Engineering a Safer World

Nancy Leveson

MIT

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Why We Need a New Approach to Safety

"Without changing our patterns of thought, we will not be able to solve the problems we created with our current patterns of thought."

Albert Einstein

- Traditional safety engineering approaches developed for relatively simple electro-mechanical systems
- Accidents in complex, software-intensive systems are changing their nature
- Role of humans in systems is changing
- We need new ways to deal with safety in complex systems

Accident Causality Models

- Underlie all our efforts to engineer for safety
- Explain why accidents occur
- Determine the way we prevent and investigate accidents
- May not be aware you are using one, but you are
- Imposes patterns on accidents

"All models are wrong, some models are useful"

George Box

Introduction to Systems Theory

Ways to cope with complexity

1. Analytic Reduction
2. Statistics

[Recommended reading: Peter Checkland, “Systems Thinking, Systems Practice,” John Wiley, 1981]

Analytic Reduction

- Divide system into distinct parts for analysis

  Physical aspects → Separate physical components

  Behavior → Events over time

- Examine parts separately

- Assumes such separation does not distort phenomenon
  - Each component or subsystem operates independently
  - Analysis results not distorted when consider components separately
  - Components act the same when examined singly as when playing their part in the whole
  - Events not subject to feedback loops and non-linear interactions

Chain-of-Events Accident Causality Model

- Explains accidents in terms of multiple events, sequenced as a forward chain over time.
  - Simple, direct relationship between events in chain

- Events almost always involve component failure, human error, or energy-related event

- Forms the basis for most safety engineering and reliability engineering analysis:

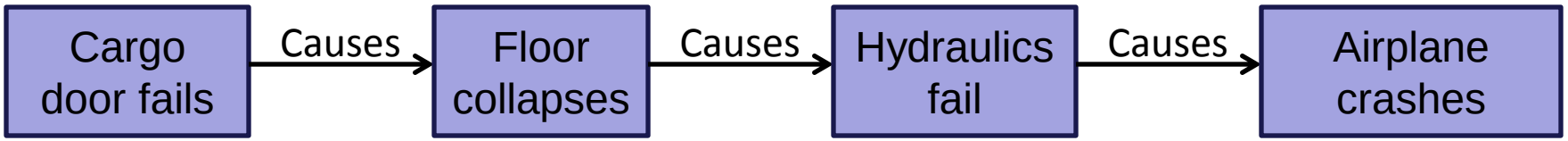
  e,g, FTA, PRA, FMECA, Event Trees, etc.

  and design:

  e.g., redundancy, overdesign, safety margins, ….

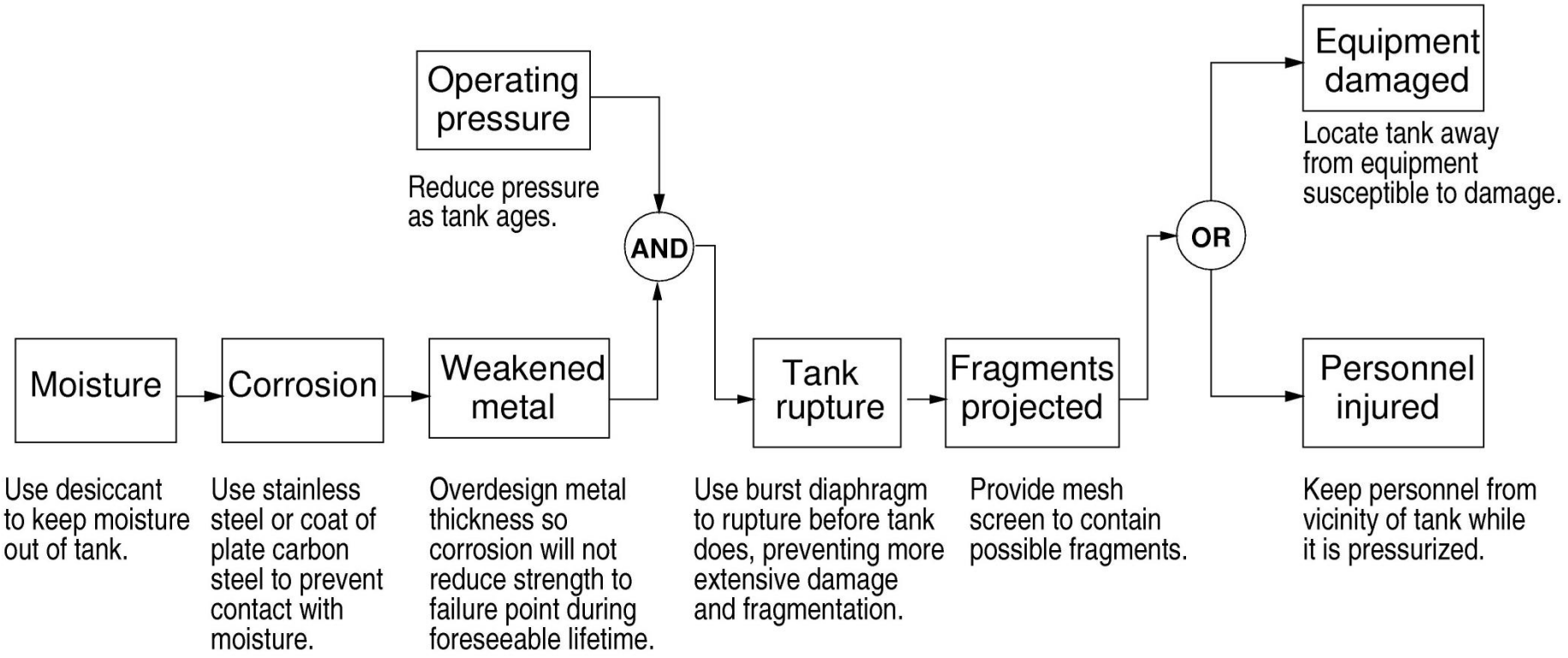
Domino “Chain of events” Model

DC-10:

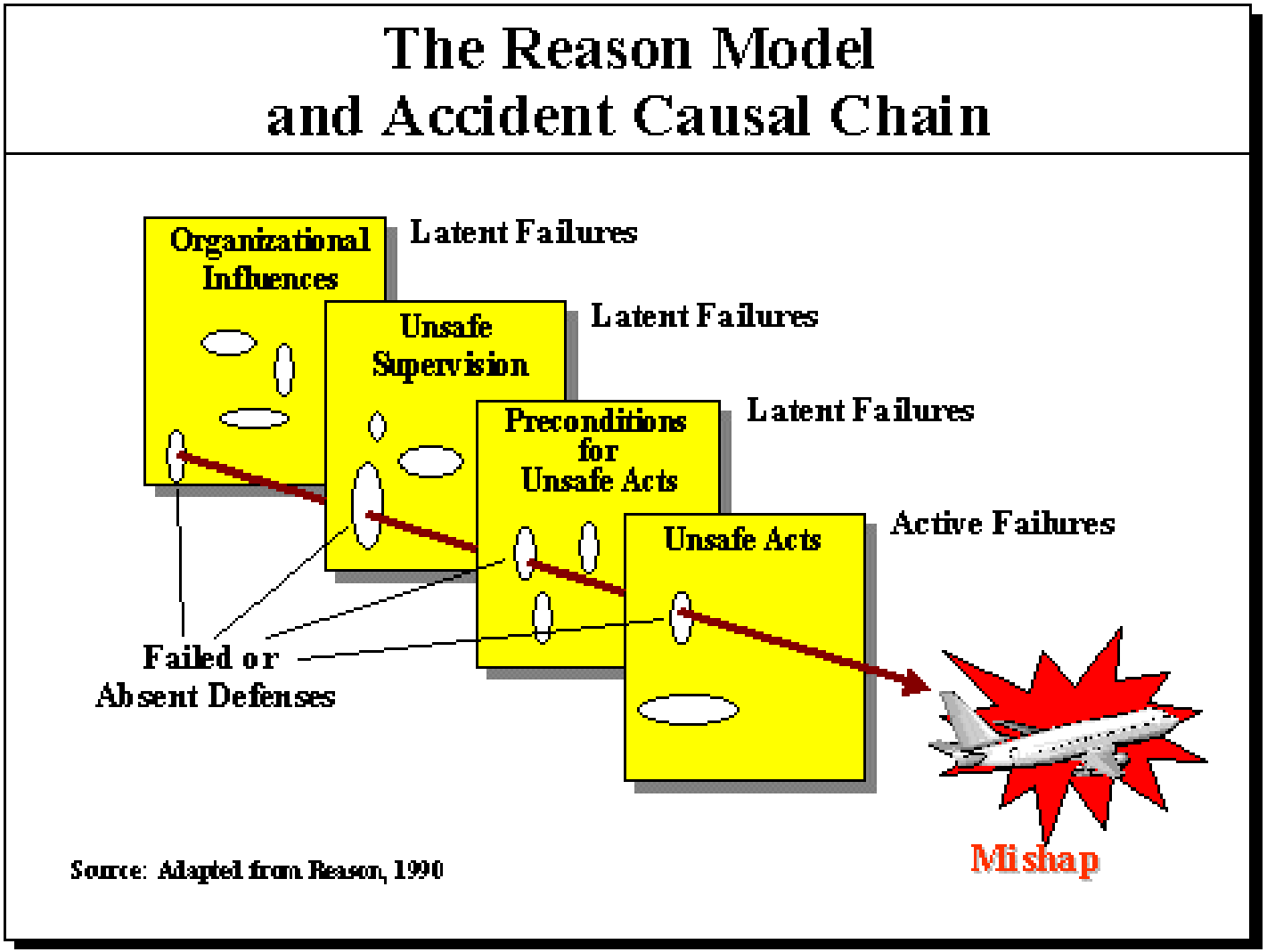


Event-based

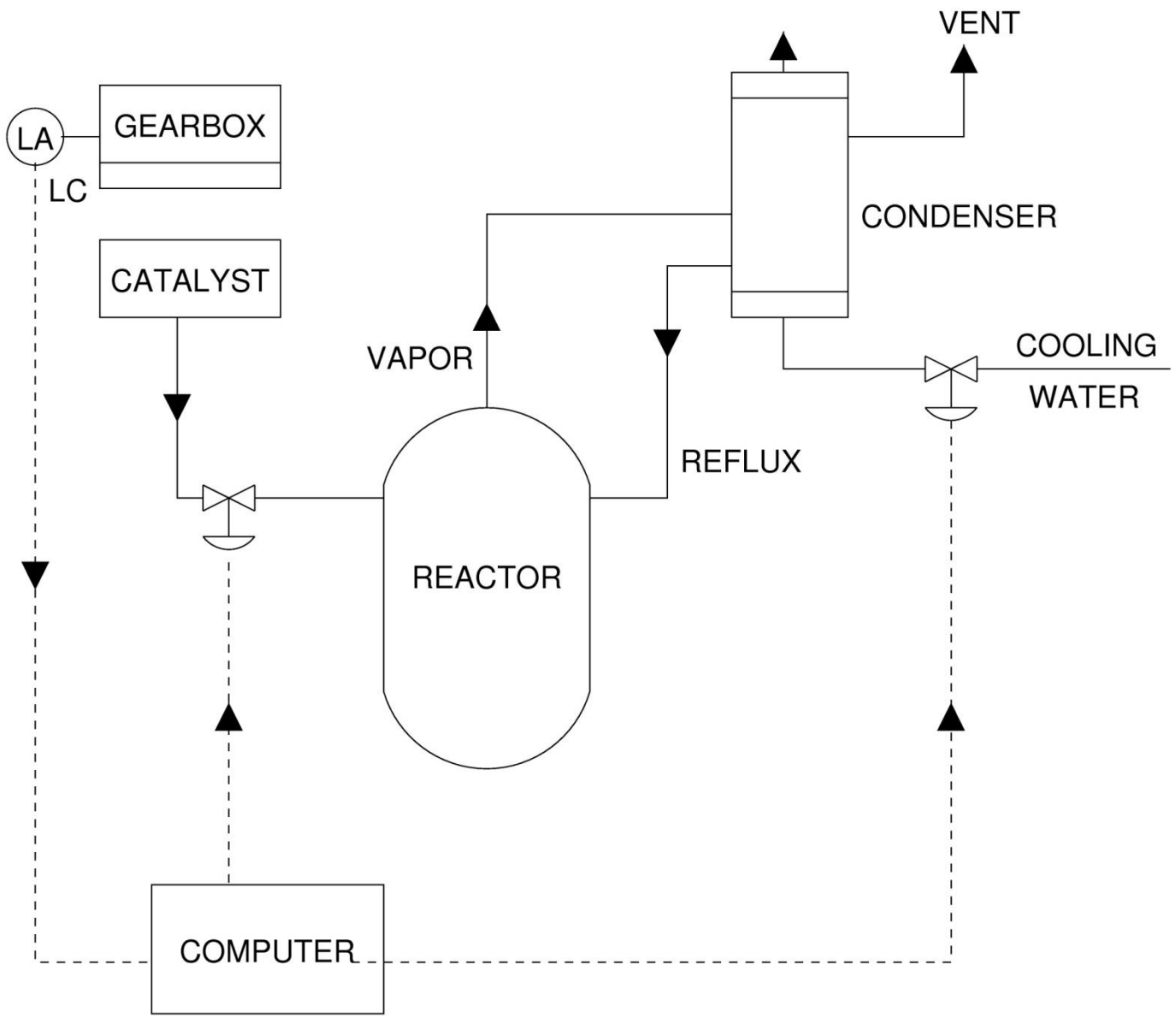
Chain-of-events example



Reason Swiss Cheese



Accident with No Component Failures



Types of Accidents

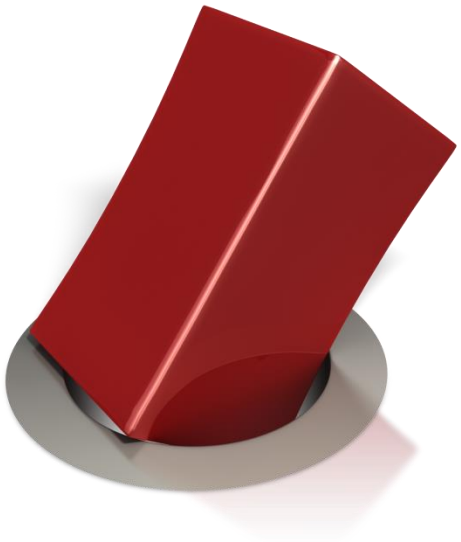
- Component Failure Accidents
  - Single or multiple component failures
  - Usually assume random failure
- Component Interaction Accidents
  - Arise in interactions among components
  - Related to interactive and dynamic complexity
  - Behavior can no longer be
    - Planned
    - Understood
    - Anticipated
    - Guarded against
  - Exacerbated by introduction of computers and software

Analytic Reduction does not Handle

- Component interaction accidents
- Systemic factors (affecting all components and barriers)
- Software
- Human behavior (in a non-superficial way)
- System design errors
- Indirect or non-linear interactions and complexity
- Migration of systems toward greater risk over time

Summary

- The world of engineering is changing.
- If safety engineering does not change with it, it will become more and more irrelevant.
- Trying to shoehorn new technology and new levels of complexity into old methods does not work



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
- Developed for biology (von Bertalanffly) and engineering (Norbert Weiner)
- Basis of system engineering and system safety

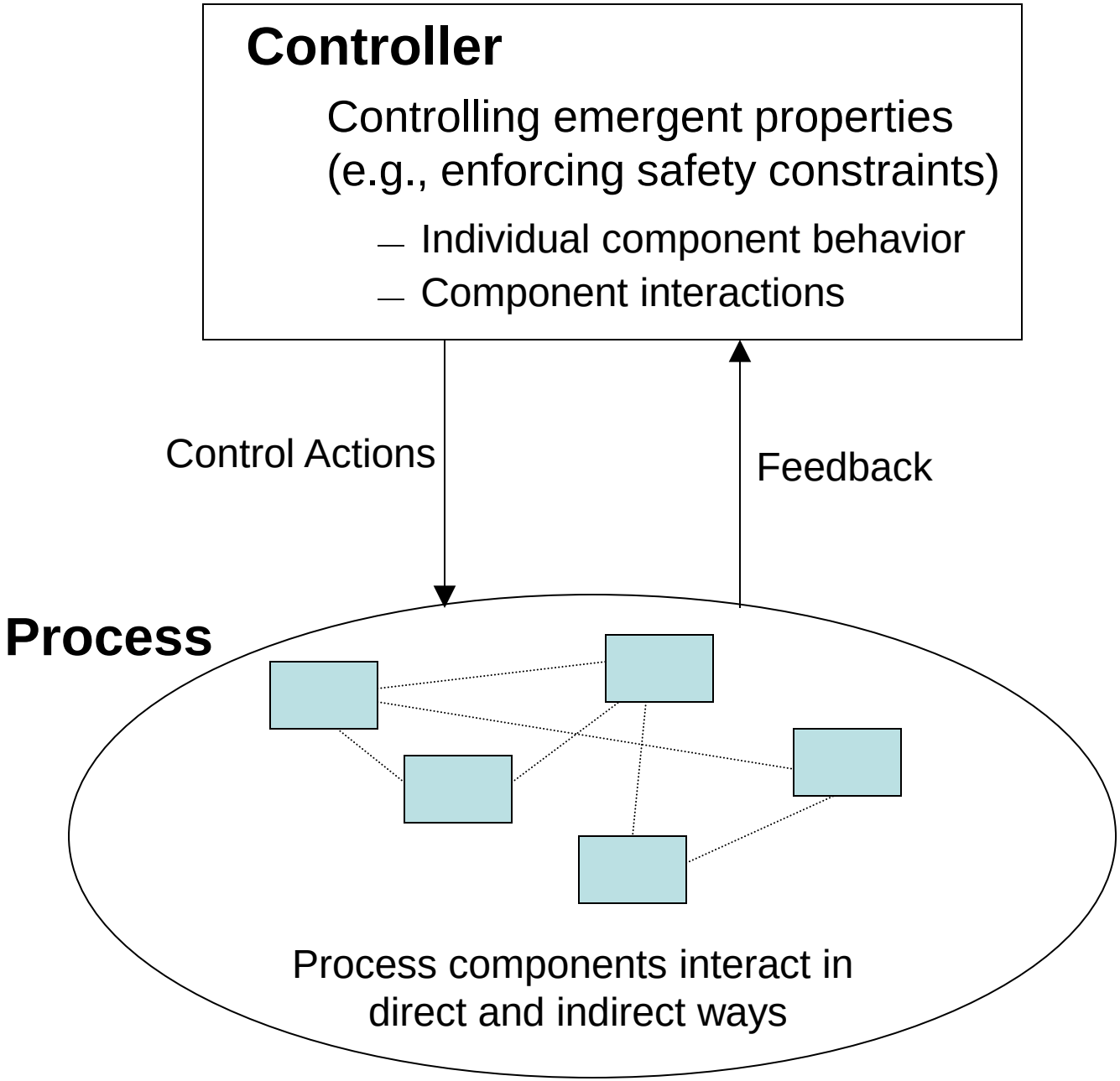
Systems Theory (2)

- 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 derive from relationships among the parts of the system

    How they interact and fit together
- Two pairs of ideas
  1. Hierarchy and emergence
  2. Communication and control


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

Controls/Controllers Enforce Safety Constraints

- Power must never be on when access door open
- Two aircraft must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Public health system must prevent exposure of public to contaminated water and food products
- Pressure in a deep water well must be controlled
- Runway incursions and operations on wrong runways or taxiways must be prevented

A Broad View of “Control”

Component failures and unsafe interactions may be “controlled” through design

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

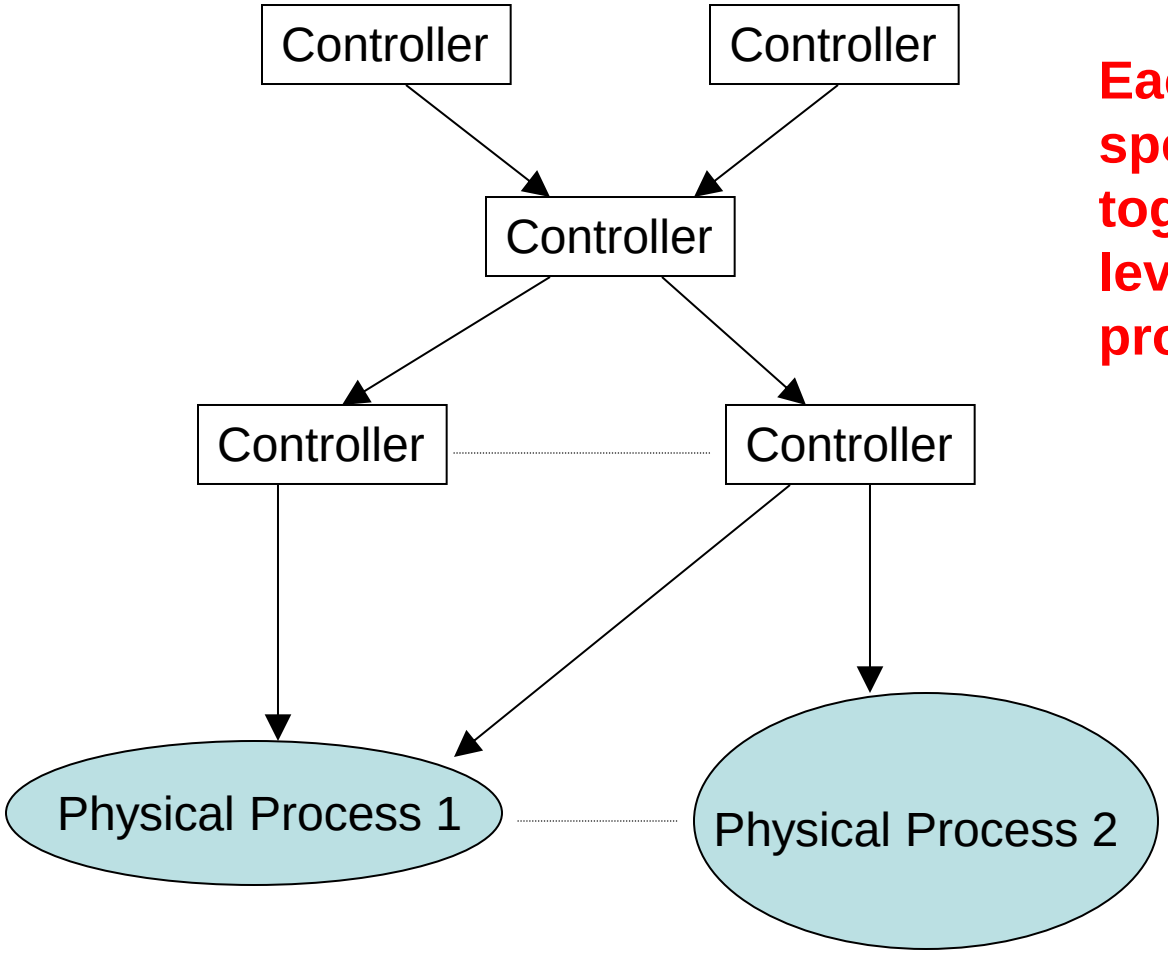
or through process

- 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)

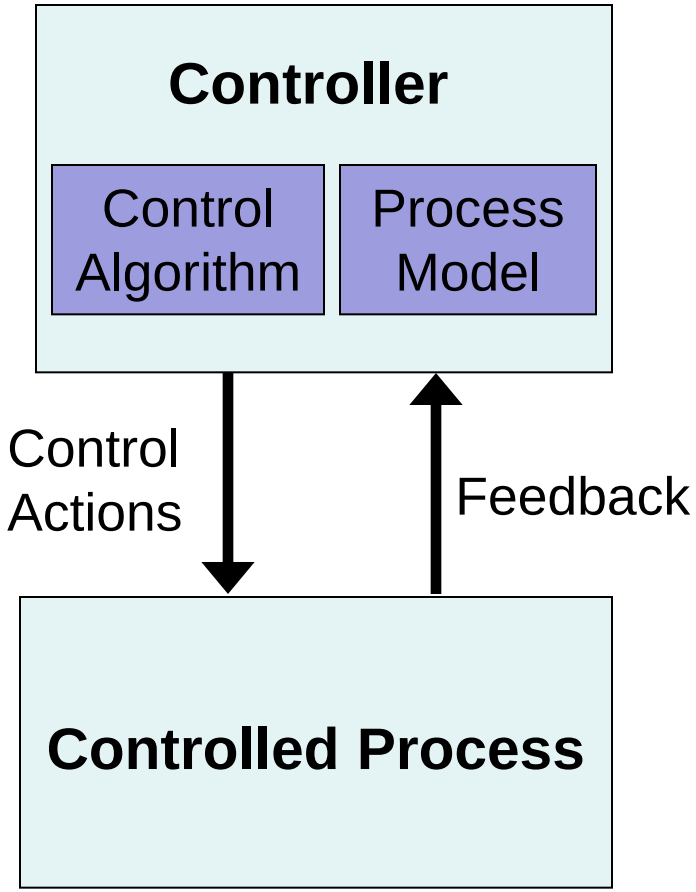
There may be multiple controllers, processes, and levels of control



Each controller enforces specific constraints, which together enforce the system level constraints (emergent properties)

(with various types of communication between them)

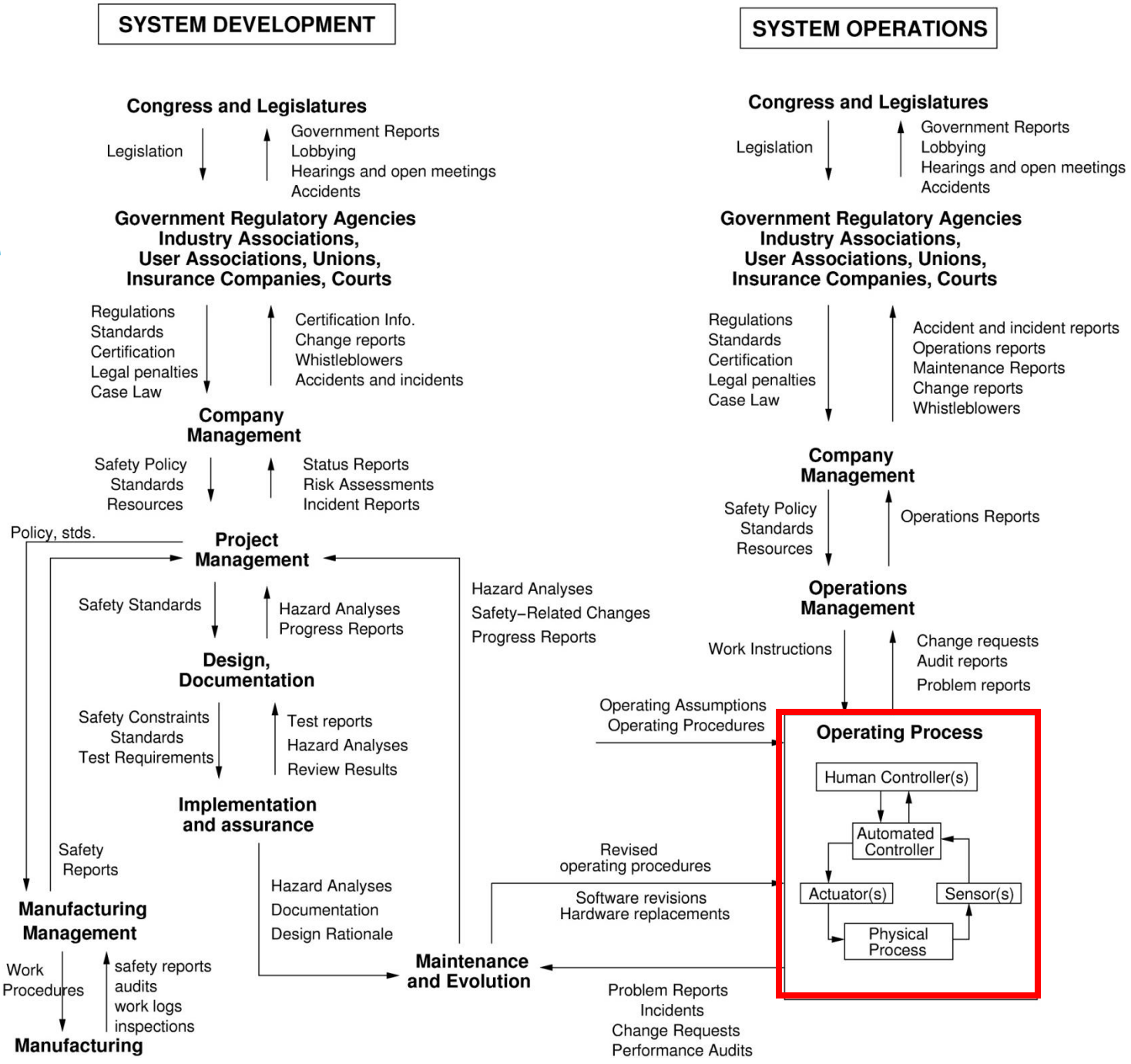
Role of Process Models in Control



- Controllers use a **process model** to determine control actions
- Accidents often occur when the process model is incorrect
  - How could this happen?
- Four types of unsafe control actions:
  - Control commands required for safety are not given
  - Unsafe ones are given
  - Potentially safe commands given too early, too late
  - Control stops too soon or applied too long

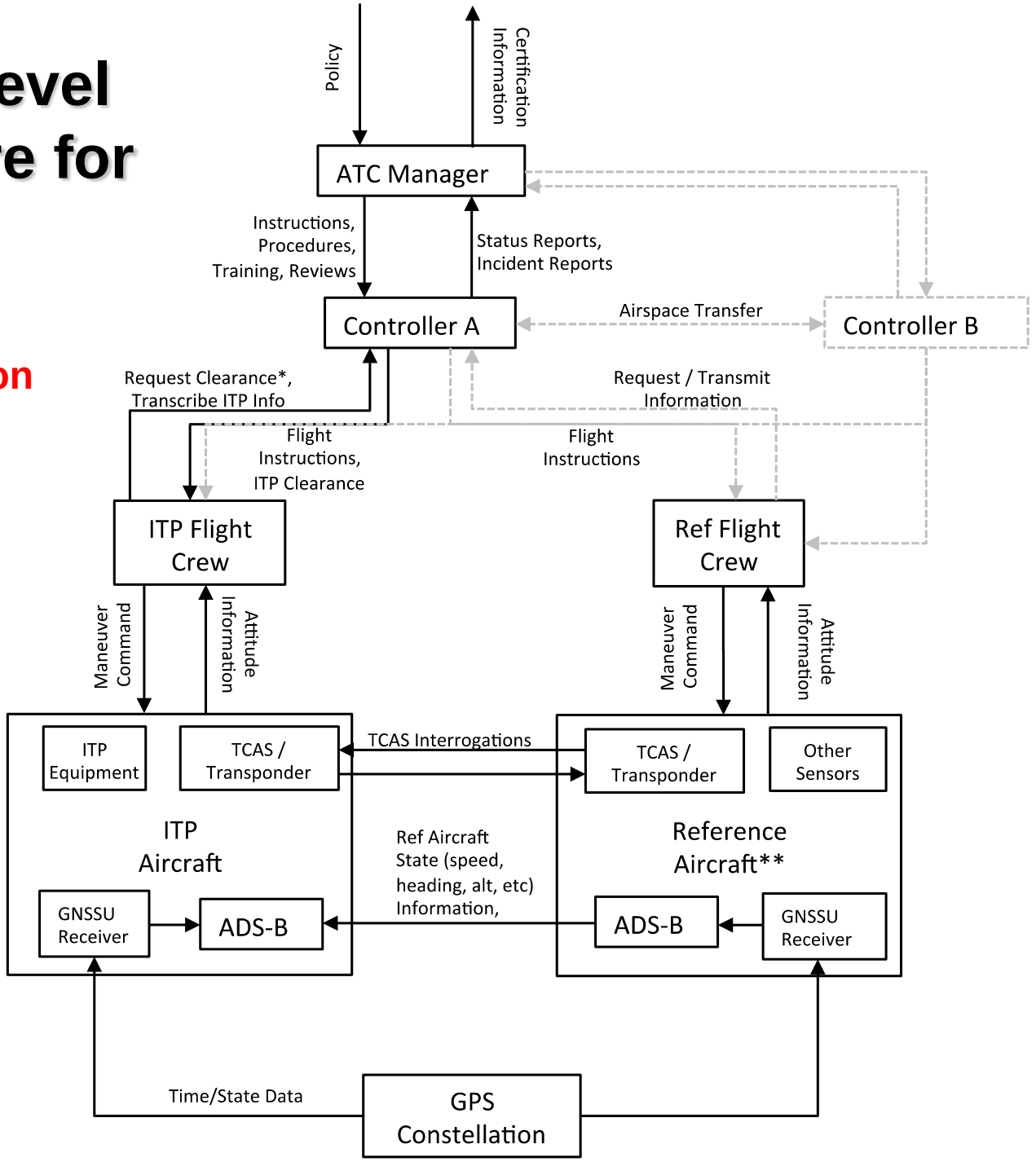
(Leveson, 2003); (Leveson, 2011)

Example Safety Control Structure

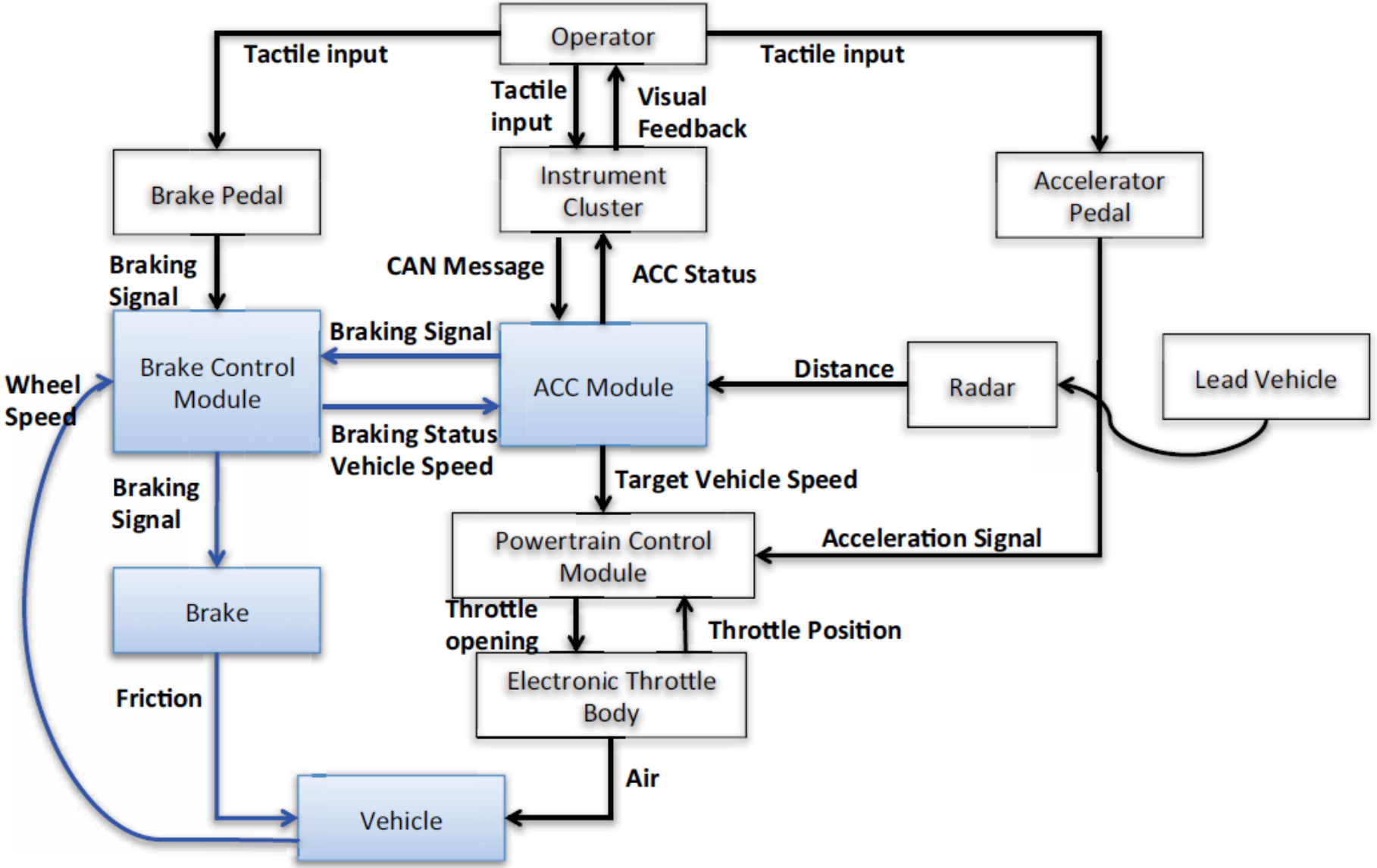


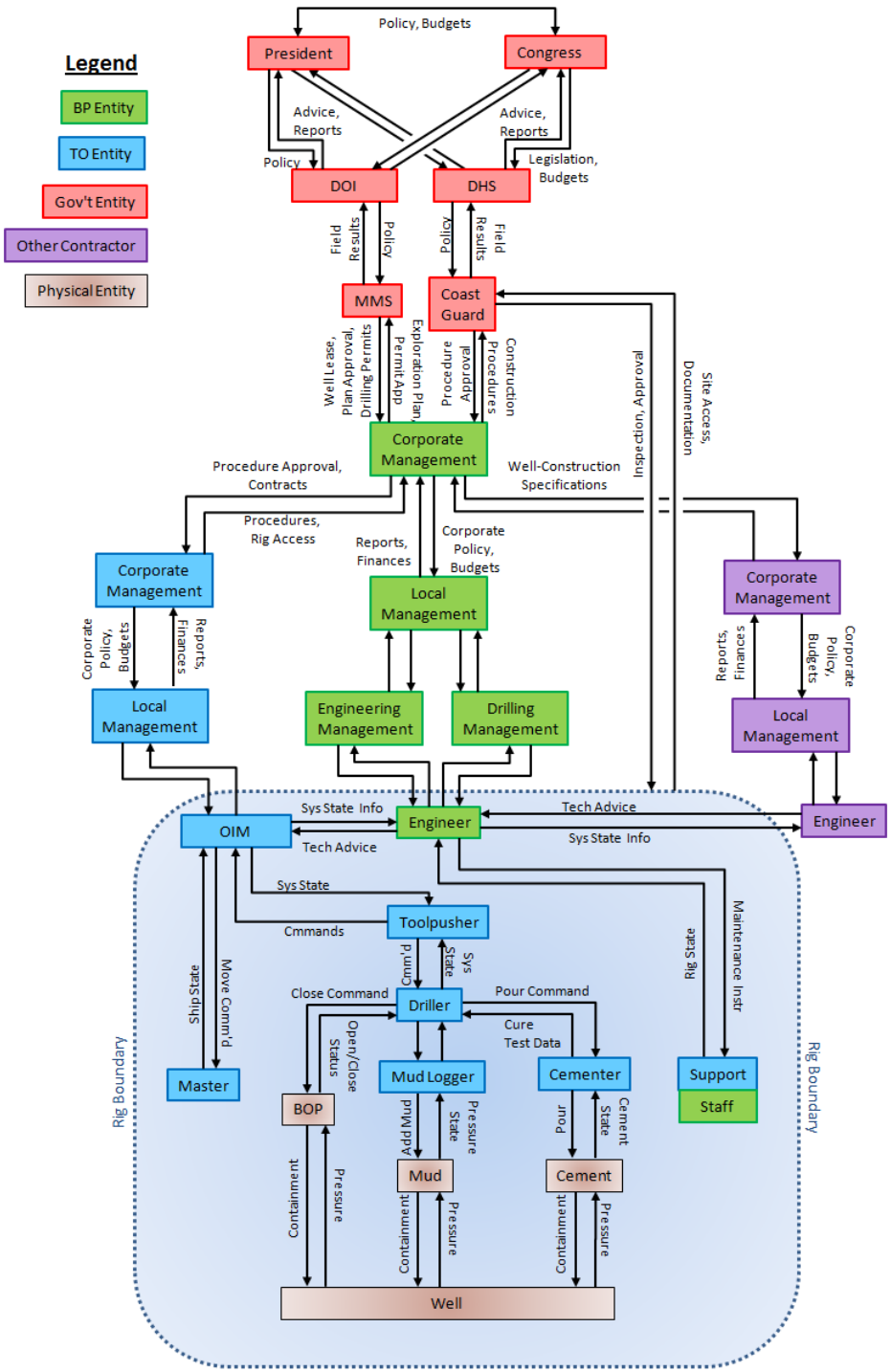
Example High-Level Control Structure for ITP

CONSTRAINTS:
Enforce minimum separation
Maximize throughput



Example: ACC – BCM Control Loop




Legend
BP Entity
TO Entity
Gov't Entity
Other Contractor
Physical Entity
Policy, Budgets
President
Congress
Advice, Reports
Policy
Advice, Reports
Legislation, Budgets
DOI
DHS
Field Results
Policy
Policy
Field Results
MMS
Coast Guard
Well Lease, Plan Approval, Drilling Permits
Exploration Plan, Permit App
Procedure Approval
Construction Procedures
Inspection, Approval
Site Access, Documentation
Corporate Management
Procedure Approval, Contracts
Well-Construction Specifications
Procedures, Rig Access
Reports, Finances
Corporate Policy, Budgets
Corporate Management
Local Management
Corporate Management
Corporate Policy, Budgets
Reports, Finances
Corporate Policy, Budgets
Reports, Finances
Local Management
Engineering Management
Drilling Management
Local Management
Rig Boundary
Sys State Info
Tech Advice
OIM
Engineer
Tech Advice
Sys State Info
Engineer
Sys State
Cmmands
Toolpusher
Ship State
Move Comm'd
Comm'd
Sys State
Rig State
Maintenance Instr
Close Command
Driller
Pour Command
Cure Test Data
Open/Close Status
Add Mud
Pressure State
Pour
Cement State
Master
BOP
Mud Logger
Cementer
Support Staff
Containment
Pressure
Mud
Containment
Pressure
Cement
Containment
Pressure
Well
Rig Boundary

STAMP: System-Theoretic Accident Model and Processes

Based on Systems Theory
(vs. Reliability Theory)

STAMP: Safety as a Control Problem

- Safety is an emergent property that arises when system components interact with each other within a larger environment
  - A set of constraints related to behavior of system components (physical, human, social) enforces that property
  - Accidents occur when interactions violate those constraints (a lack of appropriate constraints on the interactions)
- Goal is to control the behavior of the components and systems as a whole to ensure safety constraints are enforced in the operating system.

Safety as a Dynamic Control Problem

- Examples
  - O-ring did not control propellant gas release by sealing gap in field joint of Challenger Space Shuttle
  - Software did not adequately control descent speed of Mars Polar Lander
  - At Texas City, did not control the level of liquids in the ISOM tower;
  - In DWH, did not control the pressure in the well;
  - Financial system did not adequately control the use of financial instruments

Safety as a Dynamic Control Problem

- Events are the result of the inadequate control
  - Result from lack of enforcement of safety constraints in system design and operations
- Systems are dynamic processes that are continually changing and adapting to achieve their goals
- A change in emphasis:

~~"prevent failures"~~

"enforce safety constraints on system behavior"

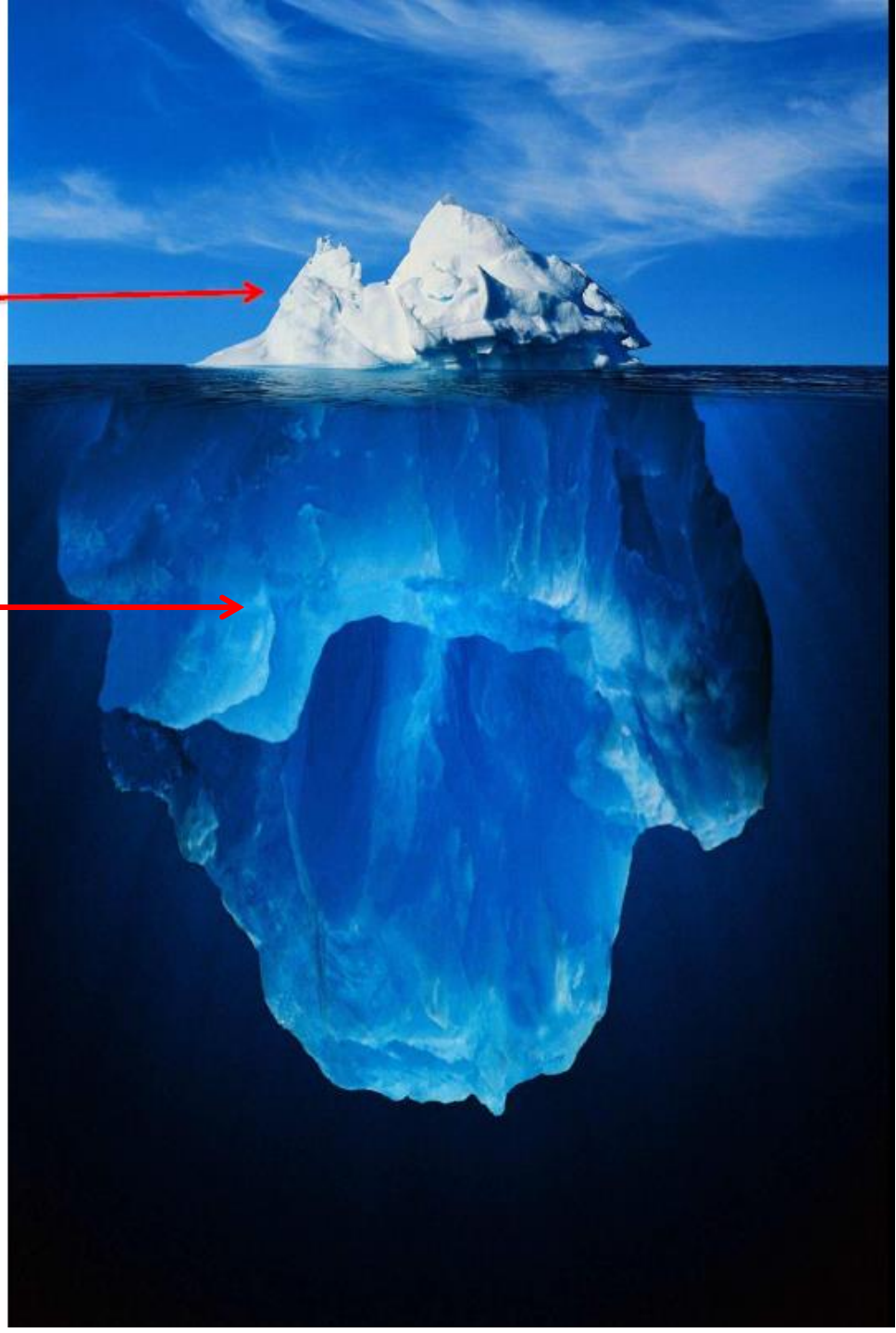
Changes to Analysis Goals

- Hazard analysis:
  - Ways that safety constraints might not be enforced
    (vs. chains of failure events leading to accident)
- Accident Analysis (investigation)
  - Why control structure was not adequate to prevent loss
    (vs. what failures led to loss and who responsible)
- Security Analysis
  - Potential weaknesses in security controls
    (vs. threat analysis)

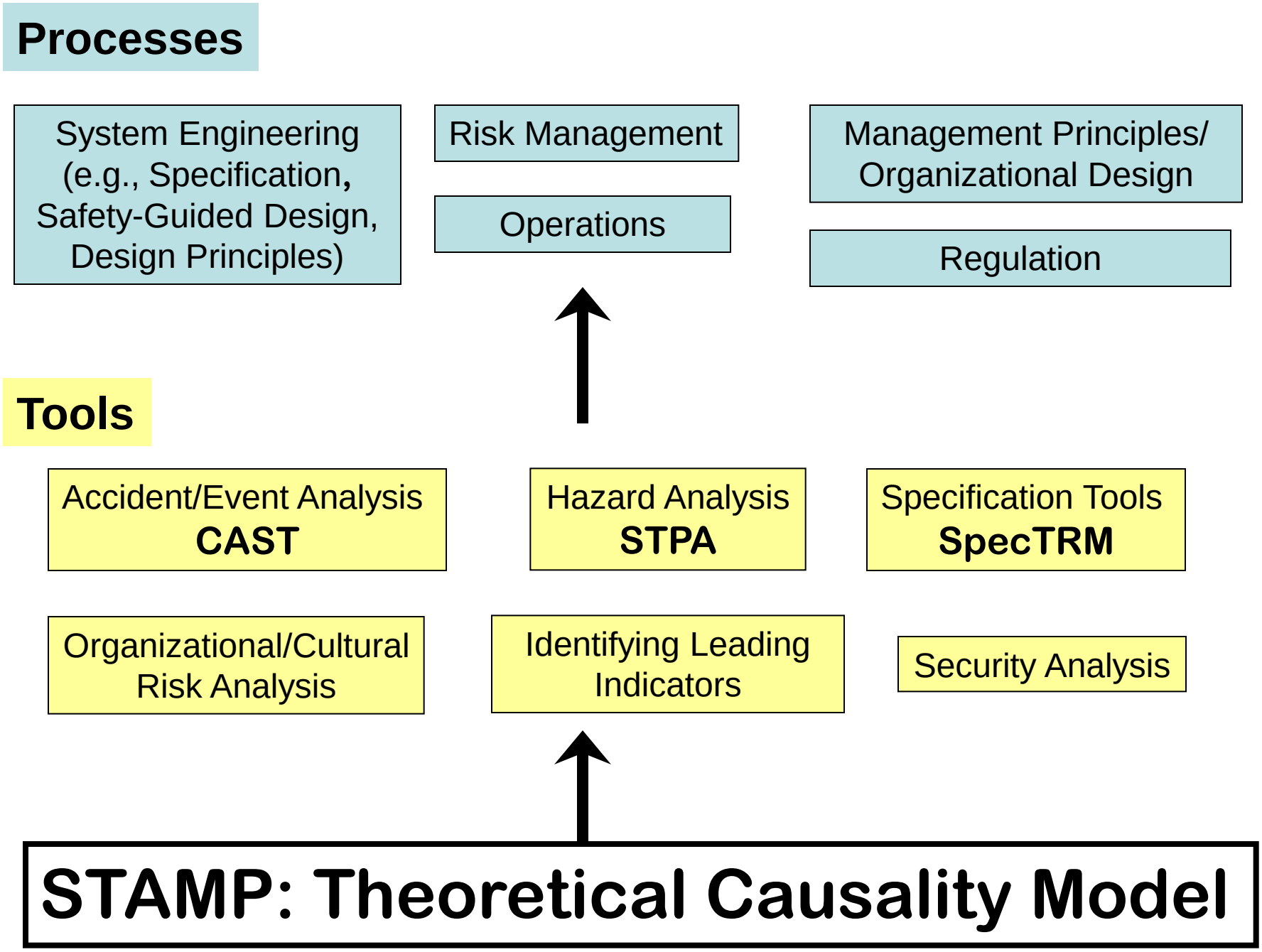
LEVERAGE

Event-based thinking

Systems Thinking



WYOUNG@MIT.EDU

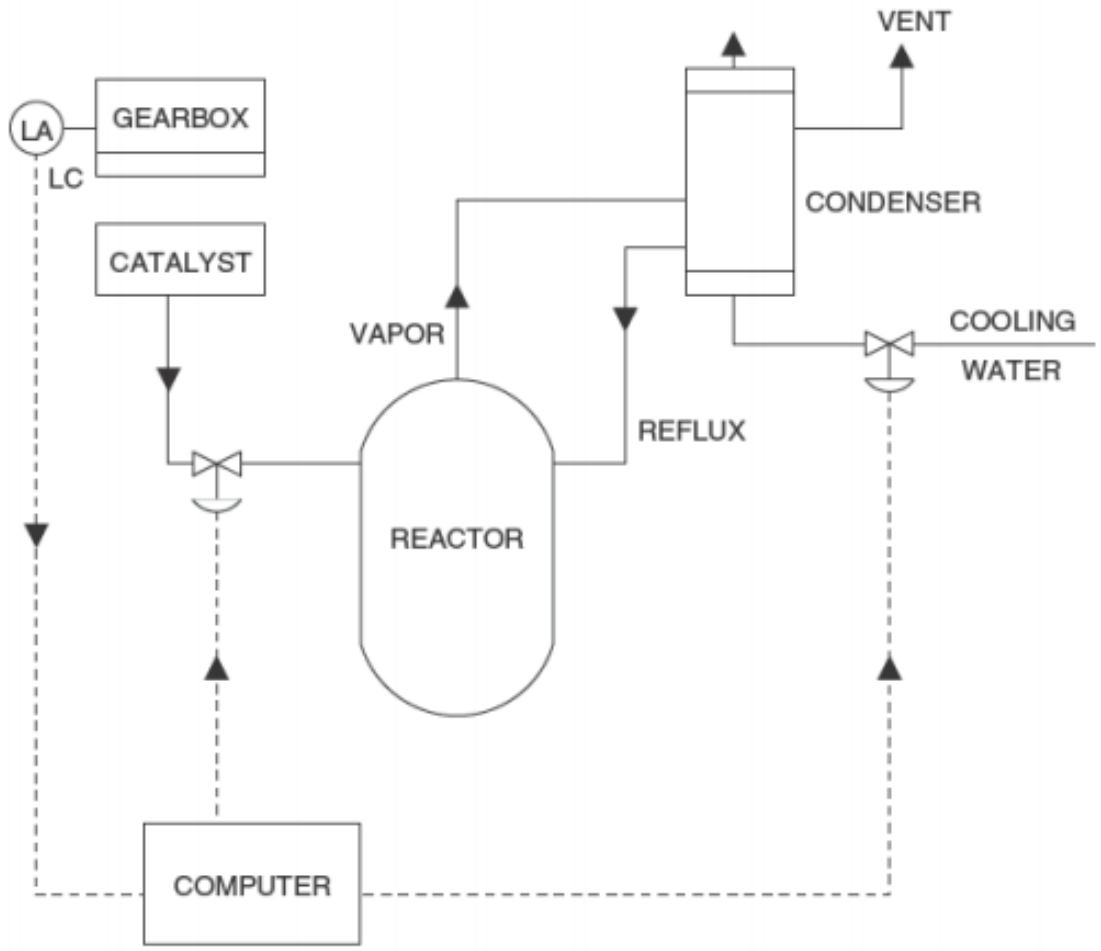

Processes
System Engineering (e.g., Specification, Safety-Guided Design, Design Principles)
Risk Management
Operations
Management Principles/ Organizational Design
Regulation
Tools
Accident/Event Analysis
CAST
Hazard Analysis
STPA
Specification Tools
SpecTRM
Organizational/Cultural Risk Analysis
Identifying Leading Indicators
Security Analysis
STAMP: Theoretical Causality Model

Outline

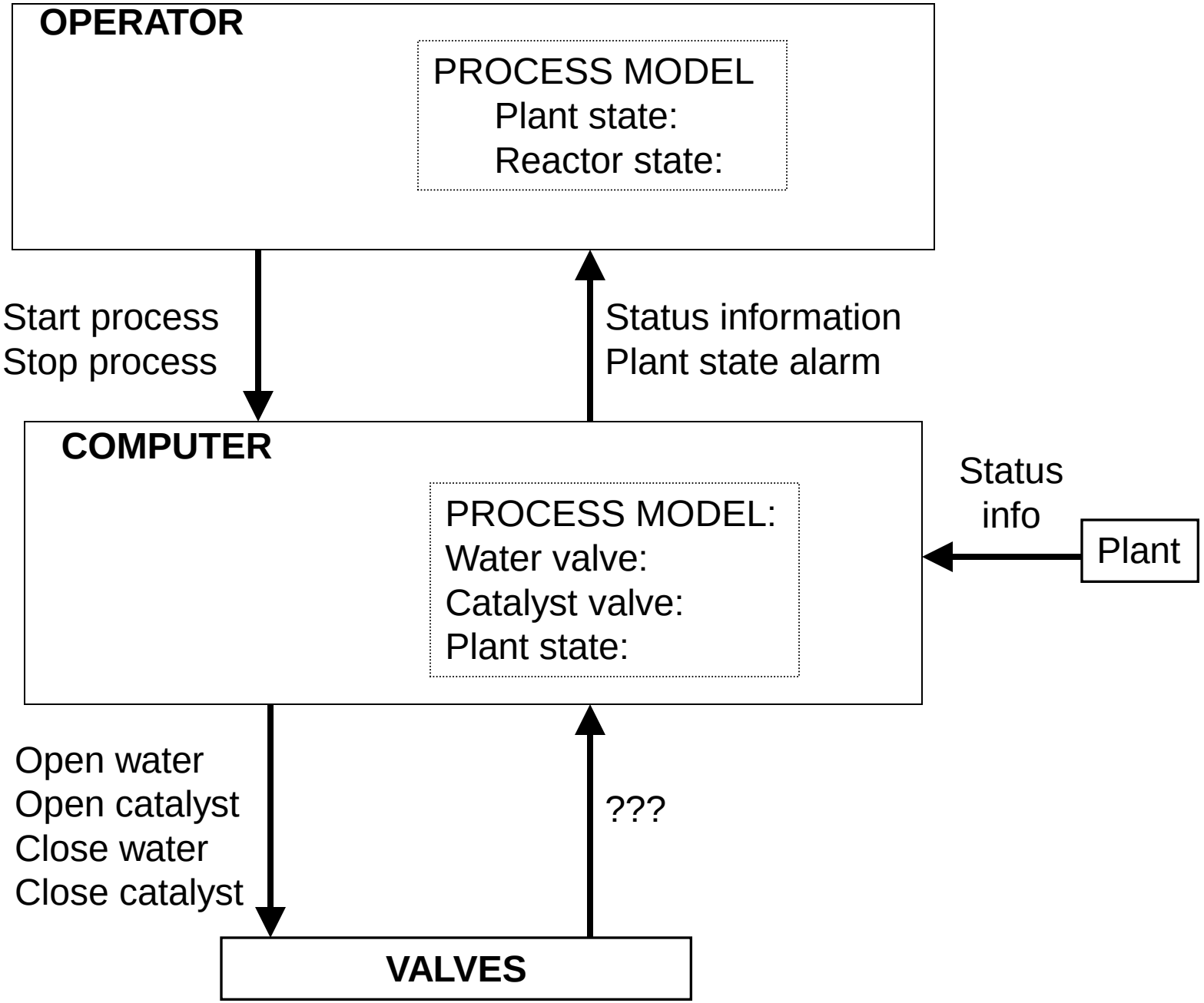
- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

STPA: System-Theoretic Process Analysis

- Integrated into system engineering
  - Can be used from beginning of project
  - Safety-guided design
  - Guidance for evaluation and test
  - Incident/accident analysis
- Works also on social and organizational aspects of systems
- Generates system and component safety requirements (constraints)
- Identifies flaws in system design and scenarios leading to violation of a safety requirement (i.e., a hazard)



Create functional control structure for this physical structure


OPERATOR
PROCESS MODEL
Plant state:
Reactor state:
Start process
Stop process
Status information
Plant state alarm
COMPUTER
PROCESS MODEL:
Water valve:
Catalyst valve:
Plant state:
Status
info
Plant
Open water
Open catalyst
Close water
Close catalyst
???
VALVES

Identifying Unsafe Control Actions

Hazard: Catalyst in reactor without reflux condenser operating (water flowing through it)

	Not providing causes hazard	Providing causes hazard	Incorrect Timing/ Order	Stopped Too Soon / Applied too long
Open Water Valve	Water valve not opened when catalyst open			
Close Water Valve				
Open Catalyst Valve				
Close Catalyst				

Hazard: Catalyst in reactor without reflux condenser operating (water flowing through it)

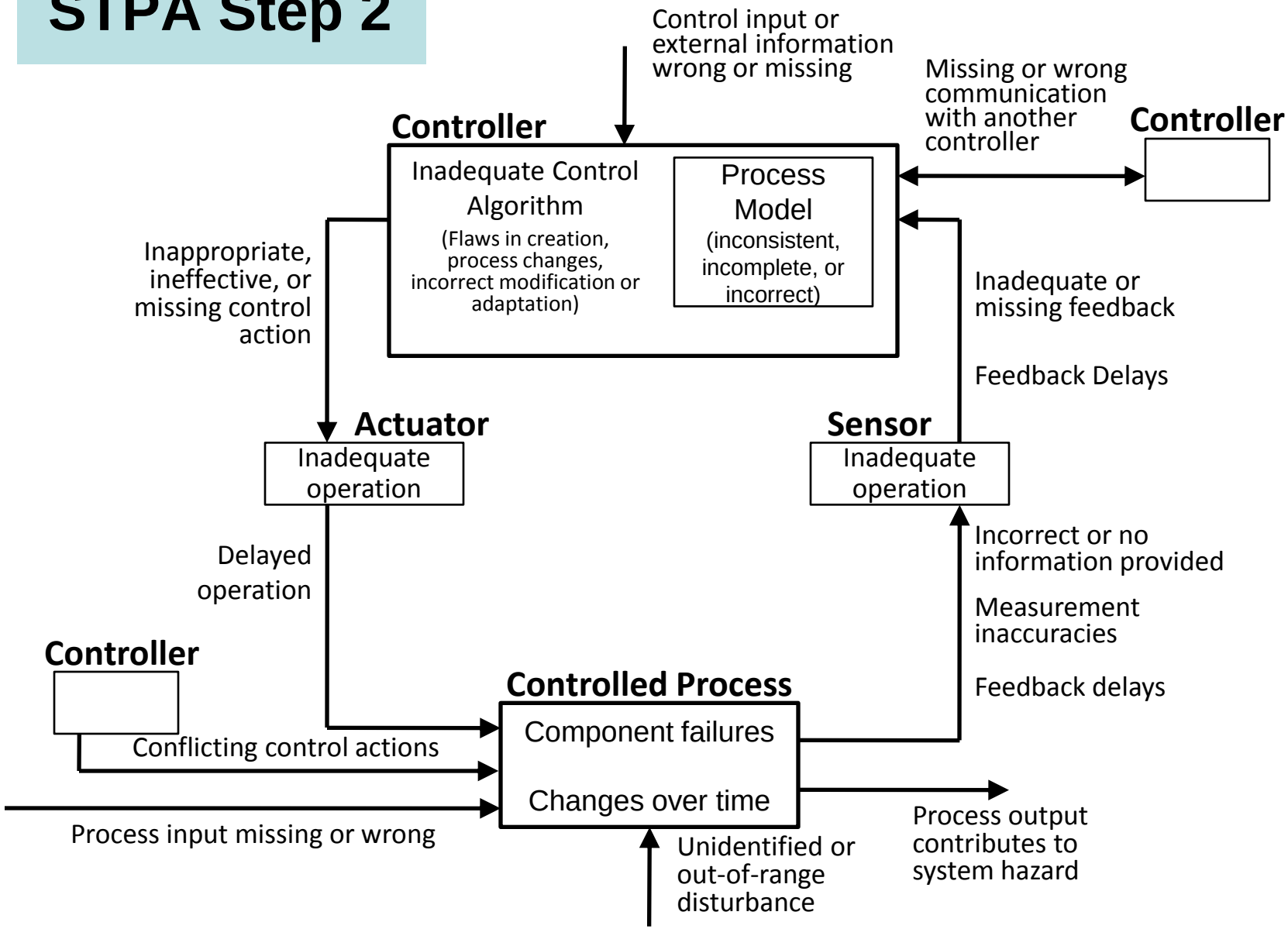
Control Action	Not providing causes hazard	Providing causes hazard	Too early/too late, wrong order	Stopped too soon/ applied too long
Open water	Not opened when catalyst open		Open water more than X seconds after open catalyst	Stop before fully opened
Close water		Close while catalyst open	Close water before catalyst closes	
Open catalyst		Open when water valve not open	Open catalyst more than X seconds before open water	
Close catalyst	Do not close when water closed		Close catalyst more than X seconds after close water	Stop before fully closed

STPA generates the following high-level requirements on the batch reactor:

- Water valve must always be fully open before catalyst valve is opened.
  - Water valve must never be opened (complete opening) more than X seconds after catalyst valve opens
- Catalyst valve must always be fully closed before water valve is closed.
  - Catalyst valve must never be closed more than X seconds after water valve has fully closed.

Next step is to identify scenarios leading to these unsafe control actions and eliminate or mitigate them

STPA Step 2


Control input or external information wrong or missing
Controller
Inadequate Control Algorithm
(Flaws in creation, process changes, incorrect modification or adaptation)
Process Model
(inconsistent, incomplete, or incorrect)
Missing or wrong communication with another controller
Controller
Inappropriate, ineffective, or missing control action
Inadequate or missing feedback
Feedback Delays
Actuator
Inadequate operation
Sensor
Inadequate operation
Delayed operation
Incorrect or no information provided
Measurement inaccuracies
Feedback delays
Controller
Controlled Process
Component failures
Changes over time
Conflicting control actions
Process input missing or wrong
Unidentified or out-of-range disturbance
Process output contributes to system hazard

Some scenarios for "Software issues open catalyst command when water valve is closed"

1. Nobody tells software engineers this is a safety constraint/requirement
2. Software engineers told but erroneously think water valve is open so issue "Open catalyst" Why?

   a. Previously issued an Open Water Valve command but valve did not open (jammed, failed, etc.) Assumed that command had been executed. Why?

      i. No feedback about affect of previous command (Control: put feedback in design)

      ii. Feedback not received. [could go on to determine why here if want] (Control: Assume not executed)

      iii. Feedback delayed (could go on to determine why if want) (Control: wait predetermined time and then assume not opened)

      iv. Incorrect feedback received. Why? (maybe assumed that if reached valve, it would open [design error] (Control: add flow meter to detect water flow through pipe)

etc.

Generates Potential New Requirements:

- Include feedback for Open Valve and Close Valve commands.
- Software shall check for feedback after issuing an Open/Close command. If not received in a specified time period, then assume valve not opened or closed and …
- A flow meter shall be used as feedback to controller to determine that water is actually flowing through pipe before issuing an Open Catalyst command.
- …

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Common Traps in Understanding Accident Causes

- Root cause seduction and oversimplification
- Narrow views of human error
- Hindsight bias
- Focus finding someone or something to blame

Root Cause Seduction

- Assuming there is a root cause gives us an illusion of control.
  - Usually focus on operator error or technical failures
  - Ignore systemic and management factors
  - Leads to a sophisticated “whack a mole” game
    - Fix symptoms but not process that led to those symptoms
    - In continual fire-fighting mode
    - Having the same accident over and over

Oversimplification of Causes

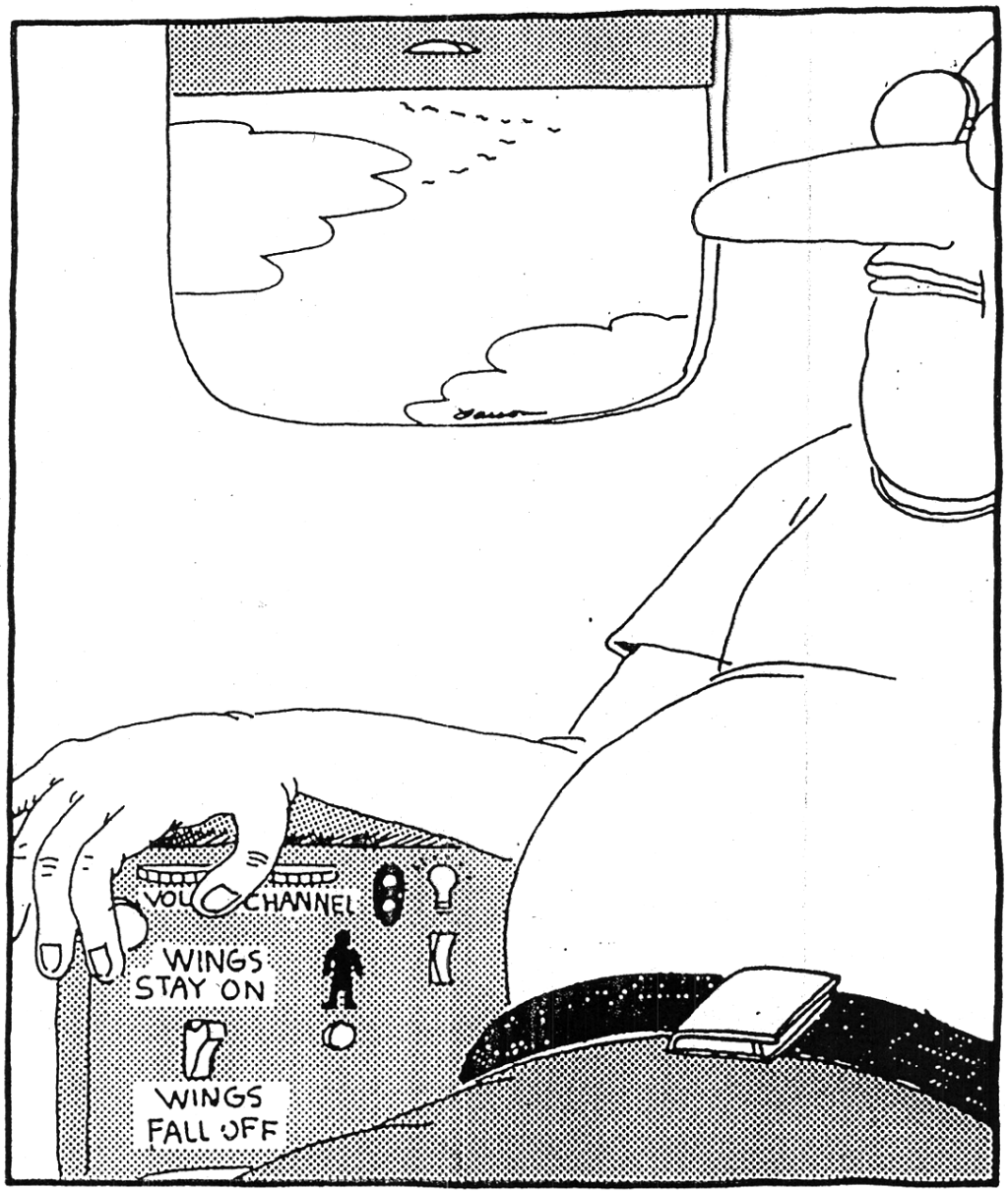
- Almost always there is:
  - Operator “error”
  - Flawed management decision making
  - Flaws in the physical design of equipment
  - Safety culture problems
  - Regulatory deficiencies
  - Etc.
- Need to determine why safety control structure was ineffective in preventing the loss.

Blame is the Enemy of Safety

- Two possible goals for an accident investigation:
  - Find who to blame
  - Understand why occurred so can prevent in future
- Blame is a legal or moral concept, not an engineering one
- Focus on blame can:
  - Prevent openness during investigation
  - Lead to finger pointing and cover ups
  - Lead to people not reporting errors and problems before accidents

Human Error: Traditional View

- Operator error is cause of most incidents and accidents
- So do something about human involved (fire them, retrain, admonish)
- Or do something about humans in general
  - Marginalize them by putting in more automation
  - Rigidify their work by creating more rules and procedures

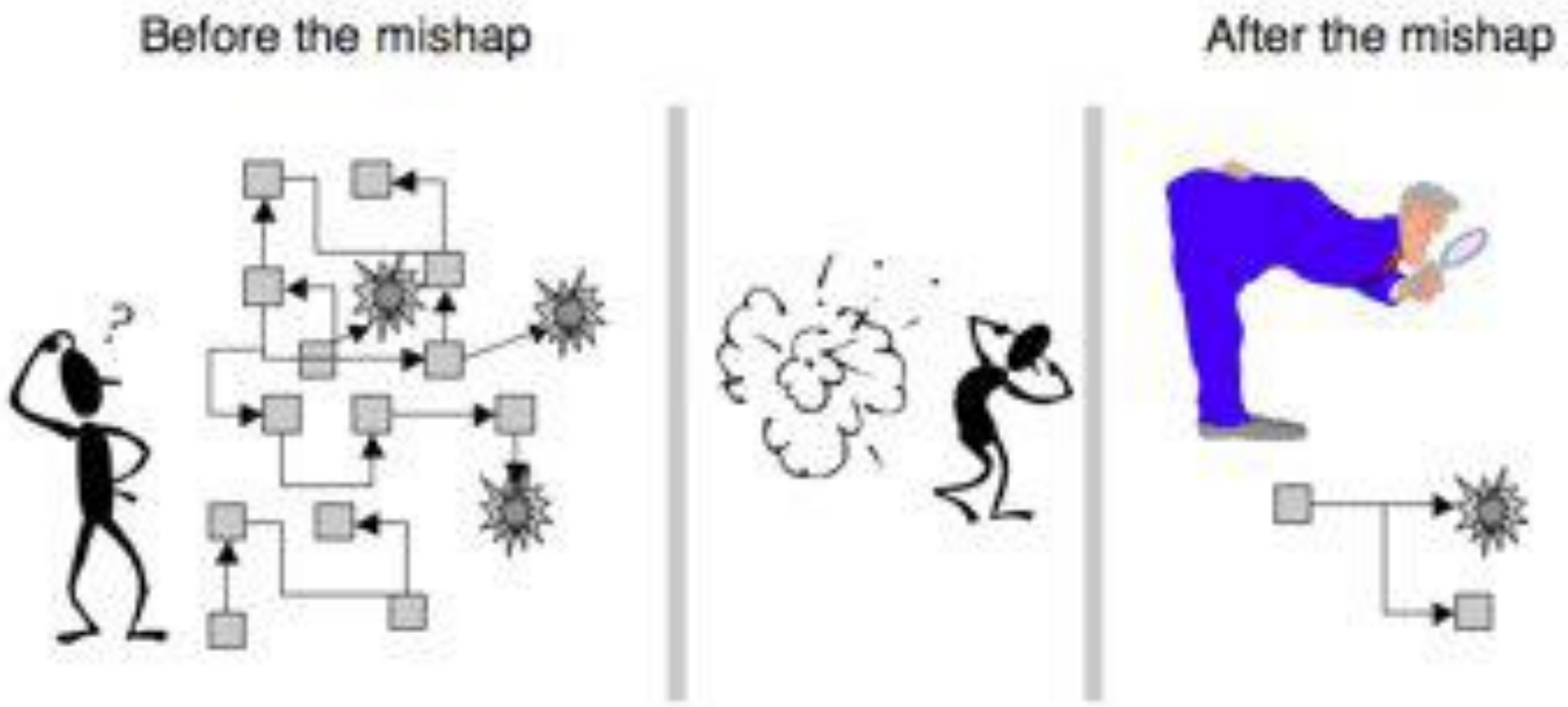


Fumbling for his recline button Ted unwittingly instigates a disaster

Human Error: Systems View
(Sydney Dekker, Jens Rasmussen, Leveson)

- Human error is a symptom, not a cause
- All behavior affected by context (system) in which occurs
- To do something about error, must look at system in which people work:
  - Design of equipment
  - Usefulness of procedures
  - Existence of goal conflicts and production pressures
- Human error is a sign that a system needs to be redesigned

Hindsight Bias



Sidney Dekker, 2009

Overcoming Hindsight Bias

- Assume nobody comes to work to do a bad job.
  - Assume were doing reasonable things given the complexities, dilemmas, tradeoffs, and uncertainty surrounding them.
  - Simply finding and highlighting people's mistakes explains nothing.
  - Saying what did not do or what should have done does not explain why they did what they did.
- Investigation reports should explain
  - Why it made sense for people to do what they did rather than judging them for what they allegedly did wrong and
  - What changes will reduce likelihood of happening again

CAST (Causal Analysis using System Theory)

- Identify system hazard violated and the system safety design constraints
- Construct the safety control structure as it was designed to work
  - Component responsibilities (requirements)
  - Control actions and feedback loops
- For each component, determine if it fulfilled its responsibilities or provided inadequate control.
  - If inadequate control, why? (including changes over time)
  - Context
  - Process Model Flaws

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Strategy vs. Tactics

- Primarily focus on tactics
  - Cyber security often framed as battle between adversaries and defenders (tactics)
  - Requires correctly identifying attackers motives, capabilities, targeting
- Can reframe problem in terms of strategy
  - Identify and control system vulnerabilities (vs. reacting to potential threats)
  - Top-down vs. bottom-up tactics approach
  - Tactics tackled later

Integrated Approach to Safety and Security:

- Safety: prevent losses due to unintentional actions by benevolent actors
- Security: prevent losses due to intentional actions by malevolent actors
- Key difference is intent
- Common goal: loss prevention
  - Ensure that critical functions and services provided by networks and services are maintained
  - 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

Top-Down Approach

- Starts with identifying losses
- Identify vulnerabilities and system safety/security constraints
- Build functional control model
  - Controlling constraints whether safety or security
  - Includes physical, social, logical and information, operations, and management aspects
- Identify unsafe/unsecure control actions and causes for them
  - May have to add new causes, but rest of process is the same

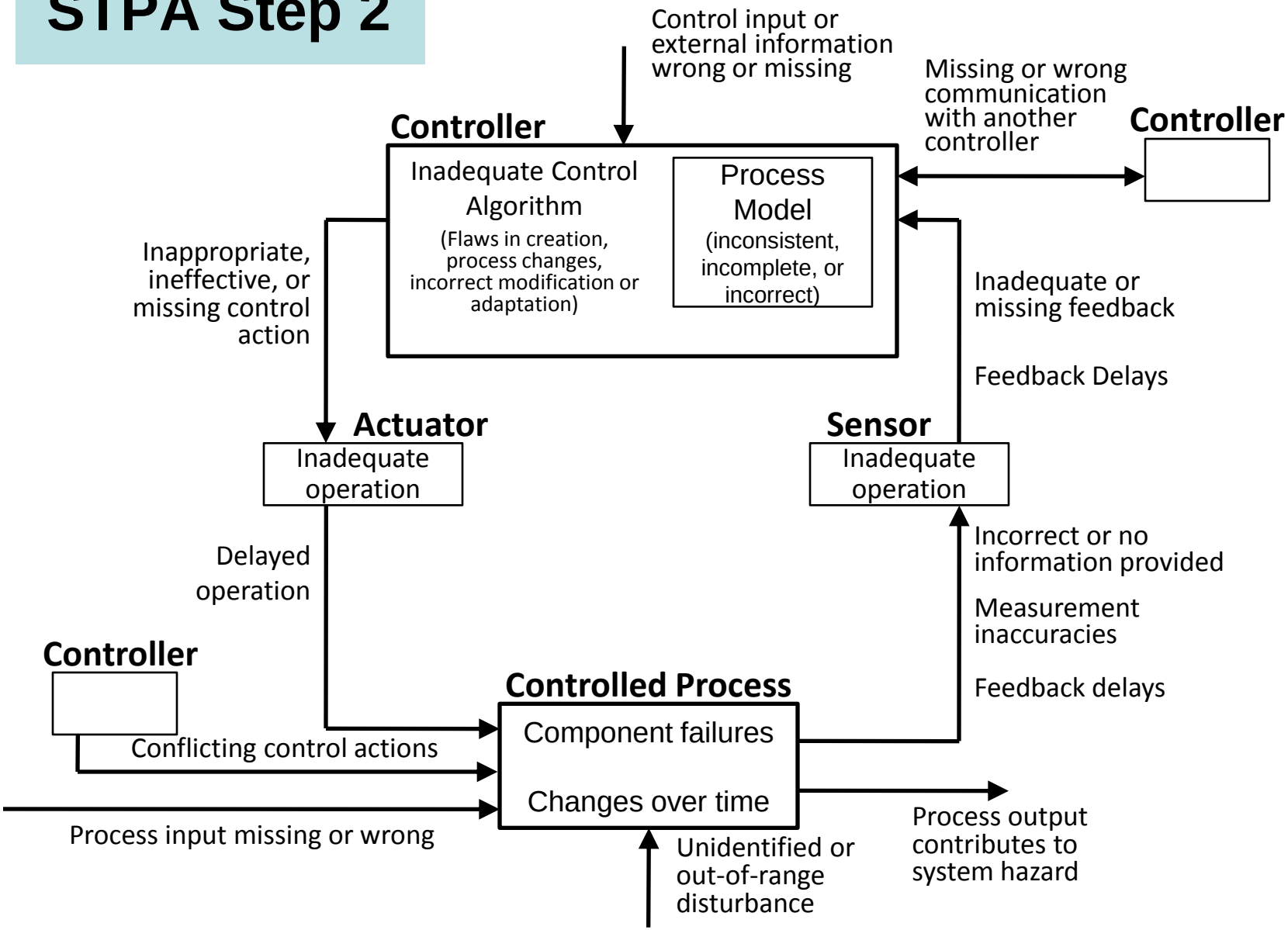
Example: Stuxnet

- Loss: Damage to reactor (in this case centrifuges)
- Hazard/Vulnerability: Centrifuges are damaged by spinning too fast
- Constraint: Centrifuges must never spin above maximum speed
- Hazardous control action: Issuing *increase speed* command when already spinning at maximum speed
- One potential cause:
  - *Incorrect process model*: thinks spinning at less than maximum speed
  - Could be inadvertent or advertent

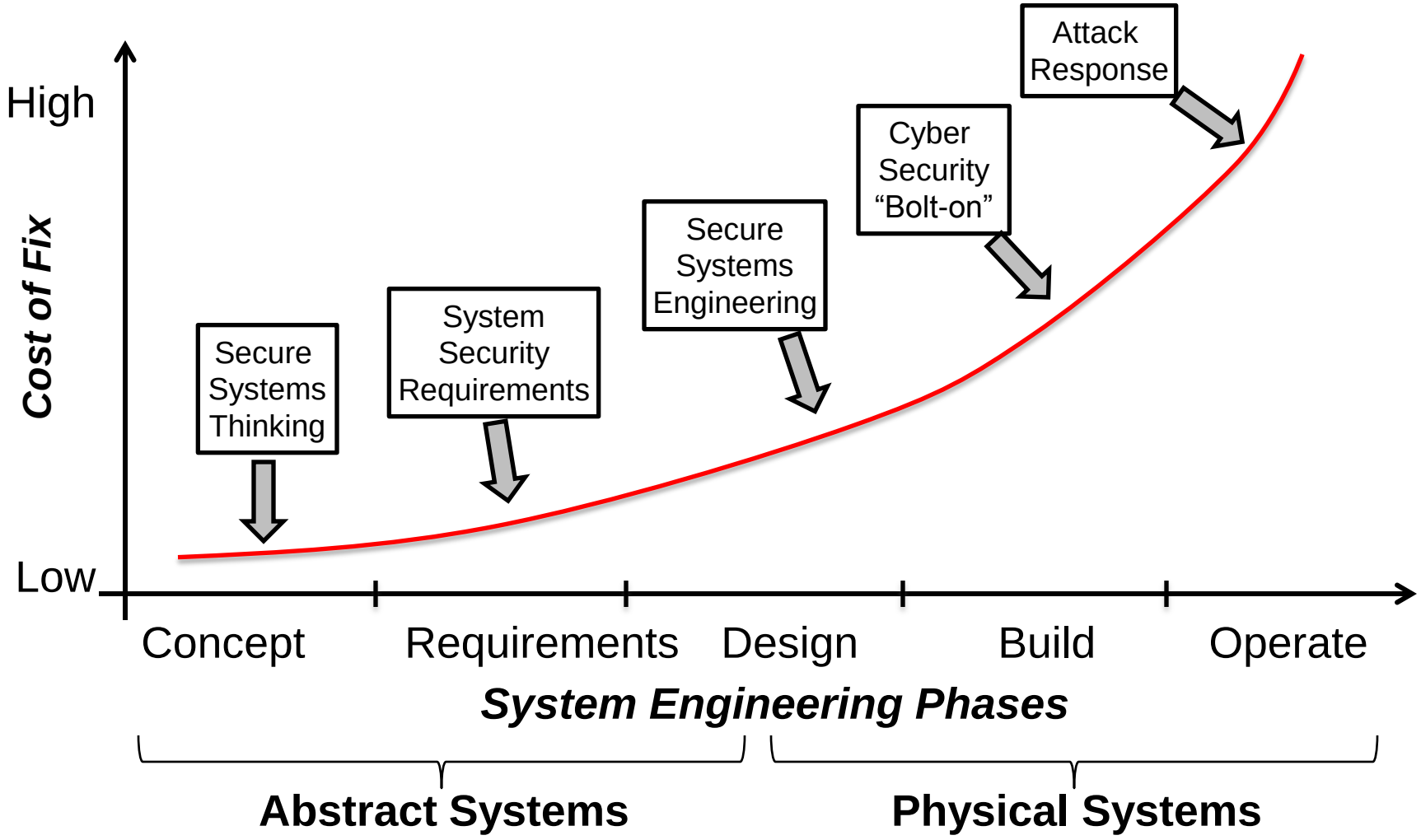
Analysis

- Step 1 is same as for safety
- Step 2 may require adding new causes

STPA Step 2


Control input or external information wrong or missing
Missing or wrong communication with another controller
Controller
Controller
Inadequate Control Algorithm
(Flaws in creation, process changes, incorrect modification or adaptation)
Process Model
(inconsistent, incomplete, or incorrect)
Inappropriate, ineffective, or missing control action
Inadequate or missing feedback
Feedback Delays
Actuator
Sensor
Inadequate operation
Inadequate operation
Delayed operation
Incorrect or no information provided
Measurement inaccuracies
Feedback delays
Controller
Controlled Process
Component failures
Changes over time
Conflicting control actions
Process input missing or wrong
Unidentified or out-of-range disturbance
Process output contributes to system hazard

STPA-Sec Allows us to Address Security "Left of Design" [Bill Young]



Build security into system like safety

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Is it Practical?

- STPA has been or is being used in a large variety of industries
  - Spacecraft
  - Aircraft
  - Air Traffic Control
  - UAVs (RPAs)
  - Defense
  - Automobiles (GM, Ford, Nissan?)
  - Medical Devices and Hospital Safety
  - Chemical plants
  - Oil and Gas
  - Nuclear and Electrical Power
  - $C0_2$ Capture, Transport, and Storage
  - Etc.

Is it Practical? (2)

Social and Managerial

- Analysis of the management structure of the space shuttle program (post-Columbia)
- Risk management in the development of NASA's new manned space program (Constellation)
- NASA Mission control ─ re-planning and changing mission control procedures safely
- Food safety
- Safety in pharmaceutical drug development
- Risk analysis of outpatient GI surgery at Beth Israel Deaconess Hospital
- Analysis and prevention of corporate fraud

Does it Work?

- Most of these systems are very complex (e.g., the U.S. Missile Defense System)
- In all cases where a comparison was made:
  - STPA found the same hazard causes as the old methods
  - Plus it found more causes than traditional methods
  - In some evaluations, found accidents that had occurred that other methods missed
  - Cost was orders of magnitude less than the traditional hazard analysis methods

Example STPA Evaluations on Real Systems

- Non-advocate safety assessment of U.S. Ballistic Missile Defense System
  - 2 people for 3 months
  - Deployment and testing held up for 6 months because so many scenarios identified for inadvertent launch.
  - In many of these scenarios:
    - All components were operating exactly as intended but complexity of component interactions led to unanticipated system behavior
    - Examples: missing case in software requirements, timing problem in sending and receiving messages, etc.
  - STPA also identified component failures that could cause inadequate control (most analysis techniques consider only these failure events)

Example Comparisons

- JAXA HTV
  - Found everything found in fault tree analysis and more (mostly related to system design and software)
- Nuclear Power Plants
  - Experimental comparison performed by EPRI and experts on each technique
  - Results not available yet but informally STPA was only one that found a real accident scenario that had occurred (and none of analysts knew about)

Example Comparisons

- Blood Gas Analyzer (Vincent Balgos)
  - 75 scenarios found by FMEA
  - 175 identified by STPA
  - Took much less time and resources (mostly human)
  - Only STPA found scenario that had led to a Class 1 recall by FDA (actually found nine scenarios leading to it)
- Lots more comparisons on real systems
- Biggest surprise (to me) was required much less resources

Real World Evaluation of STPA-Sec to Date

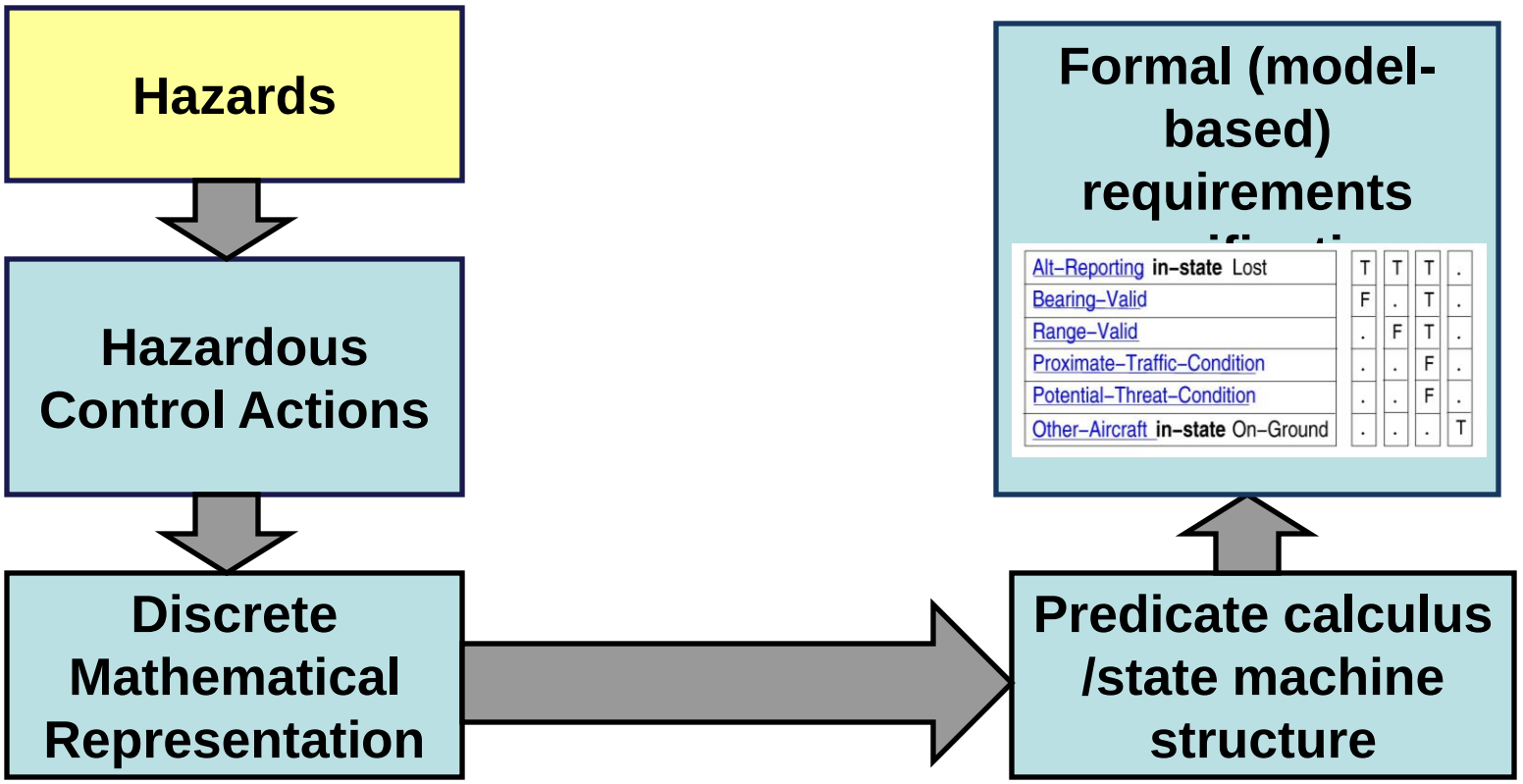
- Demonstrated ability to identify previously unknown vulnerabilities in a global DoD mission
  - Created model based on actual planning documents
- Demonstrated ability to identify high-level vulnerabilities in early system concept documents
  - Required security constraints missing
- Demonstrated ability to improve ability of network defenders to assure a real-world space surveillance mission
  - Real mission, Real mission owner, Real network
  - Defenders able to more precisely identify what to defend & why (e.g. set of servers → integrity of a single file)
  - Defenders able to provide traceability allowing non-cyber experts to better understand mission impact of cyber disruptions

Outline

- Accident Causation in Complex Systems: STAMP
- New Analysis Methods
  - Hazard Analysis
  - Accident Analysis
  - Security Analysis
- Does it Work? Evaluations
- Extensions, Tools, Research Topics

Automating STPA (Step 1): John Thomas)

- Requirements can be derived automatically (with some user guidance) using mathematical foundation
- Allows automated completeness/consistency checking



Others (that we are doing)

- Automating Step 2
- Leading Indicators
- Sophisticated Human Factors Analysis
- Safety-Guided Development (Design)
  - Concept of Operation
  - Integrated Modular Avionics
- Feature Interactions (Automobiles)
- Safety Management Systems

Others (that we are doing)

- Changes in Complex Systems
  - Air Traffic Control (NextGen)
  - UAVs in commercial airspace
- Workplace (Occupational) Safety
- Some Current Applications
  - Medicine
  - Flight Test (Air Force)
  - Security in aviation
  - Defense systems