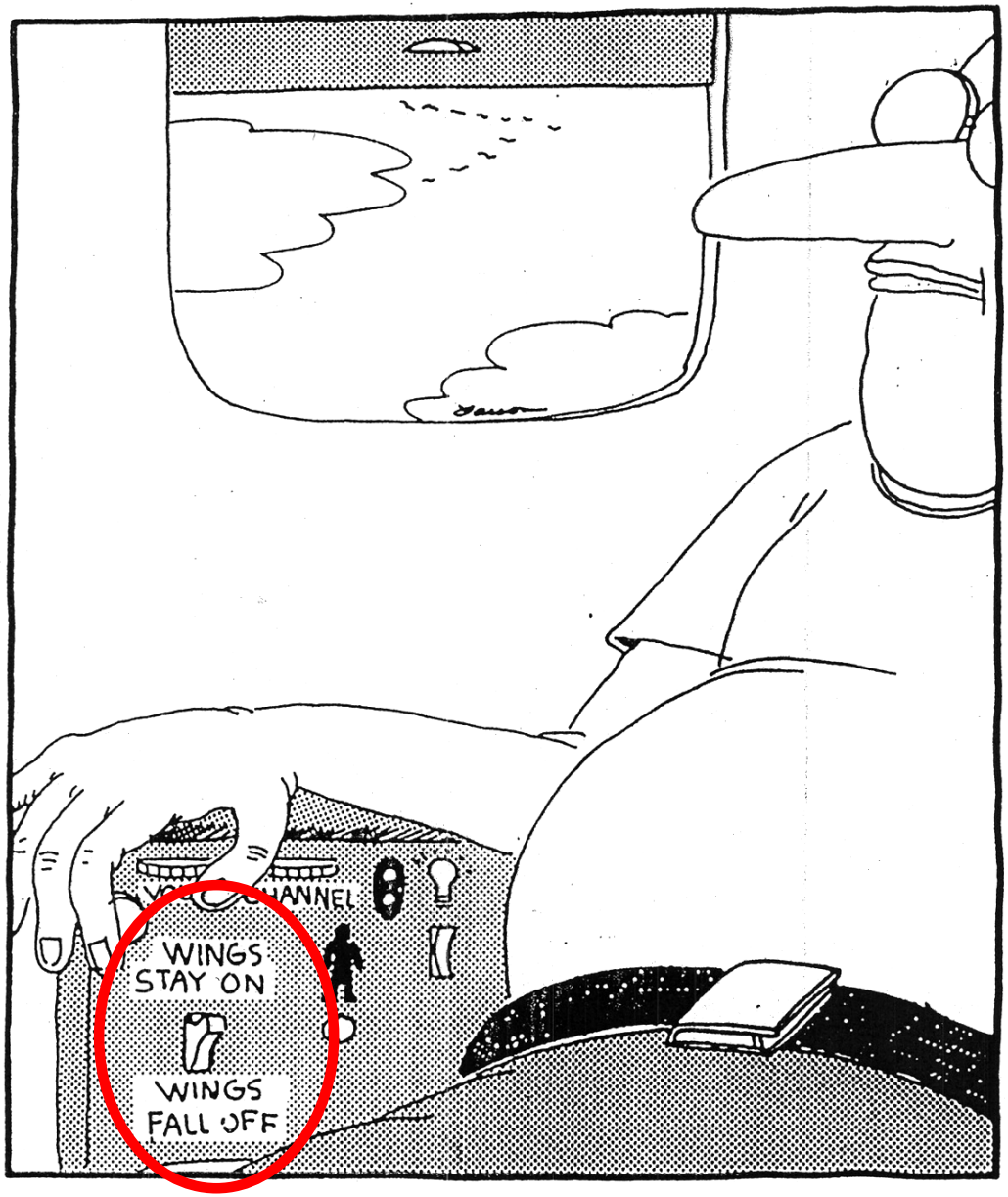
- Traditionally view safety as a failure problem
  - Chain of directly related failure events leads to loss
  - Try to prevent component failures or establish barriers between events
- Limitations
  - Systems are becoming more complex
    - Accidents often result from interactions among components
    - Cannot anticipate all potential interactions
  - 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

Confusing Safety and Reliability



Preventing Component or Functional Failures is Not Enough



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
  - **Human error is a symptom of a system that needs to be redesigned**.
- To eliminate human errors, need to change the system design

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”

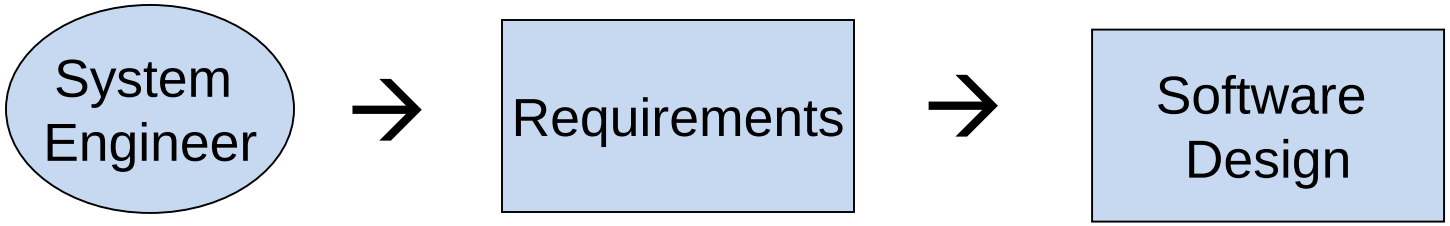
Hardware/software engineering concentrates on the “screen in”

Not enough attention on integrated system as a whole

e.g, mode confusion, situation awareness errors, inconsistent behavior, etc.

Abstraction from Physical Design

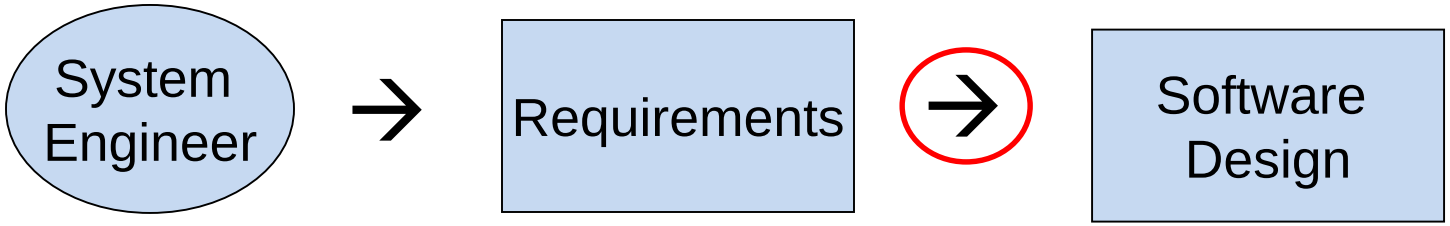
- Software Engineering starts from this perspective



- Most operational software errors related to requirements (particularly incompleteness)
- In losses related to software, it is usually doing exactly what the software engineers intended it to do

Role of Software in Losses

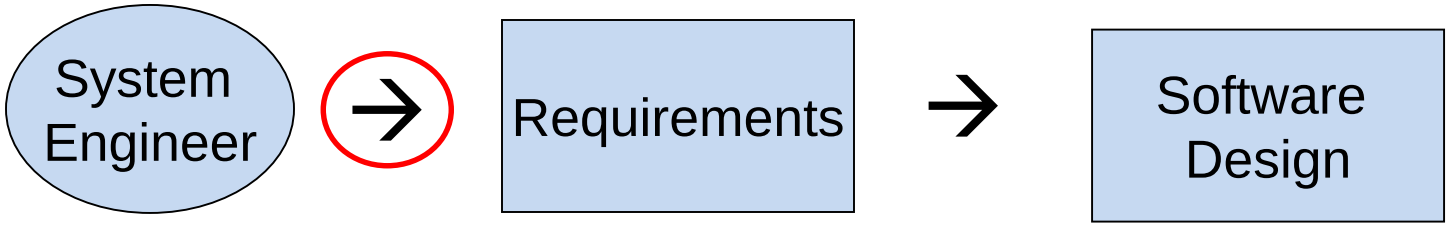
- Software Engineering starts from this perspective



- Most operational software errors related to requirements (particularly incompleteness)
- In losses related to software, it is usually doing exactly what the software engineering intended it to do

Abstraction from Physical Design

- Software Engineering starts from this perspective



- Most operational software errors related to requirements (particularly incompleteness)
- In losses related to software, it is usually doing exactly what the software engineering intended it to do

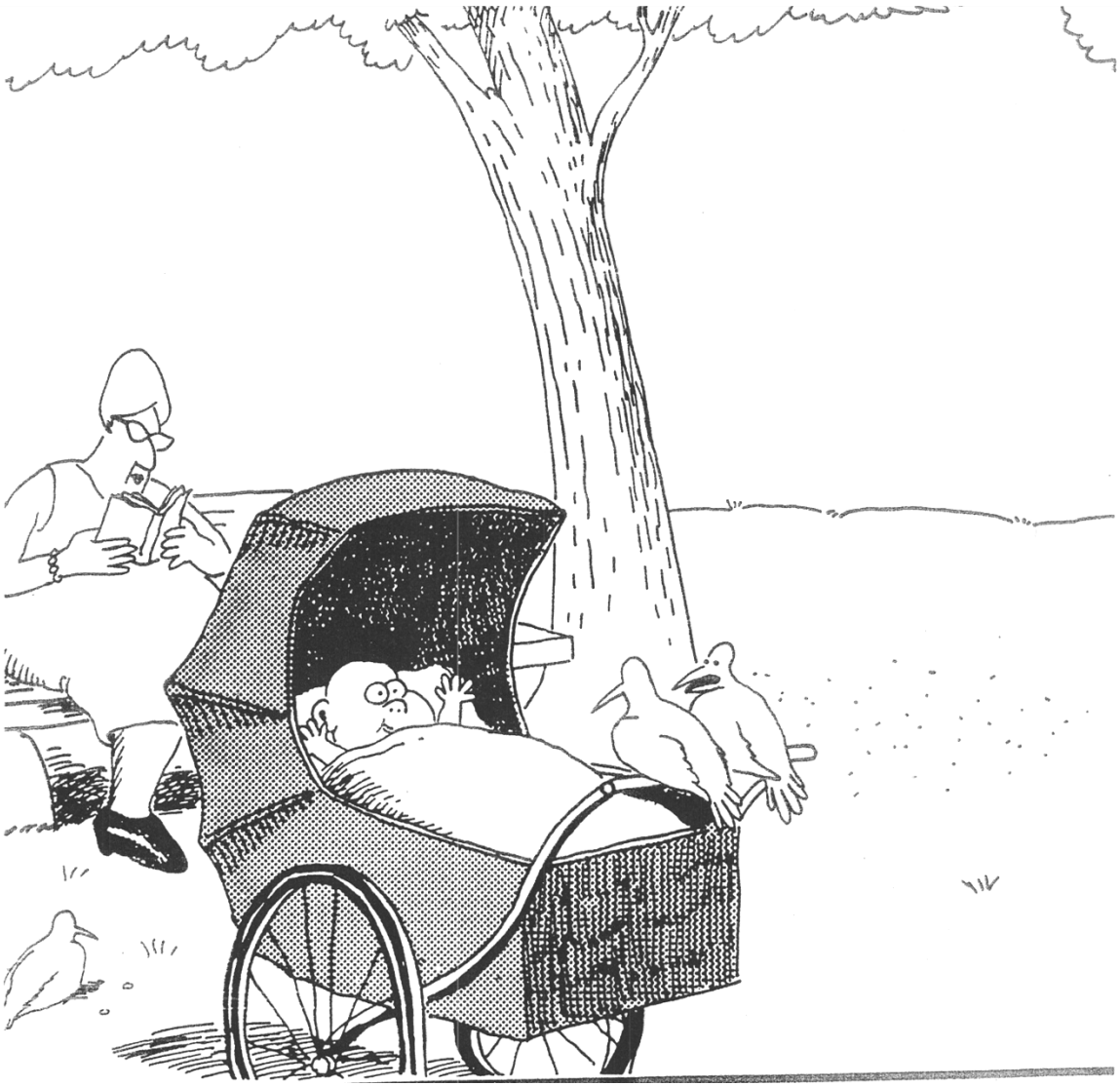
Bottom Line: We Need Something New

- Two approaches being taken now:

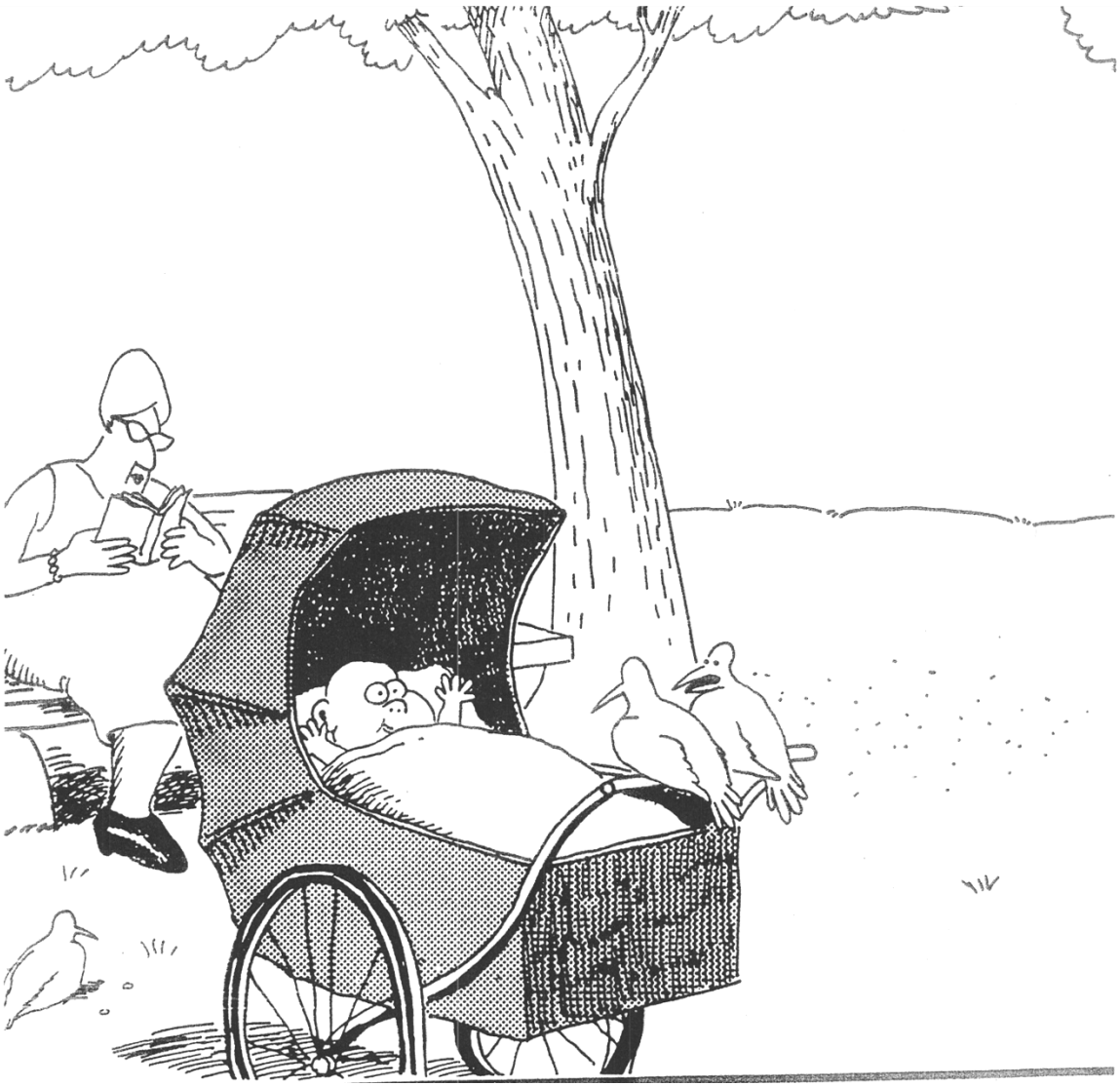
Pretend there is no problem

Shoehorn new technology and new levels of complexity into old methods

New levels of complexity are creating new problems that cannot be solved using traditional techniques.



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.

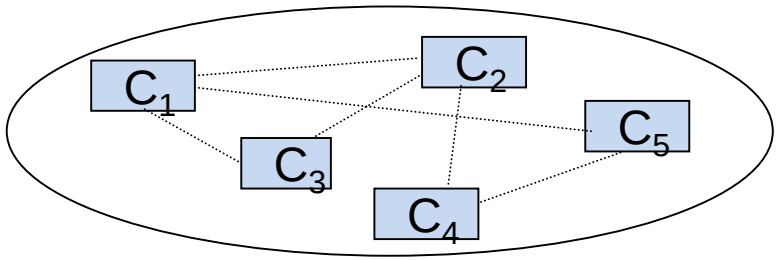
We Need New Tools for the New Problems

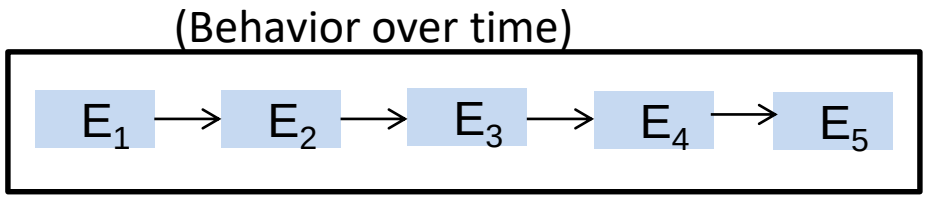
Traditional Approaches to Coping with Complexity

- Analytic Decomposition
- Statistics

Analytic Decomposition ("Divide and Conquer")

1. Divide system into separate parts





2. Analyze/examine pieces separately and combine results

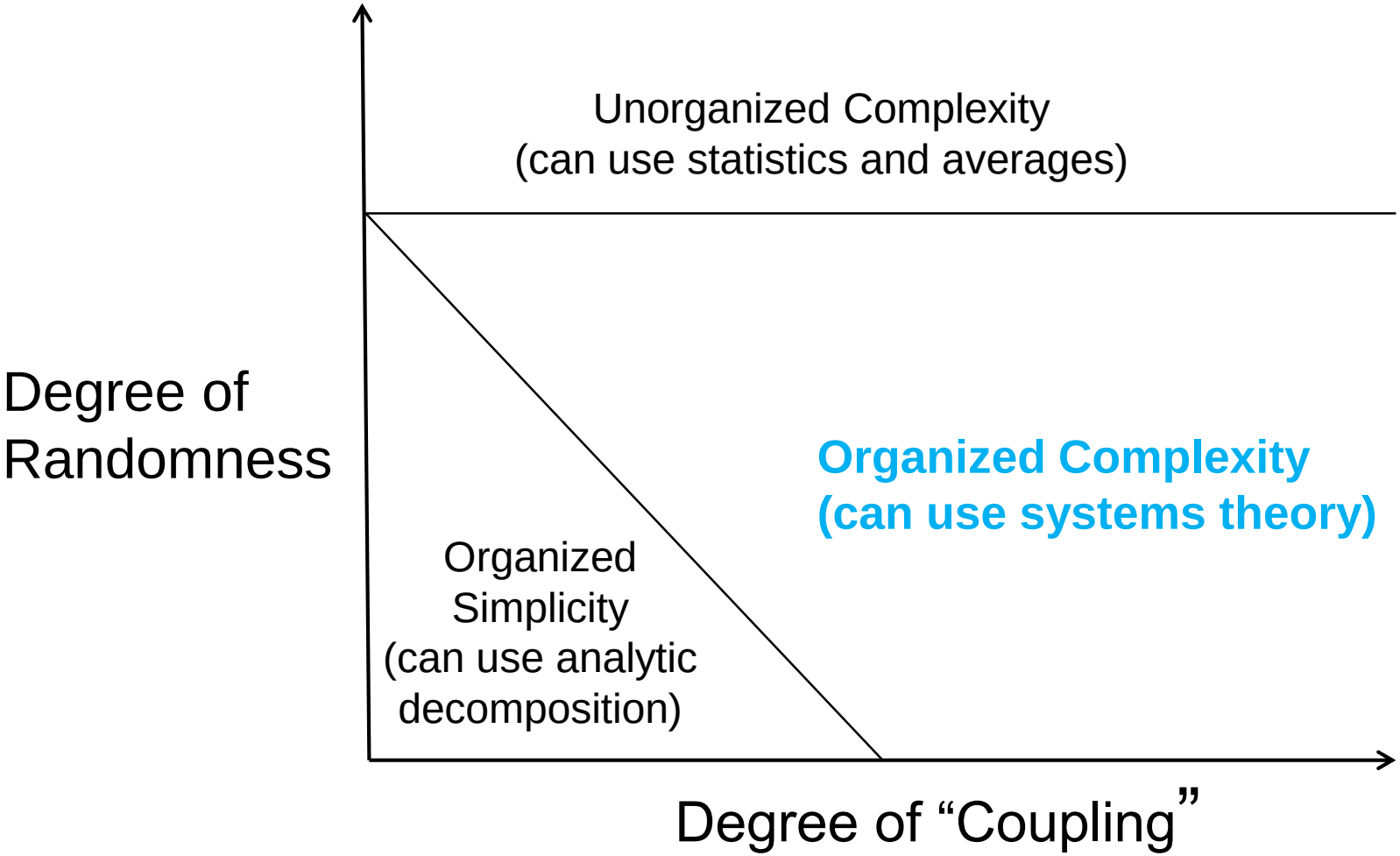
- Assumes such separation does not distort phenomenon
  - ✓ Each component or subsystem operates independently
  - ✓ Components act the same when examined singly as when playing their part in the whole
  - ✓ Components/events not subject to feedback loops and non-linear interactions
  - ✓ Interactions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

Bottom Line

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

  engineered systems today

- Need a new theoretical basis for discovering new approaches and tools
  - *System theory* can provide it



(Gerald Weinberg, An Introduction to General Systems Theory, 1975)

A Potential Way Forward

A Systems Theoretic View of Safety and Security

Systems Theory

- Developed for systems that are
  - Too complex for complete analysis
    - Separation into (interacting) subsystems distorts the results
    - The most important properties are emergent
  - Too organized for statistics
    - Too much underlying structure that distorts the statistics
    - New technology and designs have no historical information
- First used on ICBM systems of 1950s/1960s

System Theory was created to provide a more powerful way to deal with complexity

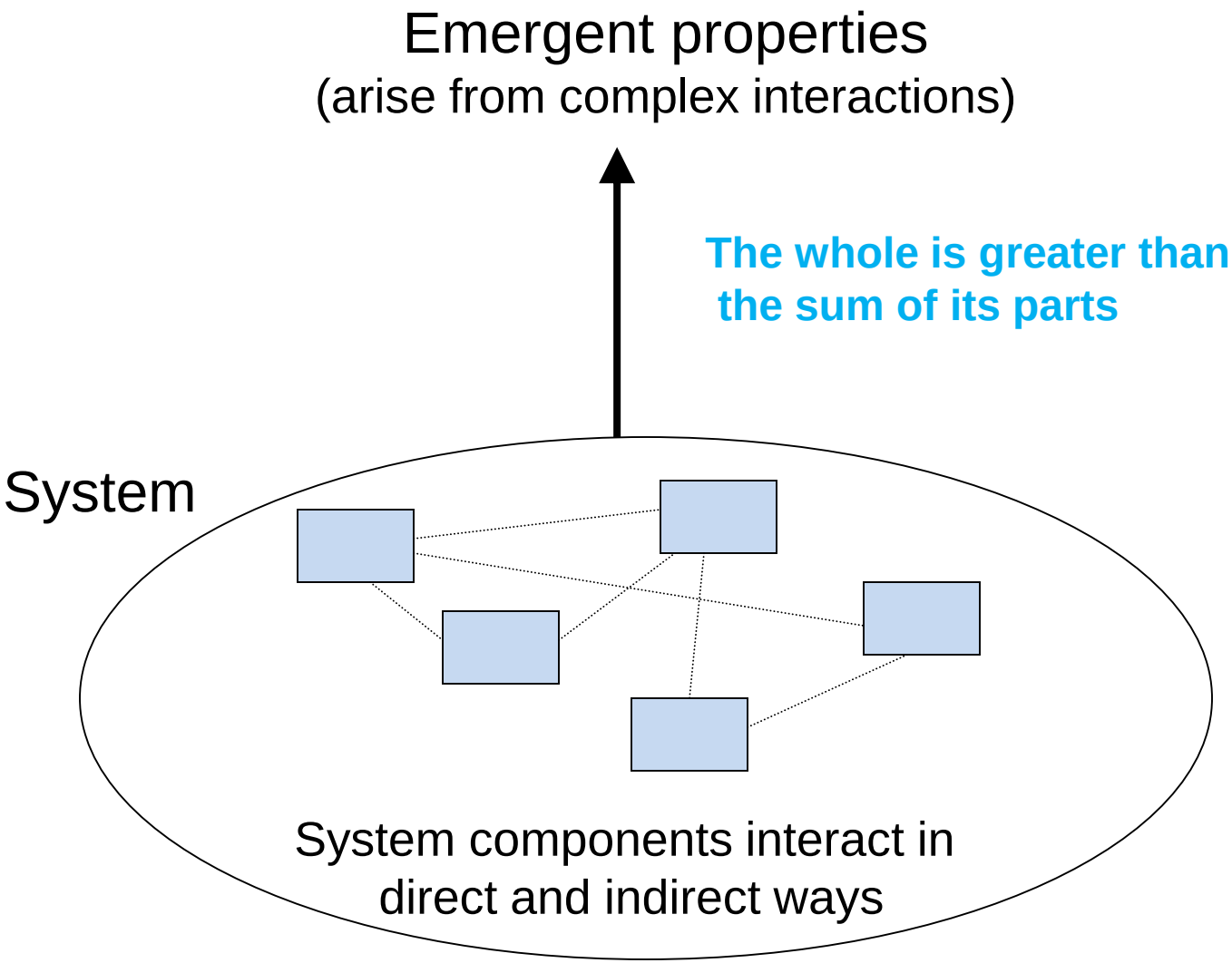
Systems Theory (2)

- Focuses on systems taken as a whole, not on parts taken separately
- Emergent properties
  - Some properties can only be treated adequately in their entirety, taking into account all social and technical aspects

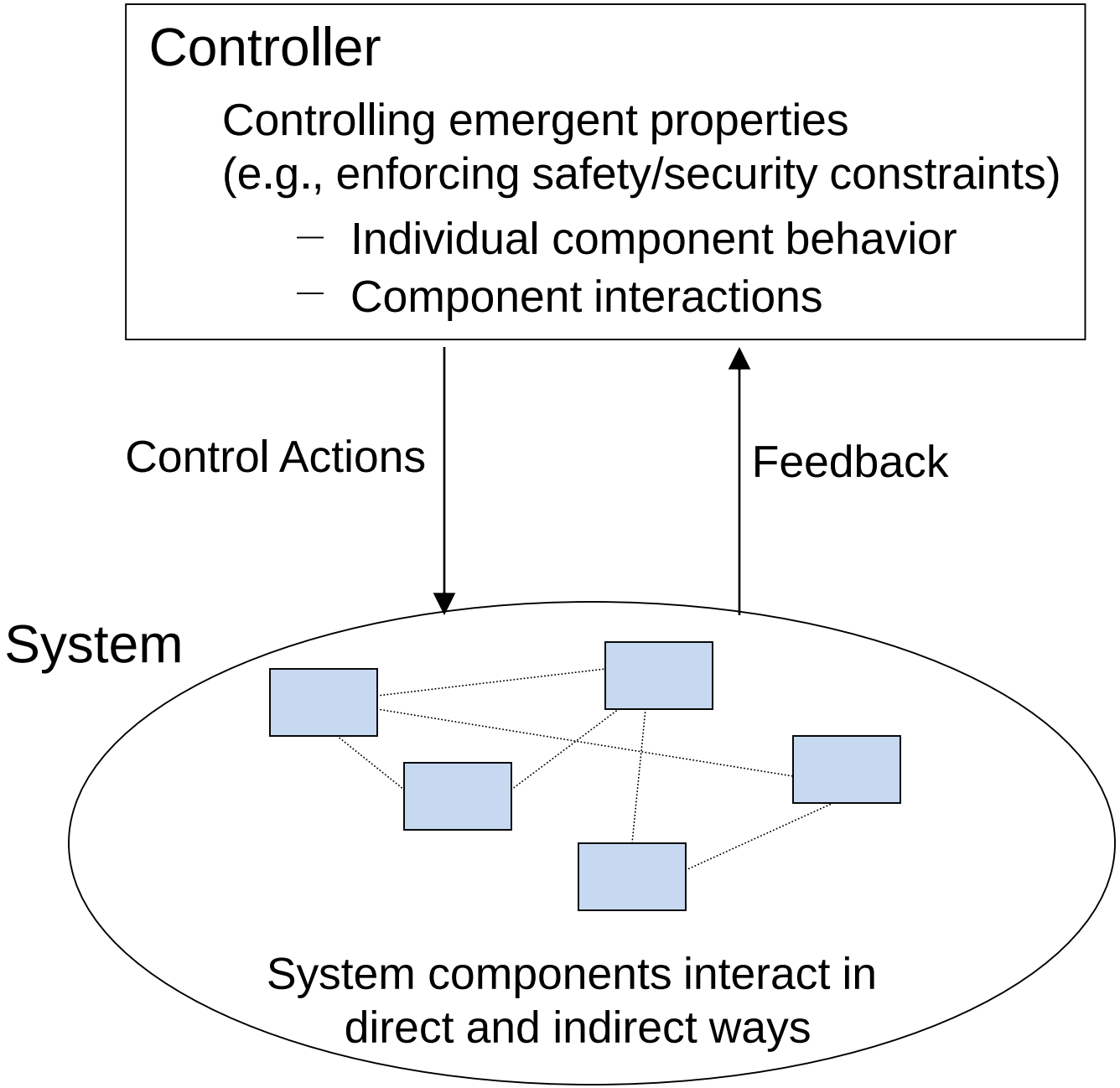
    "The whole is greater than the sum of the parts"

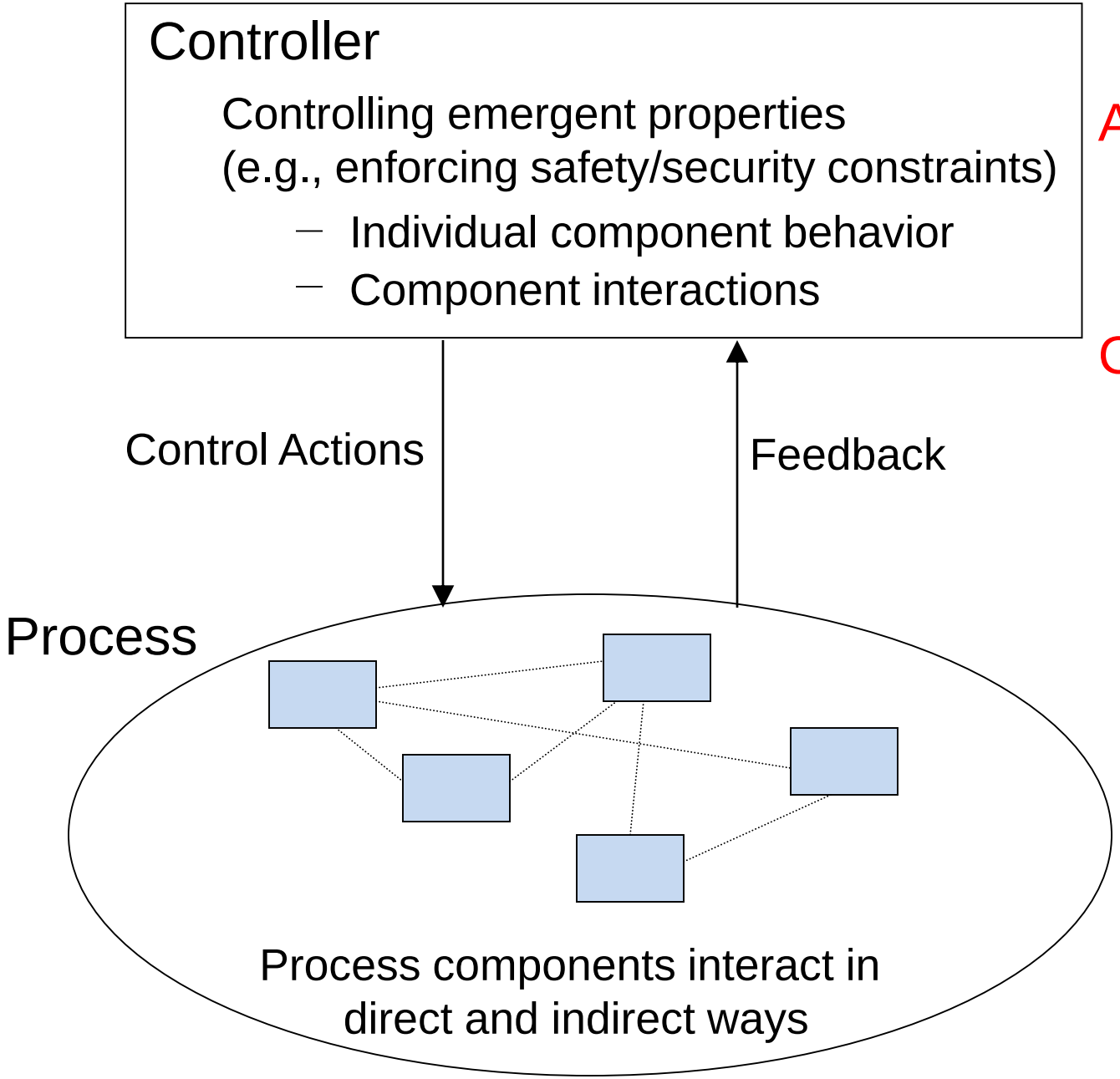
  - These properties arise from relationships among the parts of the system

    How they interact and fit together



Safety and security are emergent properties


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



Air Traffic Control:
Safety
Throughput

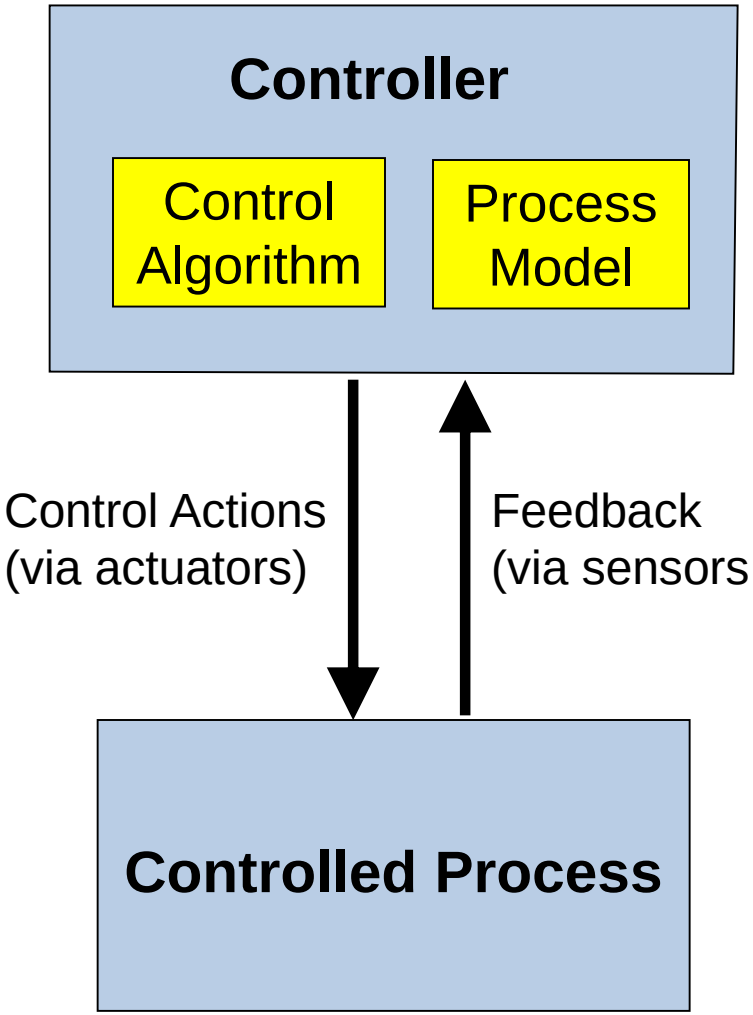
Controlling flow
over internet

Controls/Controllers Enforce Safety/Security Constraints

- Two aircraft/automobiles must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Weapons must not target friendly forces
- Toxic chemicals/radiation must not be released from the plant
- Nuclear materials must never get into the wrong hands
- Weapons must never be detonated inadvertently

Note: Functional/Mission Safety and Security (Mission Assurance)

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 …

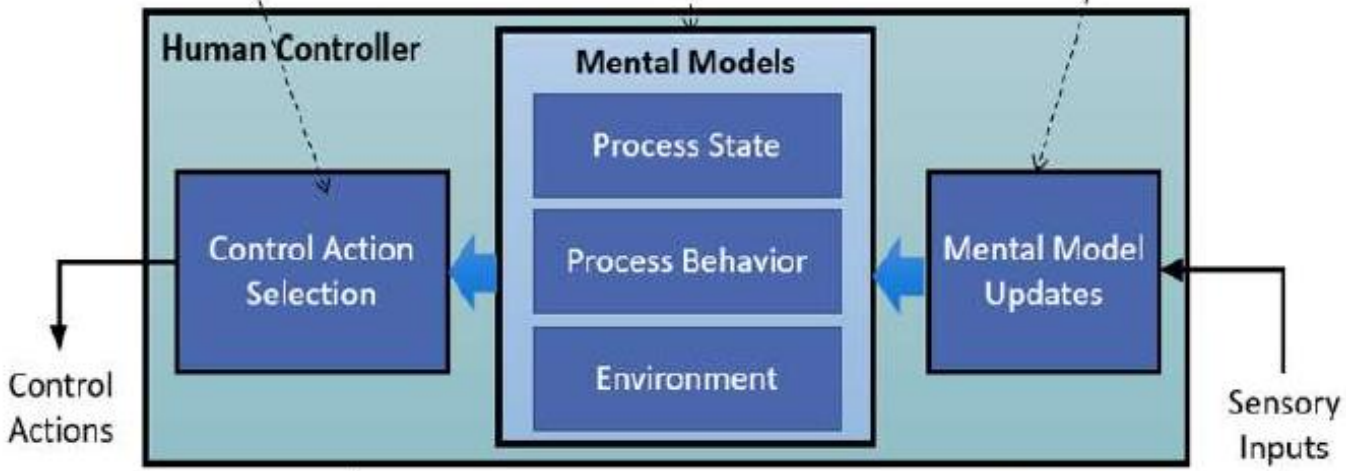
Treat safety as a control problem, not a failure problem

A NEW MODEL FOR HUMAN CONTROLLERS

Captures the controller's goals and how decisions are made based on the mental models

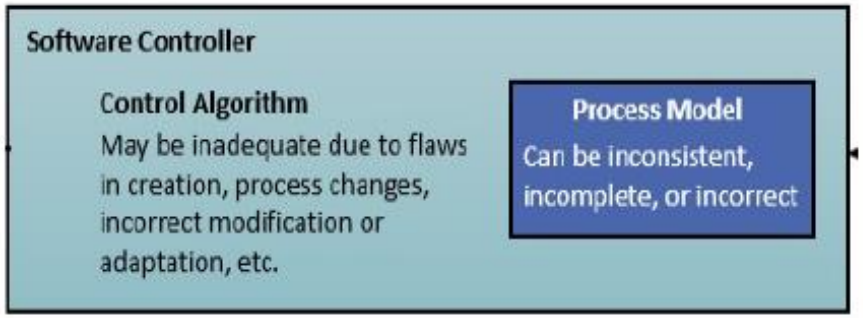
Captures specific types of flaws in the way the human controller conceptualizes the system and environment

Captures the influence of human experiences, and expectations on the processing of sensory input

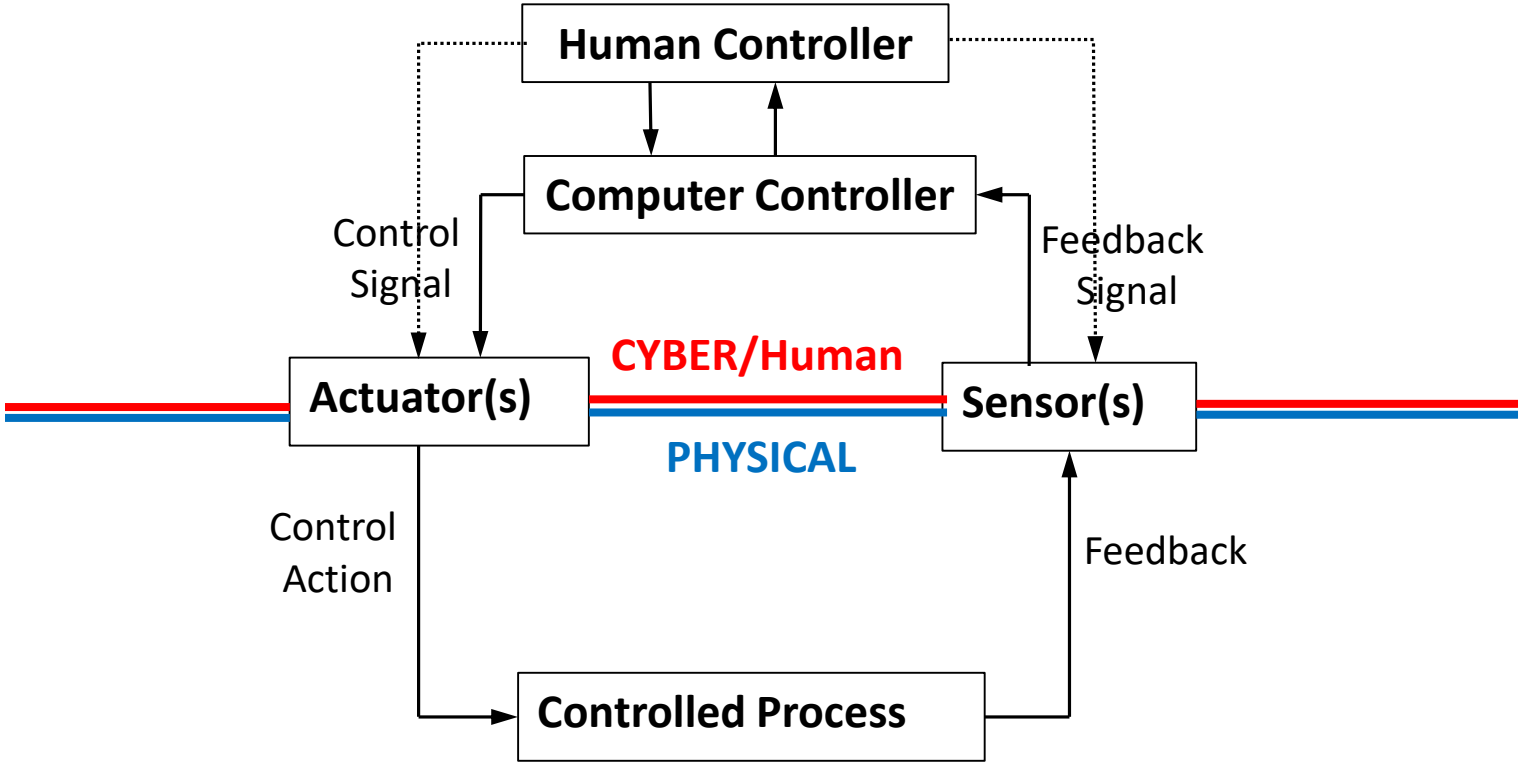


(Thomas & France, 2016)

Provides an alternative to the existing controller model which is better suited for software controllers

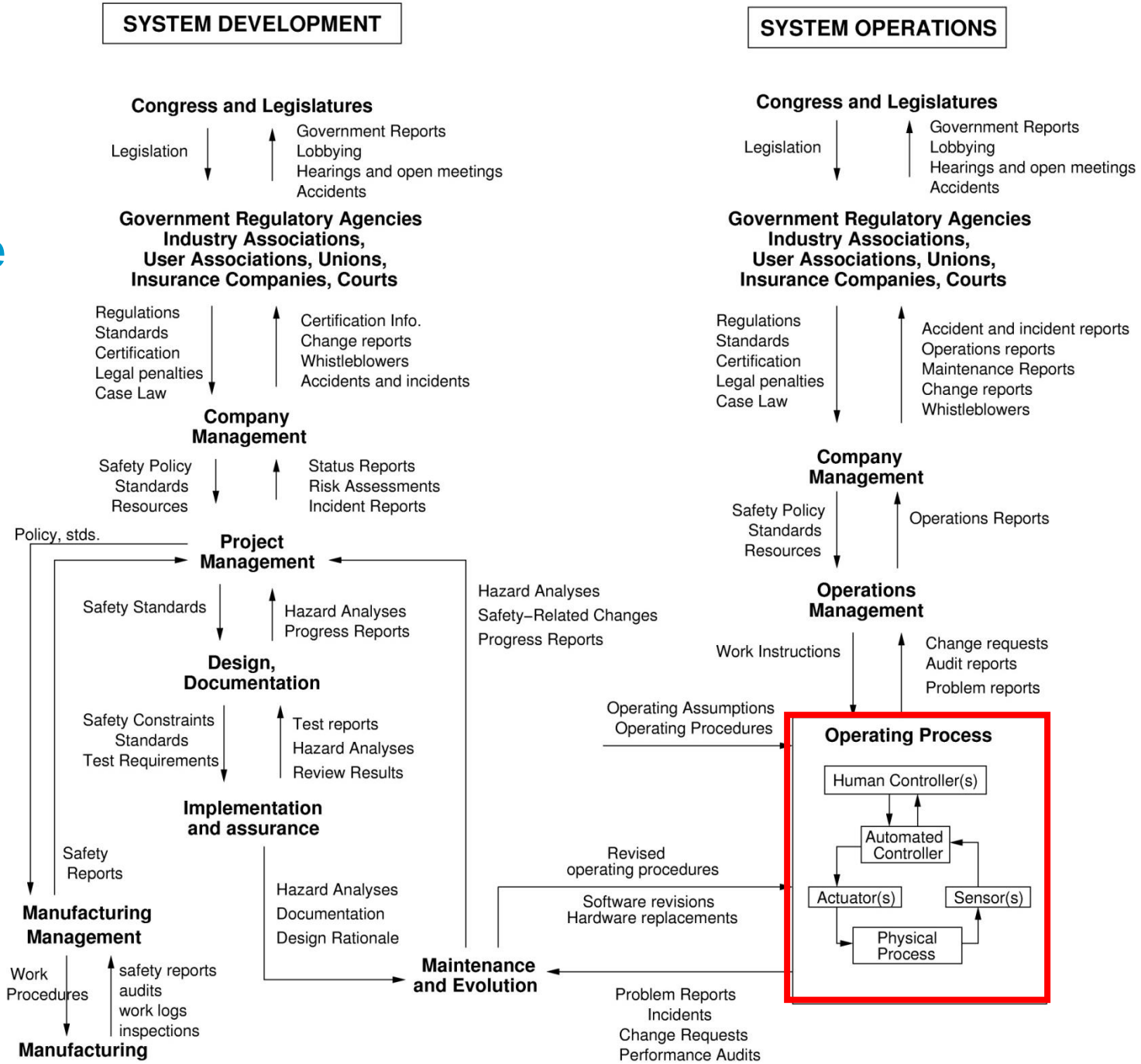


A General Model of Control



- Software is not unsafe; the control signals it generates can be
- Virtually all software-related accidents have resulted from unsafe requirements; not software design errors

Example Safety Control Structure (SMS)



System Block Diagram

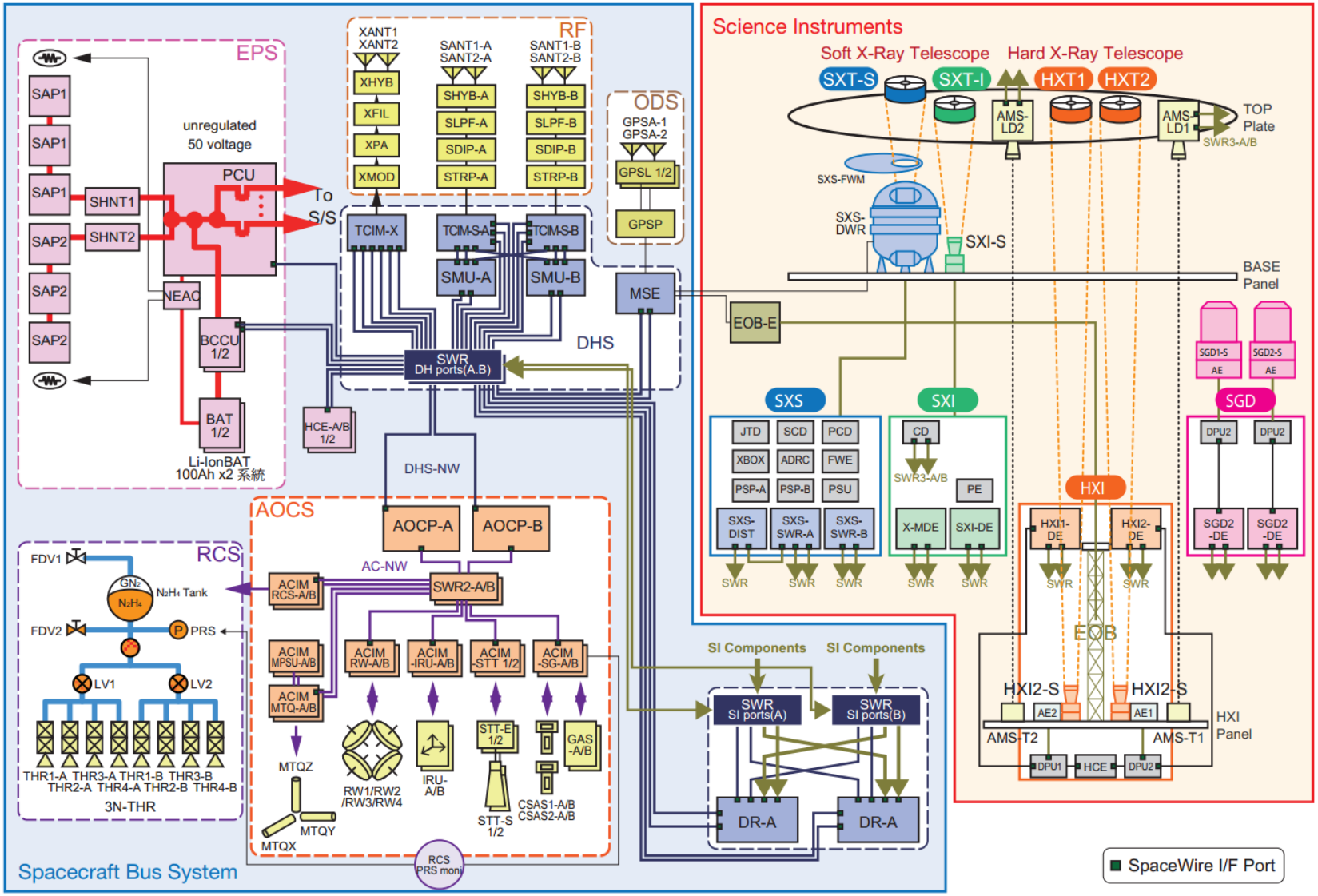
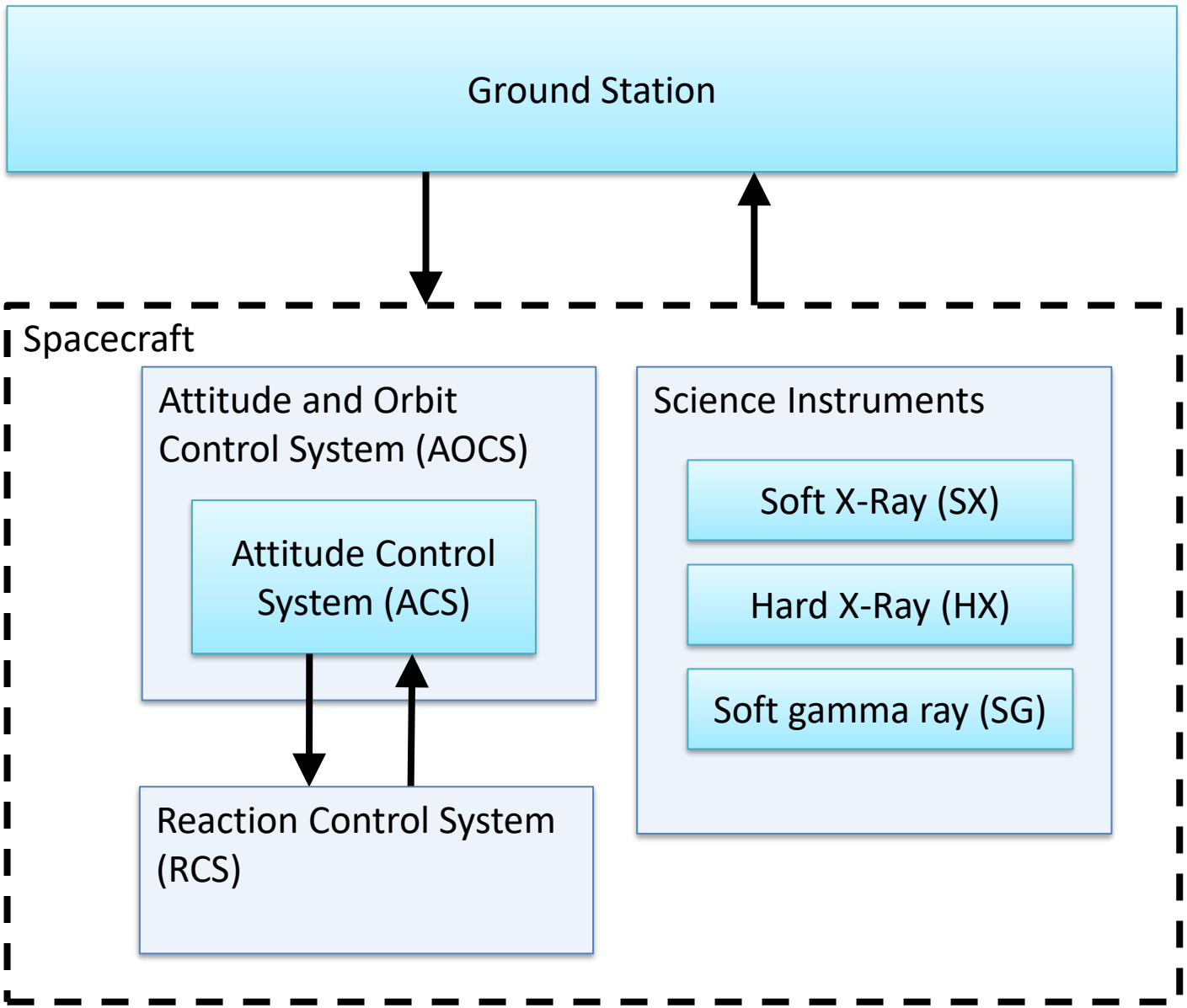
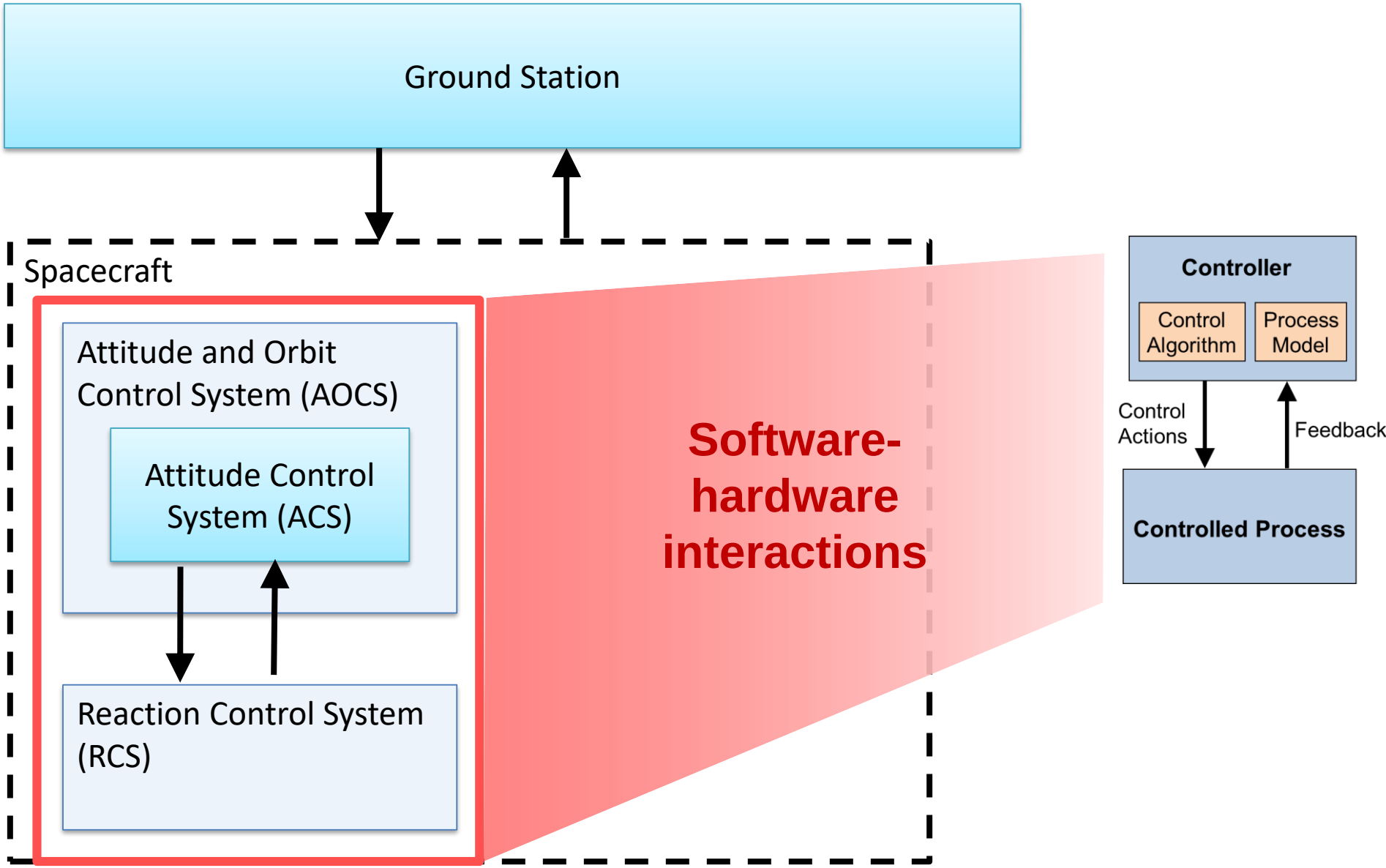


Figure 3.9: System block diagram. A is the primary and B is the redundant system.

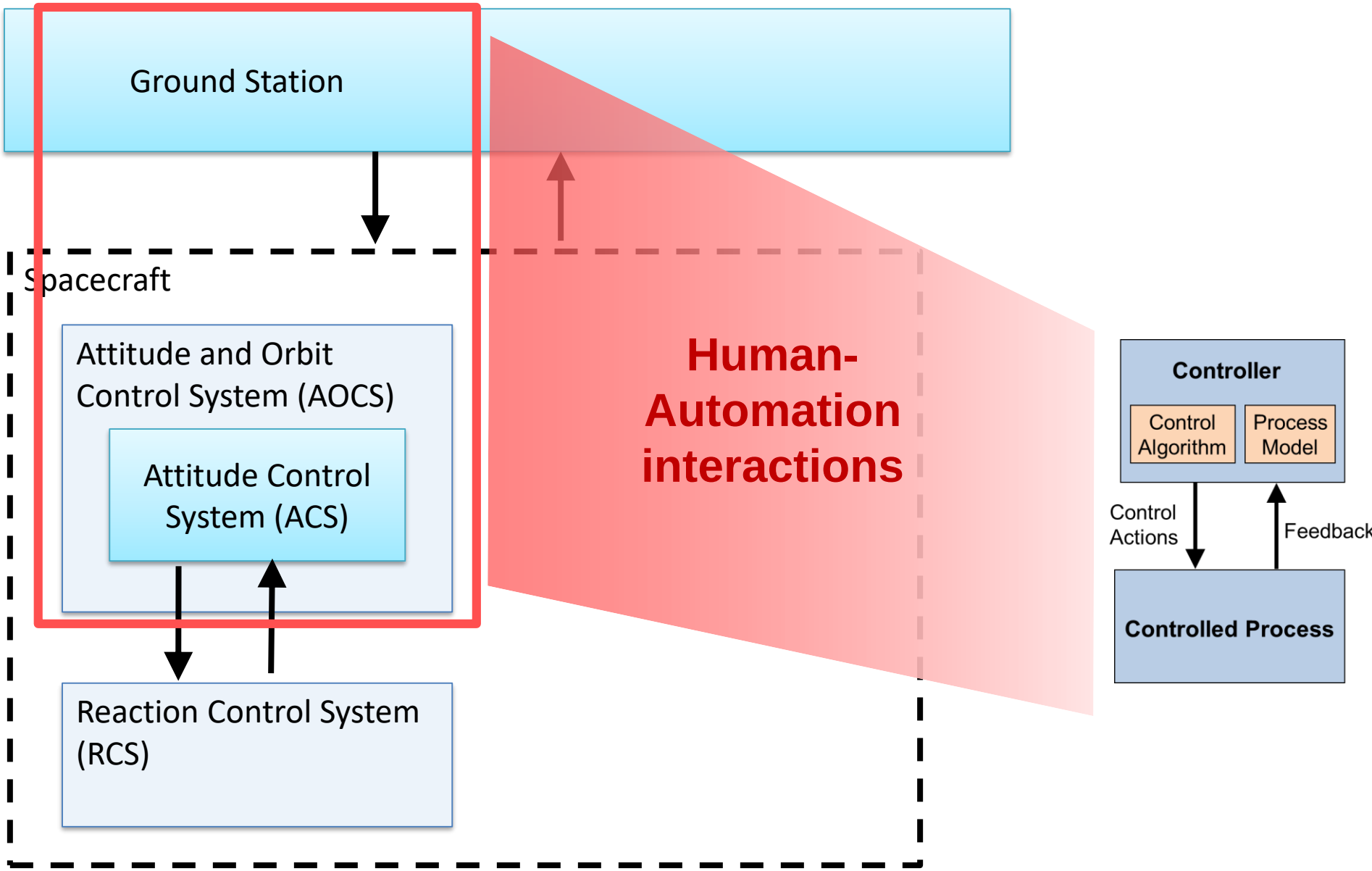
High-level control structure



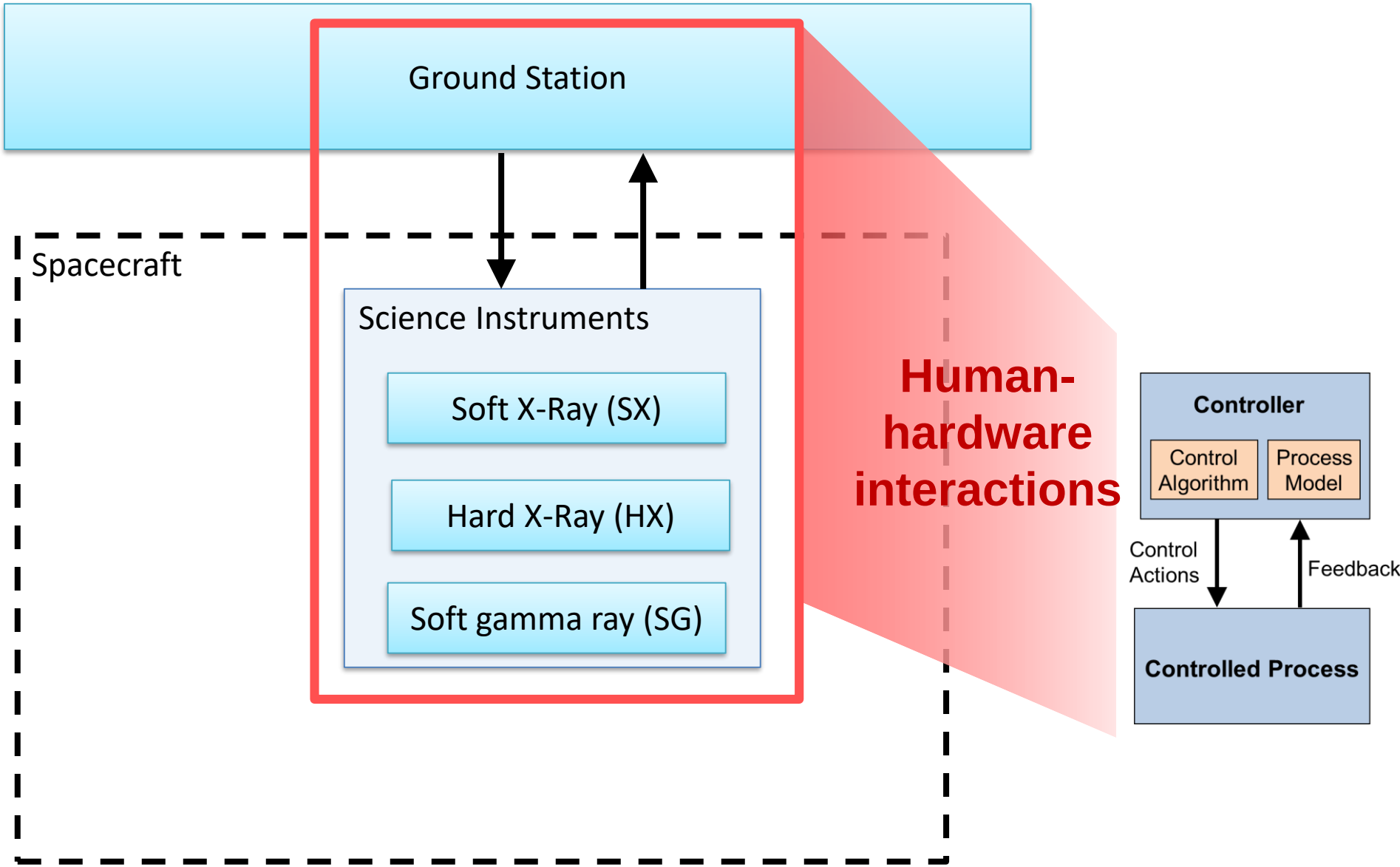
High-level control structure

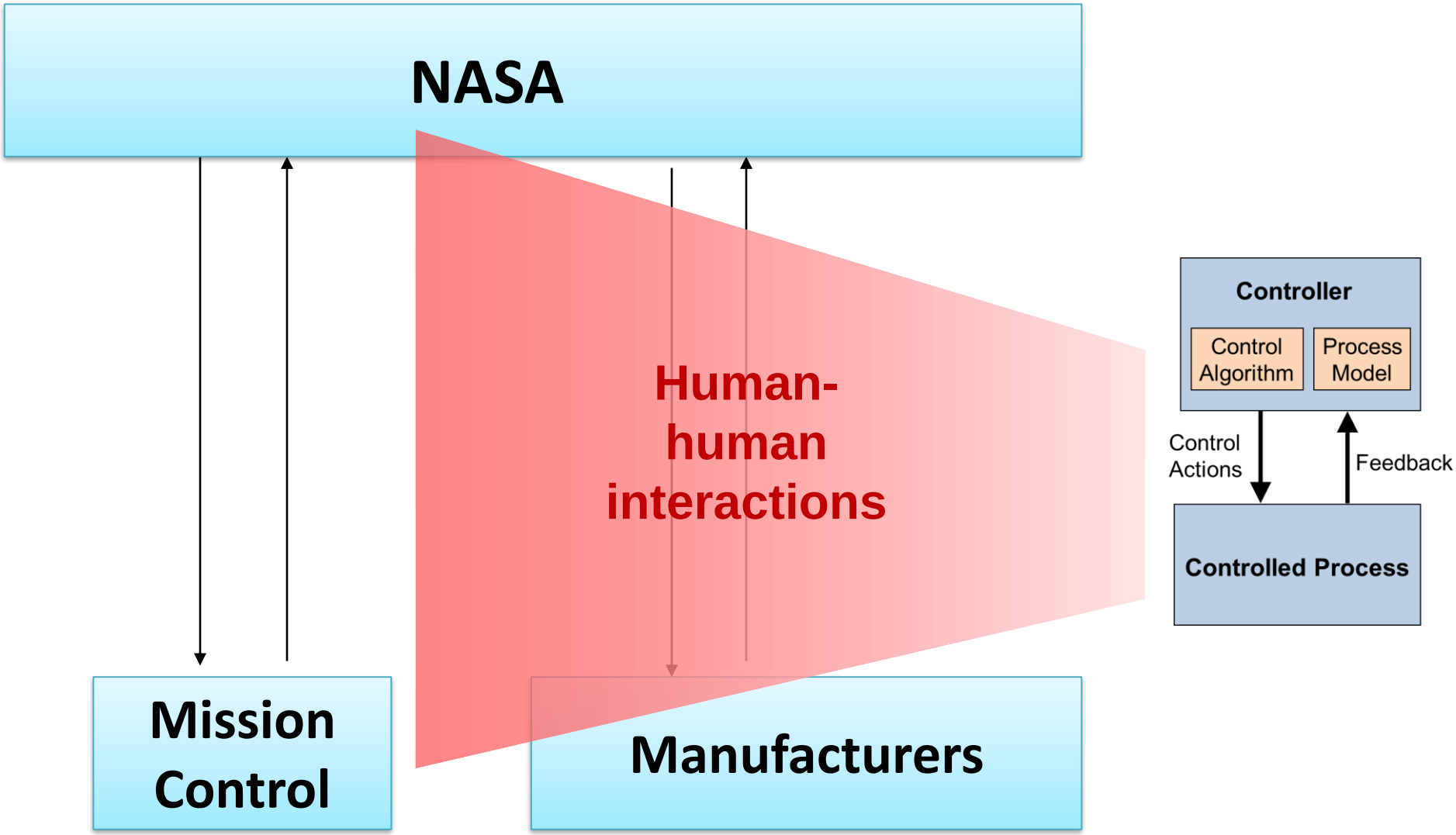


High-level control structure



High-level control structure




NASA
Human-
human
interactions
Mission
Control
Manufacturers
Controller
Control
Algorithm
Process
Model
Control
Actions
Feedback
Controlled Process

Thomas, 2017

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)
- 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

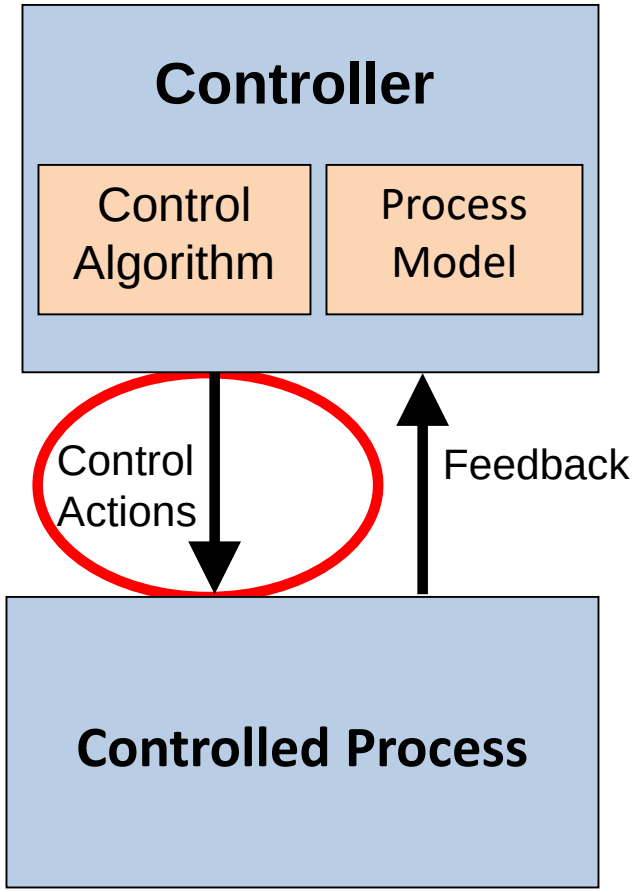
Security Analysis
STPA-Sec

STAMP: Theoretical Causality Model

STPA: System-Theoretic Process Analysis

- 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 <u>safety-driven design</u> process where
  - Analysis influences and shapes early design decisions
  - Analysis iterated and refined as design evolves

Unsafe Control Actions



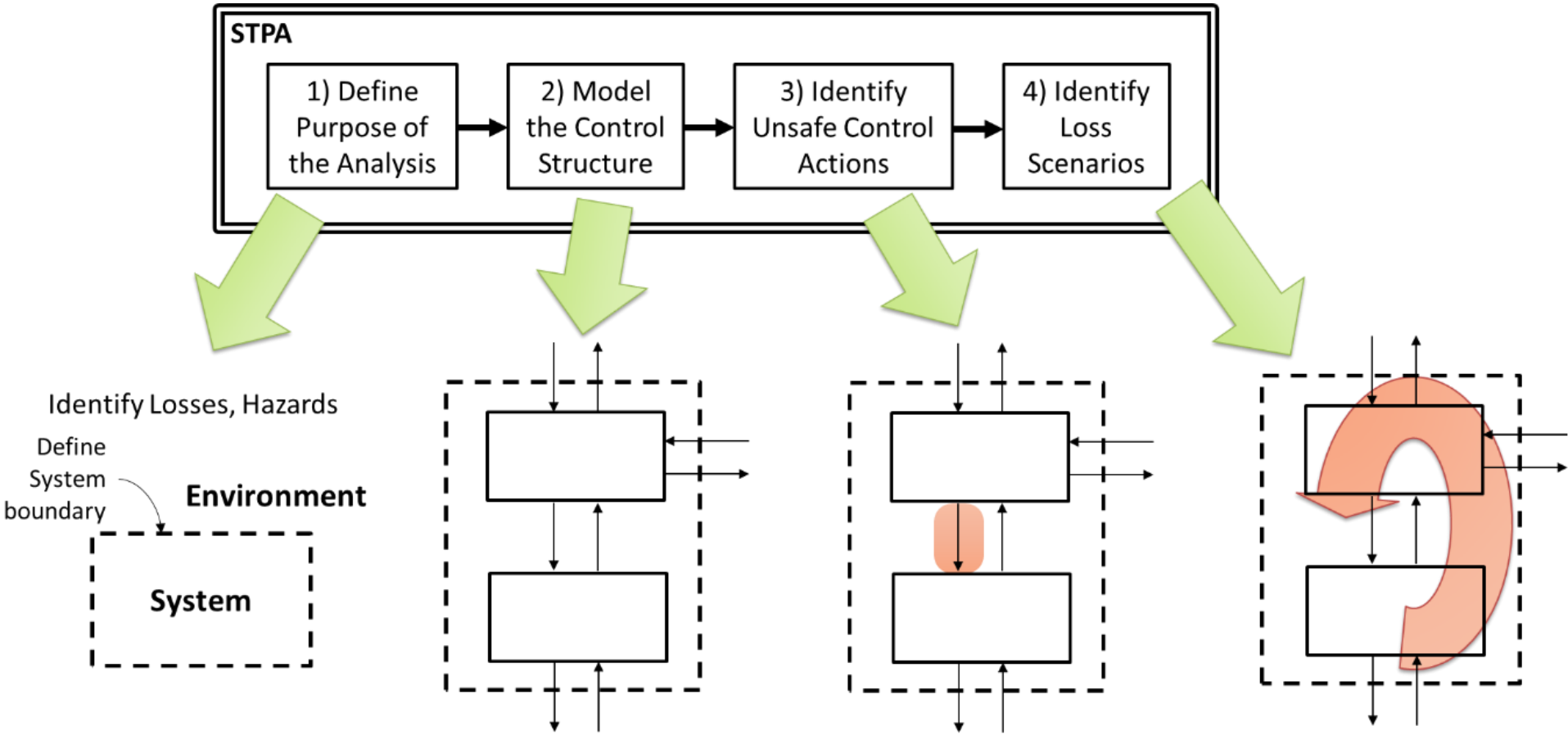
Four types of unsafe control actions

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:

1. Identify potential unsafe control actions
2. Identify why they might be given
3. If safe ones provided, then why not followed?

STPA Process



Embraer: Aircraft Air Management System

Identify Hazards: Reverse flow of air toward engine or APU [leads to power loss]

Create AMS Control Structure [not shown here]

Generate Unsafe Control Actions (UCAs) using STPA on the control structure

UCA1: The AMS does not ensure the correct air flow direction when manifold pressure is higher than engine pressure or APU pressure.

Convert UCAs into Safety Constraints:

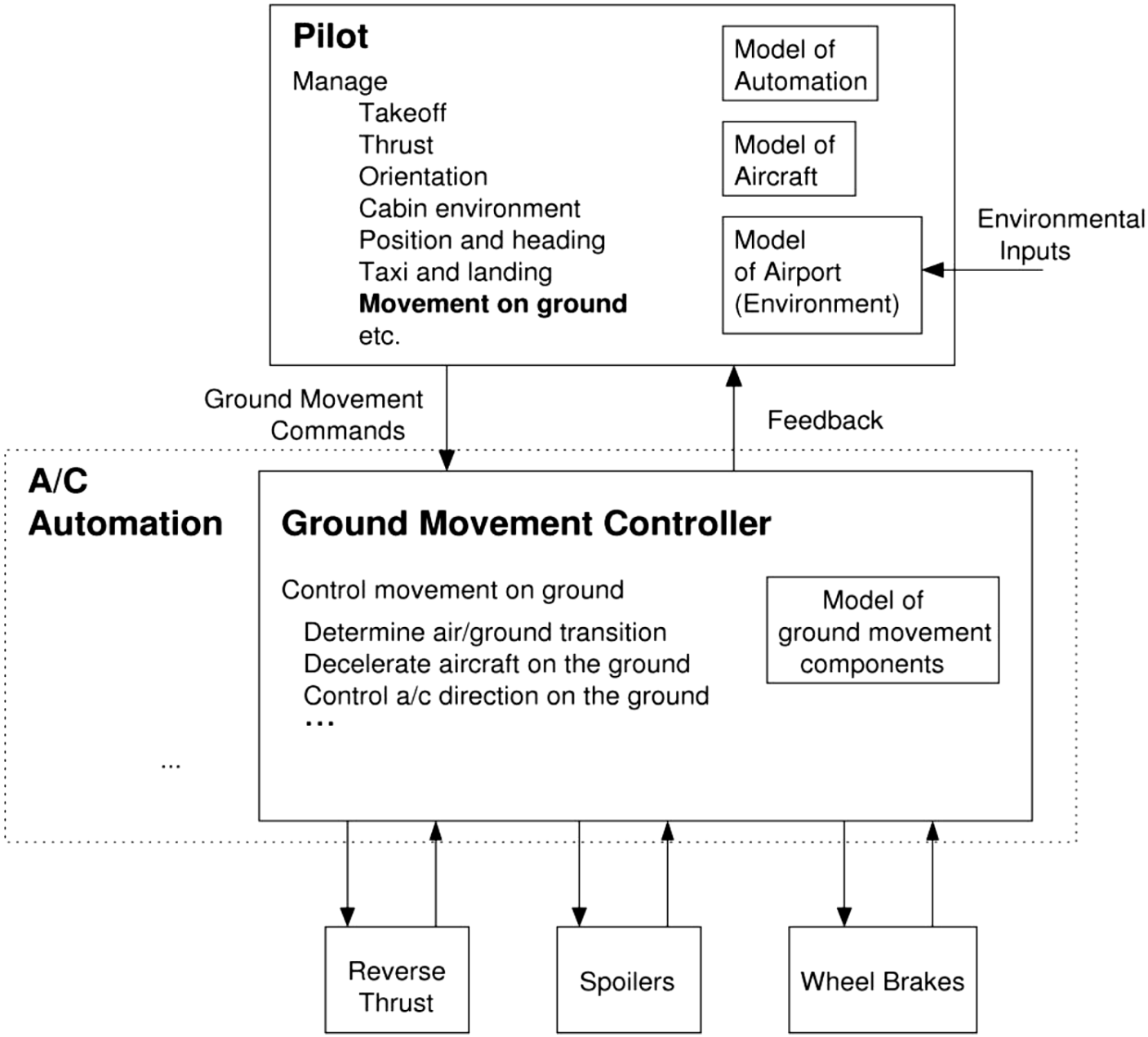
Safety constraint: The AMS must not allow reverse flow when the manifold pressure is higher than the engine pressure or APU pressure

Generate Hazard Scenarios for the UCAs using the control structure model [not shown here]

Air Management System Example (2)

<u>From Causal Scenarios, Generate Detailed Requirements</u>

- The PRSOV valve shall close to 95% of its “fully-closed” position value in 0.2 seconds when the manifold pressure is higher than the engine pressure
- The check valve shall close in 0.1 seconds in case of PRSOV failures in order to prevent reverse flow
- The AMS controller shall command the PRSOV valve to be closed at the same time as the HPRSOV
- The HPRSOV shall close to 95% of its “fully-closed” position value in 0.2 seconds after commanded to close.


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.

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

- Cybersecurity has focused on:
  - Information security vs. **mission assurance**
  - Keeping people out vs. **ensuring critical functions are maintained** if someone gets 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

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 (interlock), Analog RPM gauge

Focus on preventing hazardous state
(not keeping intruders out)

Other Uses

- Leading indicators of increasing risk during operations
- Improved risk assessment
- Generating test data, flight testing
- Production engineering
- “ilities” (other system properties), e.g., quality, servicebility
- Improved Safety Management Systems
- Workplace Safety
- etc.

Is it 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 (including UAVs)
  - Ships/Marine
  - Air Traffic Control
  - Medical Devices and Hospital Safety
  - Chemical plants
  - Oil and Gas
  - Nuclear and Electric Power
  - Robotic Manufacturing / Workplace Safety
  - Finance
- New international standards (autos) created, 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% return on investment when using STPA

Ballistic Missile Defense System (MDA)

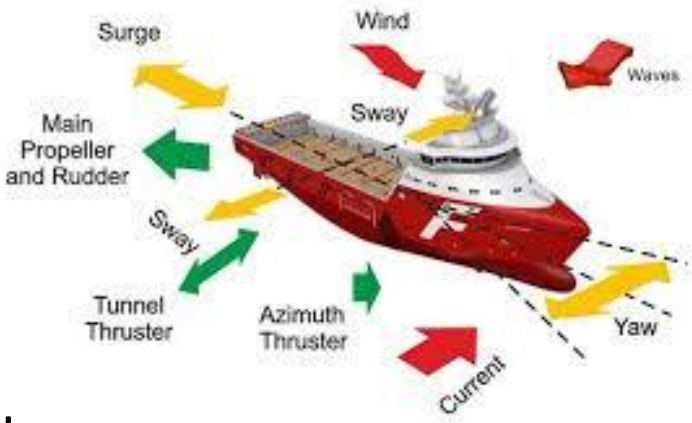
- Hazard was inadvertent launch
- Analyzed right before deployment and field testing (so done late)
  - 2 people, 5 months (unfamiliar with system)
  - Found so many paths to inadvertent launch that deployment delayed six months
- One of first uses of STPA on a real defense system (2005)

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 FMEA)
- 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, human behavior, and component integration and interactions flaws

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

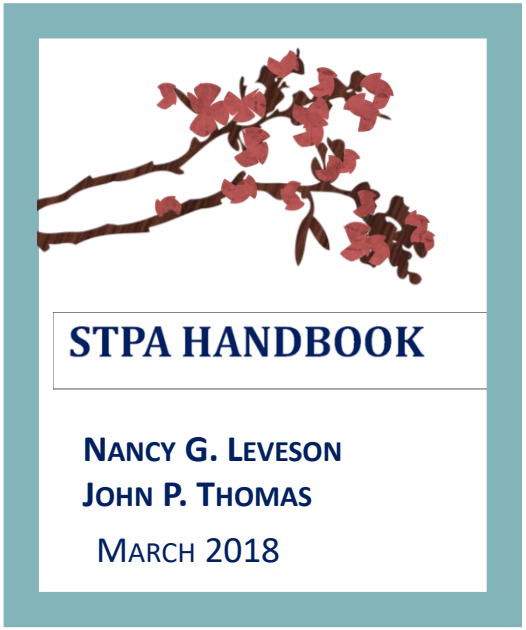
Summary: A Systems Approach to Safety and Security

- Emphasizes building in safety/security rather than measuring it or assuring it
- 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.

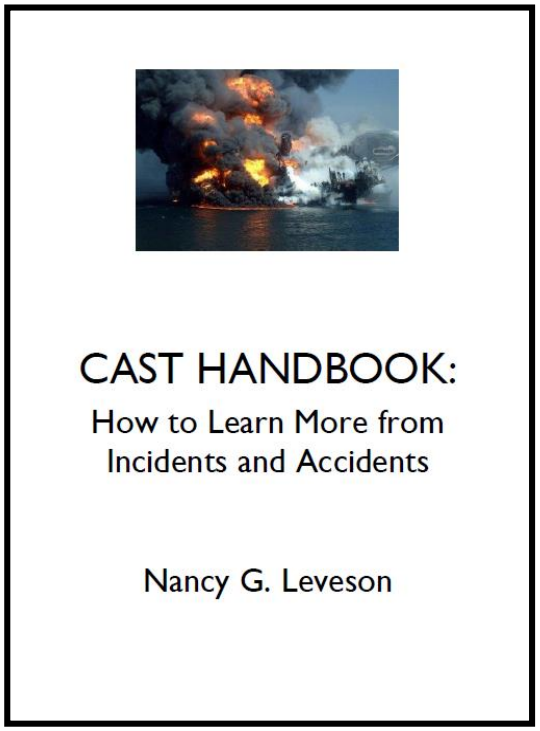
More Information

- 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
(55,000+ downloads in 24 mos.
Japanese, Chinese, and
Korean versions)



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

BACKUP

Paradigm Change

- Does not imply what previously done is wrong and new approach correct

- Einstein:

  "Progress in science (moving from one paradigm to another) is like climbing a mountain"

As move further up, can see farther than on lower points

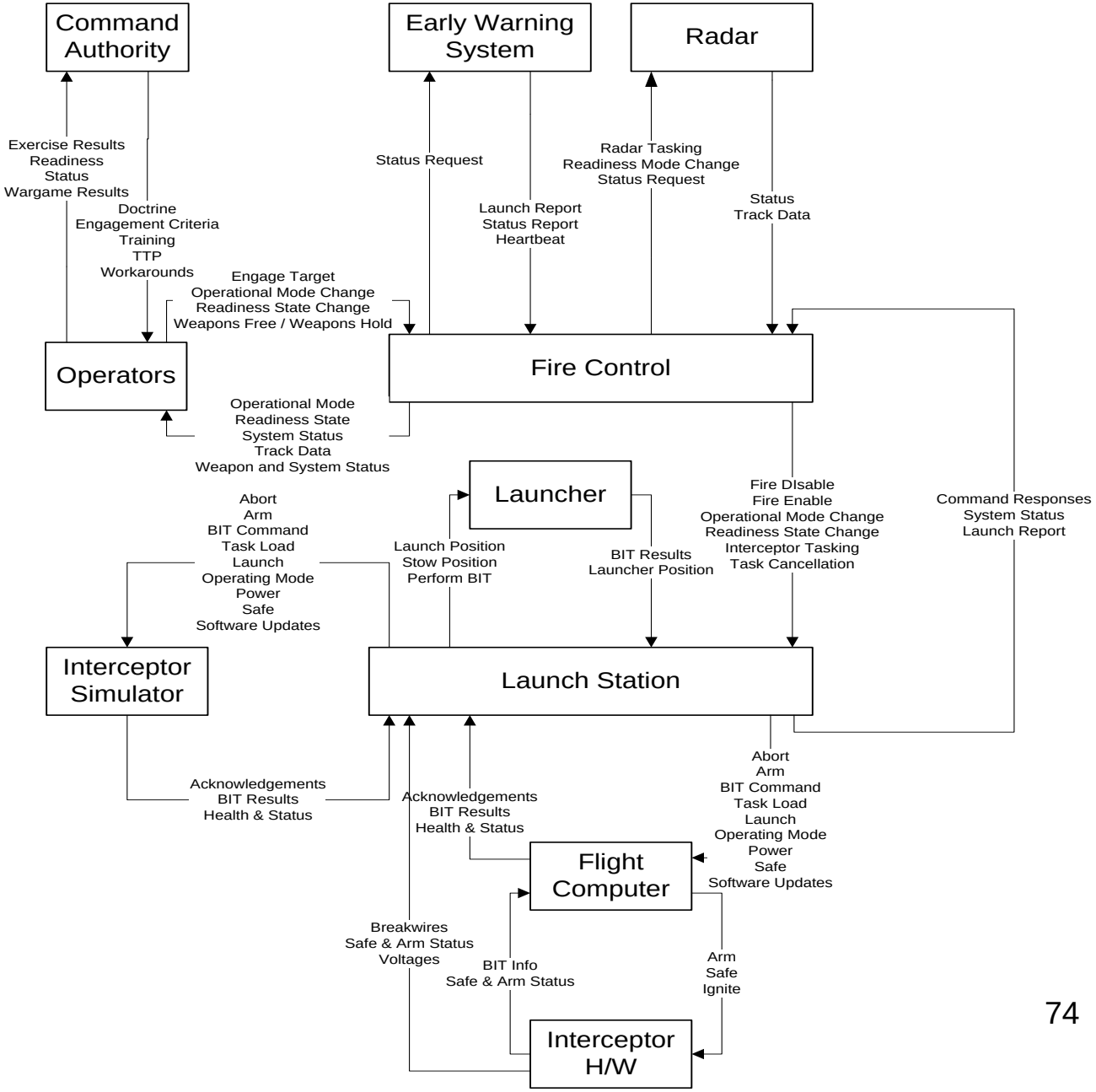
Paradigm Change (2)

New perspective does not invalidate the old one, but extends and enriches our appreciation of the valleys below

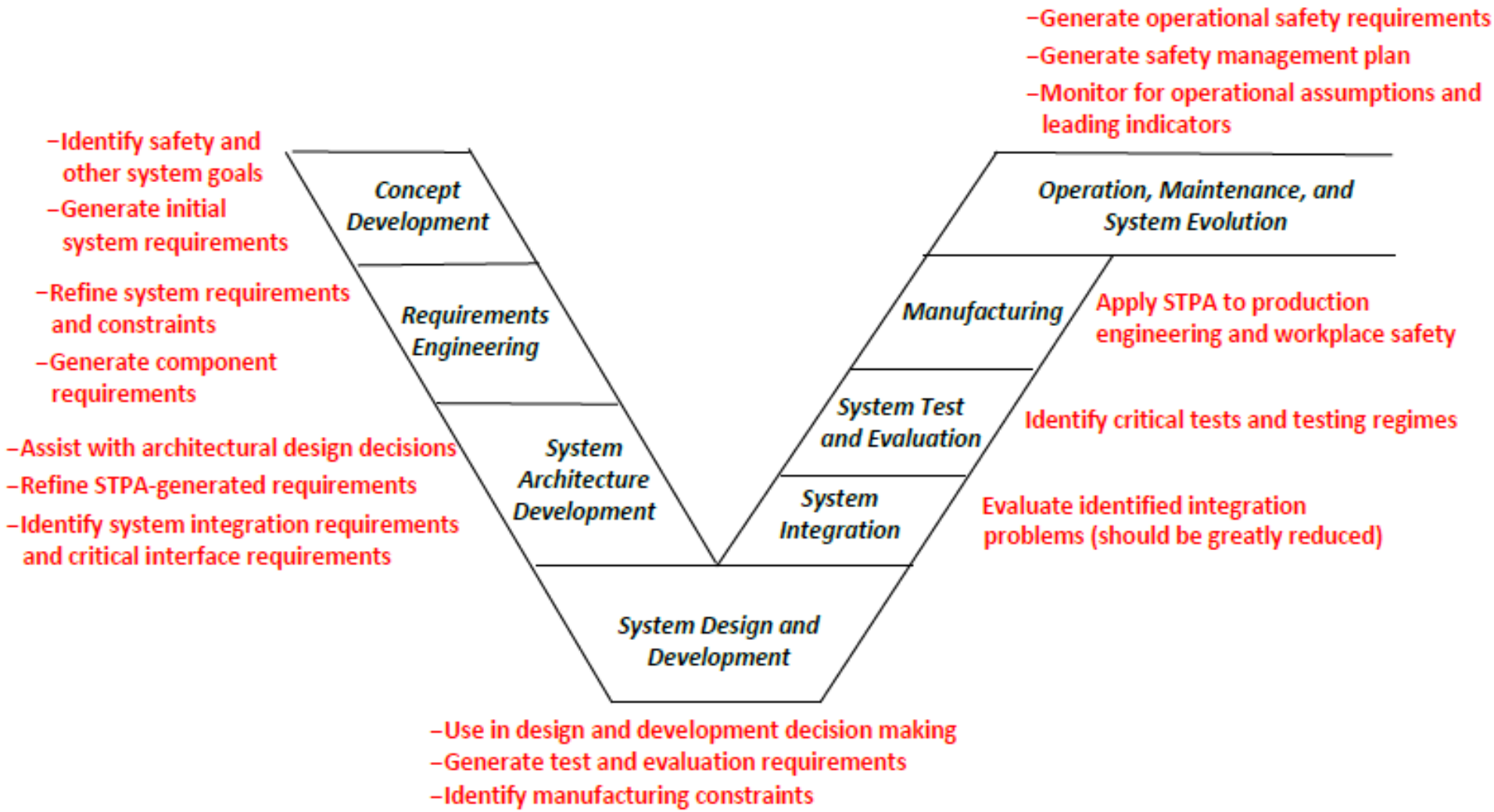
Value of new paradigm often depends on ability to accommodate successes and empirical observations made in old paradigm.

New paradigms offer a broader, richer perspective for interpreting previous answers.

Control Structure for FMIS

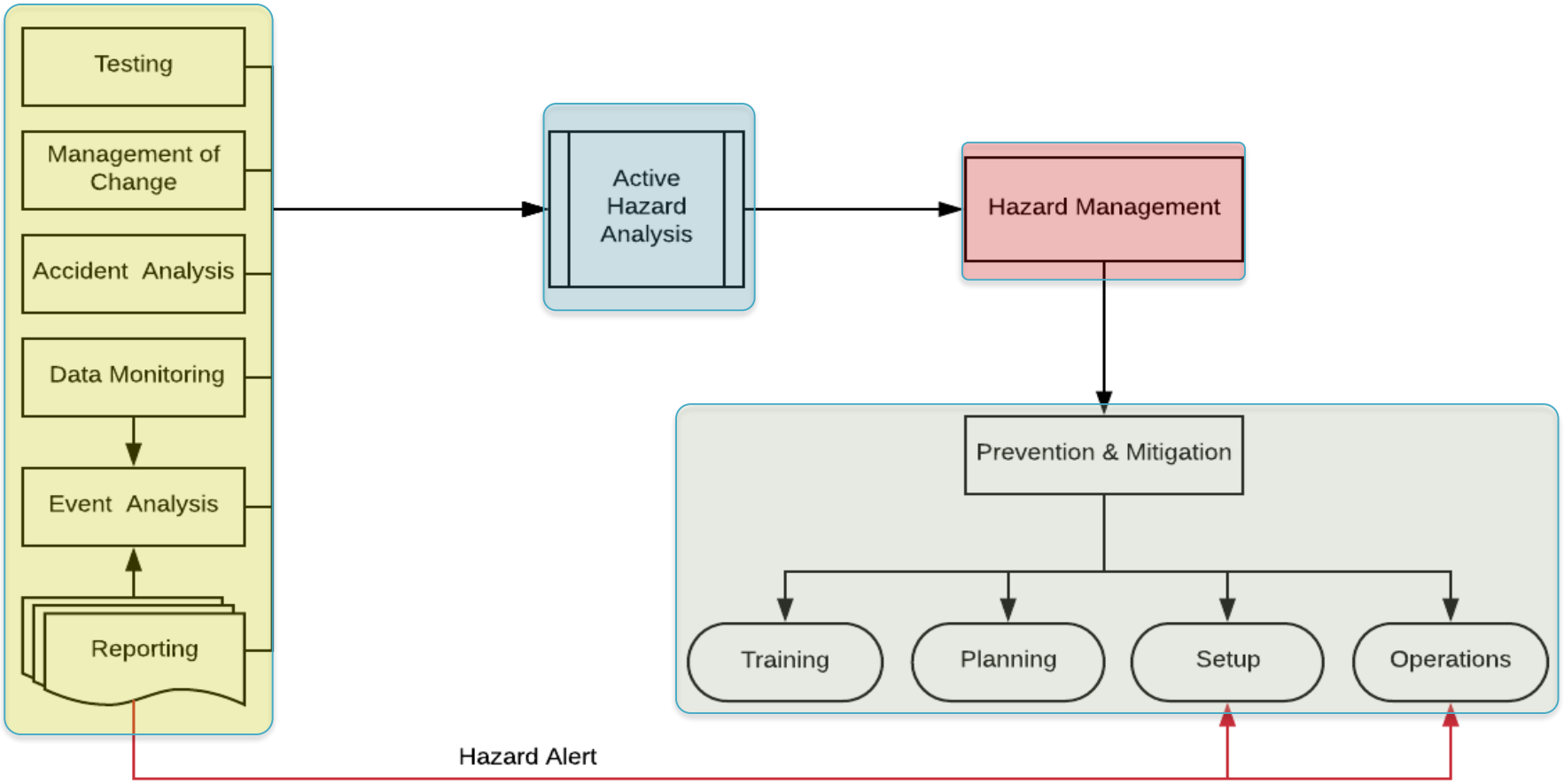


STPA can be used throughout product development and operations

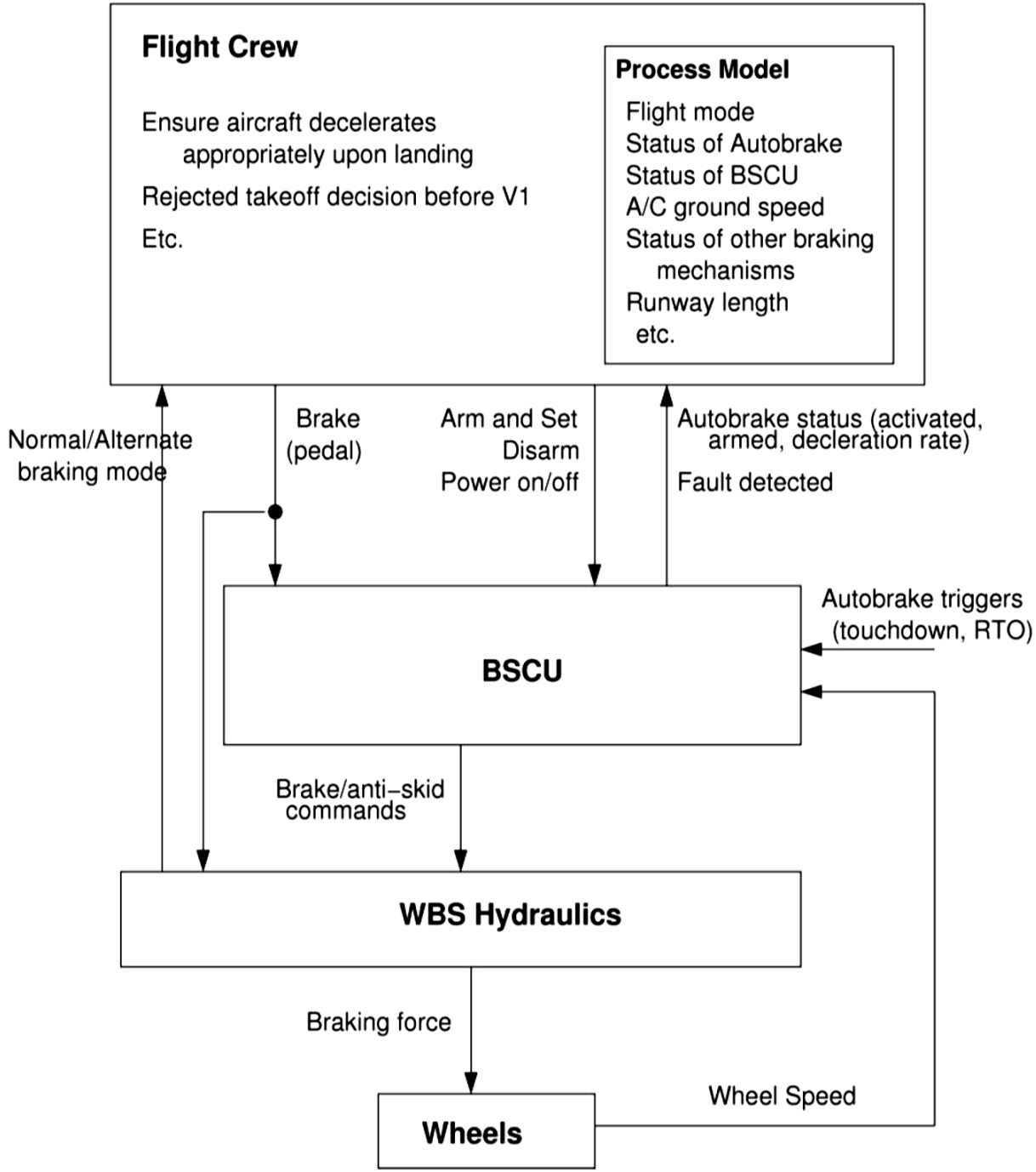


Risk Management During Operations and Leading Indicators (Maj. Diogo Castilho)

- Systems and their environments are not static
- Goal is to detect when risk is increasing (leading indicators)



Wheel Braking System Control Structure

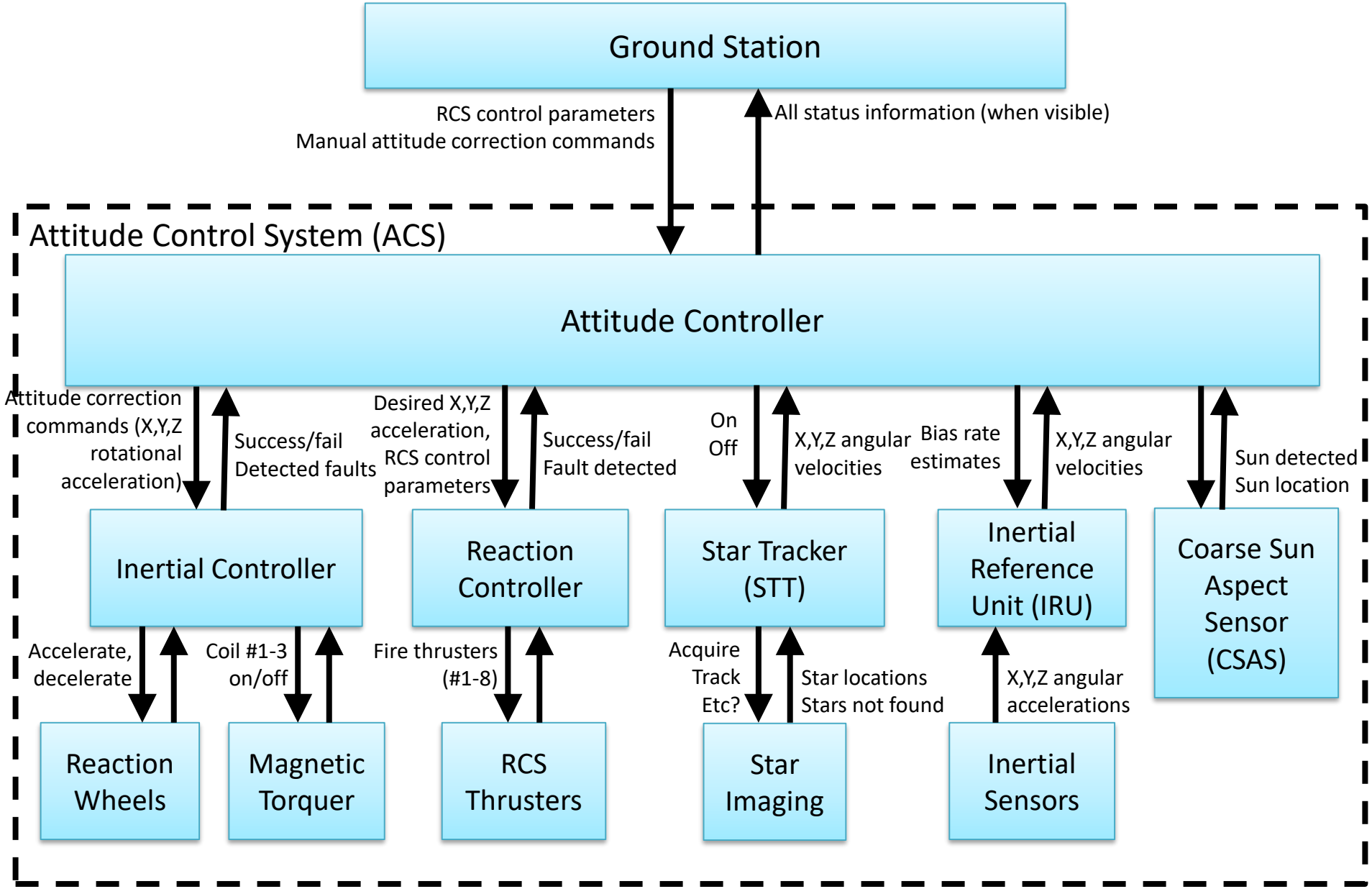


BSCU.1a2: Brake command not provided during landing roll, resulting in insufficient deceleration and potential overshoot

Scenario 1: Autobrake believes the desired deceleration rate has already been achieved or exceeded (incorrect process model). The reasons Autobrake may have this process model flaw include:

- If wheel speed feedback influences the deceleration rate determined by the Autobrake controller, inadequate wheel speed feedback may cause this scenario. Rapid pulses in the feedback (e.g. wet runway, brakes pulsed by anti-skid) could make the actual aircraft speed difficult to detect and an incorrect aircraft speed might be assumed.
- Inadequate external speed/deceleration feedback could explain the incorrect Autobrake process model (e.g. inertial reference drift, calibration issues, sensor failure, etc.)
- **[Security related scenarios, e.g., intruder changes process model]**

One Possible Requirement for S1: Provide additional feedback to Autobrake to detect aircraft deceleration rate in the event of wheel slipping (e.g. fusion of multiple sensors)


Ground Station
RCS control parameters
Manual attitude correction commands
All status information (when visible)
Attitude Control System (ACS)
Attitude Controller
Attitude correction commands (X,Y,Z rotational acceleration)
Success/fail
Detected faults
Desired X,Y,Z acceleration, RCS control parameters
Success/fail
Fault detected
On
Off
X,Y,Z angular velocities
Bias rate estimates
X,Y,Z angular velocities
Sun detected
Sun location
Inertial Controller
Reaction Controller
Star Tracker (STT)
Inertial Reference Unit (IRU)
Coarse Sun Aspect Sensor (CSAS)
Accelerate, decelerate
Coil #1-3 on/off
Fire thrusters (#1-8)
Acquire
Track
Etc?
Star locations
Stars not found
X,Y,Z angular accelerations
Reaction Wheels
Magnetic Torquer
RCS Thrusters
Star Imaging
Inertial Sensors

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
- Operations

or through social controls

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

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov