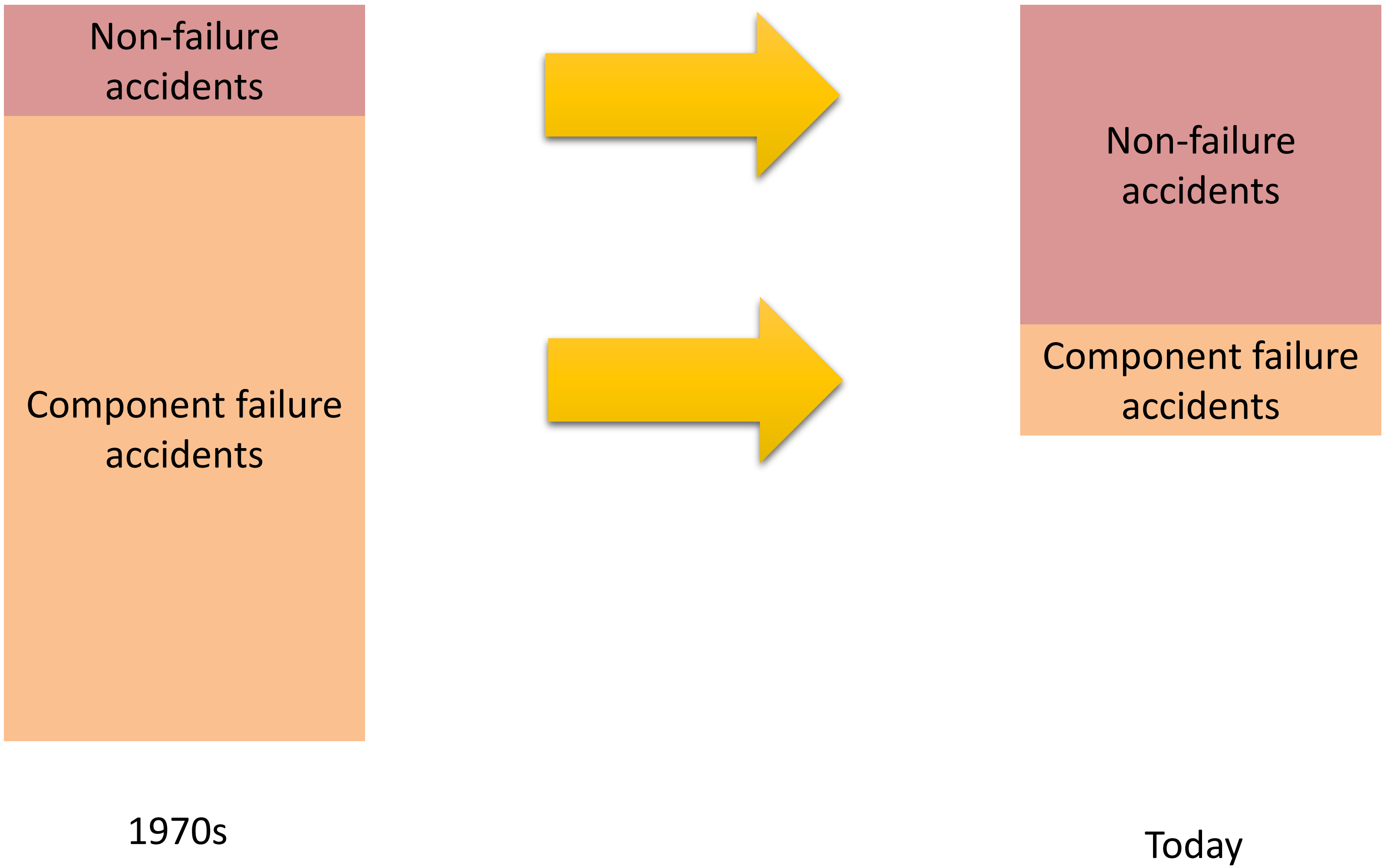
A new approach to software safety using STPA

Dr. John P. Thomas

MIT

Accident causes are changing



A320 Thrust Reversers

- Used to reverse engine thrust, help aircraft stop on ground
- Software prevents thrust reverser deployment in air

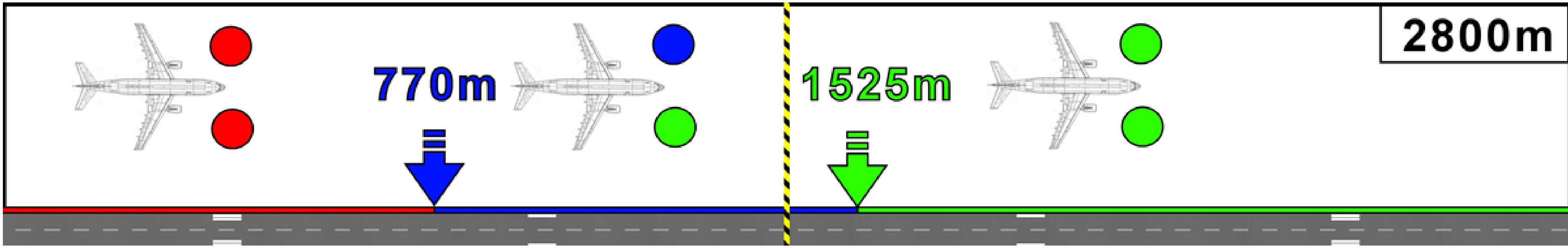
- Thrust reverser would not deploy on landing
- Software prevented manual pilot override
- 9 seconds after touchdown, software deployed thrust reversers
- Plane overruns, crashes, catches fire

Warsaw Crash

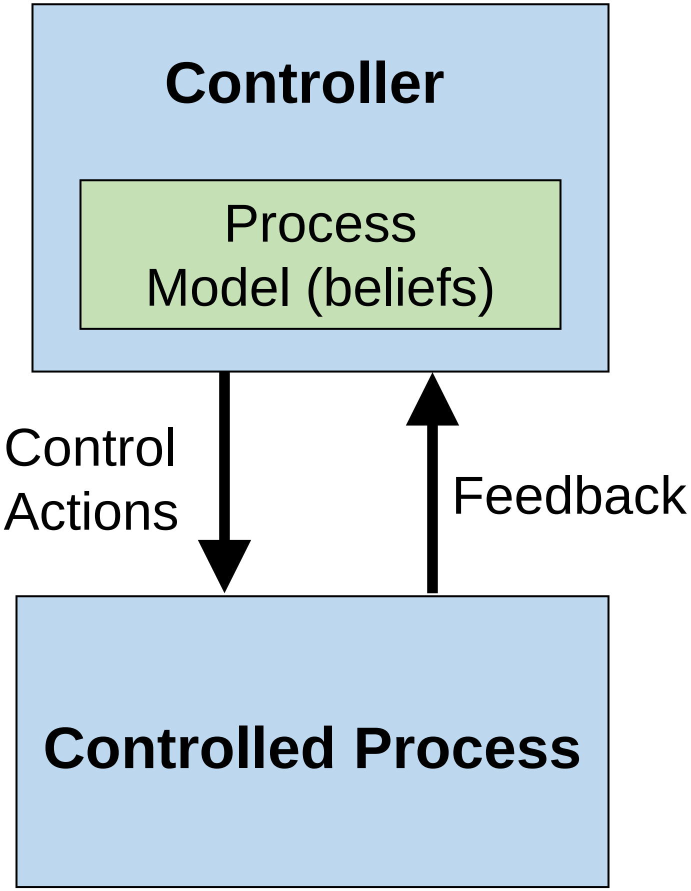
- Software algorithm to ensure aircraft has landed:
  - Must be 6.3 tons on each main landing gear strut
  - Wheel must be turning at least 72 knots
- Off-nominal landing conditions at Warsaw
  - Crosswind landing (one side first)
  - Wet runway: wheels hydroplane

Lufthansa 2904, Airbus A320

SW operated exactly as designed, no failure!



A different view



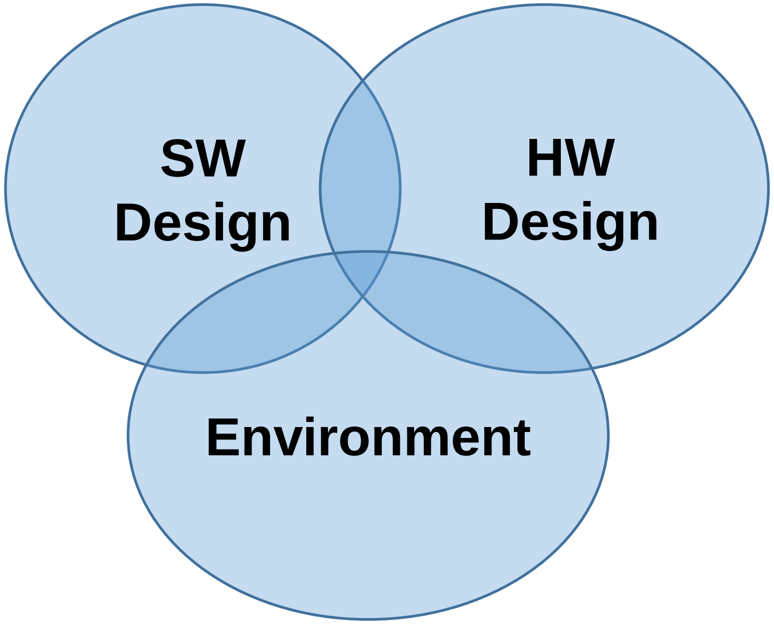
- Another way to think about accidents
- Forms foundation for STPA

How was this overlooked?

Individual parts carefully examined:

- SW Requirements
  - React within X ms
  - Detect, tolerate sensor failure
  - Respond only when multiple sensors agree
- HW Requirements:
  - Redundant WoW sensors
  - Redundant wheel speed sensors
  - Redundant computers
- HW Testing
  - Inject single WoW failure
  - Inject single wheel speed sensor failure
  - Inject single computer failure
- SW Testing
  - Verify response within X ms of inputs
  - Verify no deployment from sensor failure
  - Verify no deployment until multiple sensors agree
- *Engineering Safety Analysis:*
  use failure-based methods
- Etc.

Hard to find problem by looking at any one part



Quote

- “The hardest single part of building a software system is deciding precisely what to build.”
  -- Fred Brooks, *The Mythical Man-Month*

Software in Aviation

- Bombardier Learjet 60 Accident
  - September 19, 2008
  - Columbia Metropolitan Airport, South Carolina
- Aircraft was destroyed during rejected takeoff
- Reverse thrusters would not engage

Bombardier Learjet 60 Accident

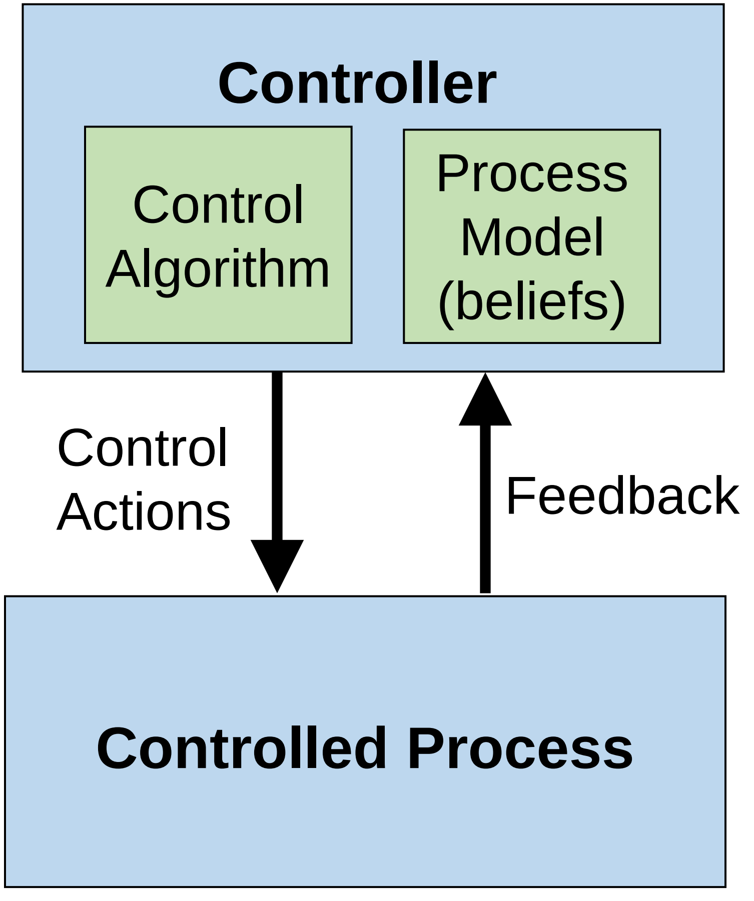
- Tires disintegrated on takeoff, pilots tried to abort
- Computer ignored pilot commands for reverse thrusters
  - The tire explosion damaged landing gear sensors
  - Computer believed aircraft in flight
  - Computer <u>increased thrust</u>
- Aircraft was destroyed

Bombardier Learjet 60 Accident

- Tires disintegrated on takeoff, pilots tried to abort
- Computer ignored pilot commands for reverse thrusters
  - The tire explosion damaged landing gear sensors
  - Computer believed aircraft in flight
  - Computer <u>increased thrust</u>
- Aircraft was destroyed

The computer operated exactly as designed!

A different view



- Another way to think about accidents
- Forms foundation for STPA

Boeing 787 Lithium Battery Fires

- Fire computer monitors for smoke in the battery bay, will activate fans and valves for venting
- Power management system detects rapid battery discharge. Begins shutting down electronics...

Boeing 787 Lithium Battery Fires

- Fire computer monitors for smoke in the battery bay, will activate fans and valves for venting
- Power management system detects rapid battery discharge. Begins shutting down electronics including ventilation computer.
- Smoke vented to cabin

Boeing 787 Lithium Battery Fires

- Fire computer monitors for smoke in the battery bay, will activate fans and valves for venting
- Power management system detects rapid battery discharge. Begins shutting down electronics including ventilation computer.
- Smoke vented to cabin

Operated as designed
Requirements met

Operated as designed
Requirements met

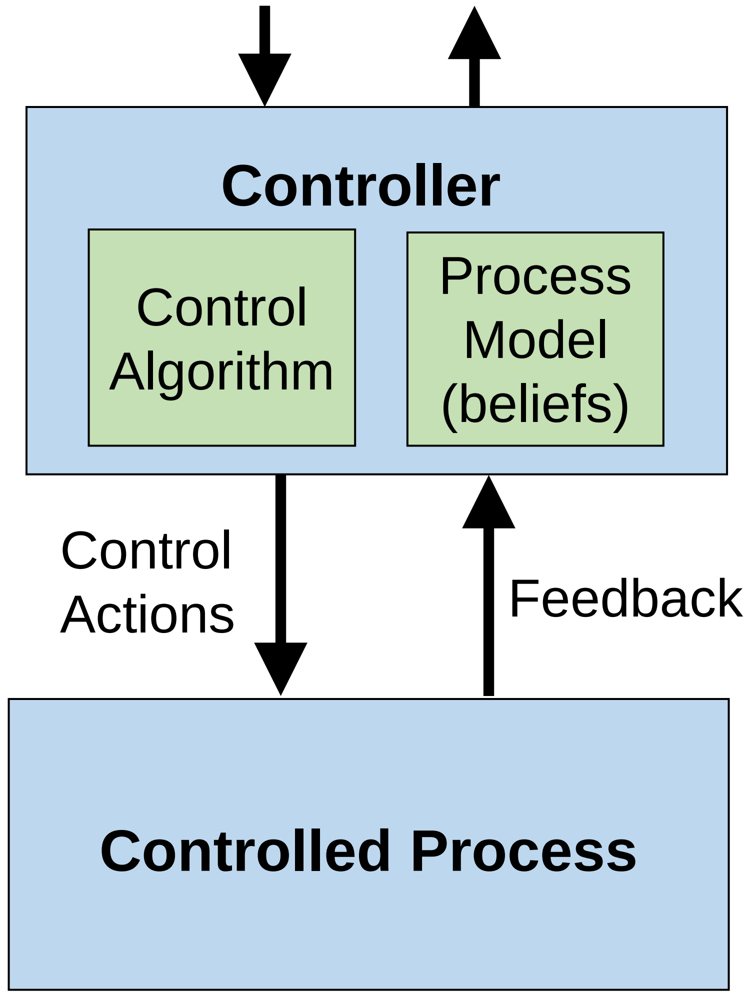
This flaw was overlooked by every software analysis, every test, safety assessment, every design review, every certification effort, etc. !!

NTSB Conclusion

- "The NTSB determines that the probable cause of this incident was an internal short circuit within a cell of the auxiliary power unit (APU) lithium-ion battery, which led to thermal runaway that cascaded to adjacent cells, resulting in the release of smoke and fire.

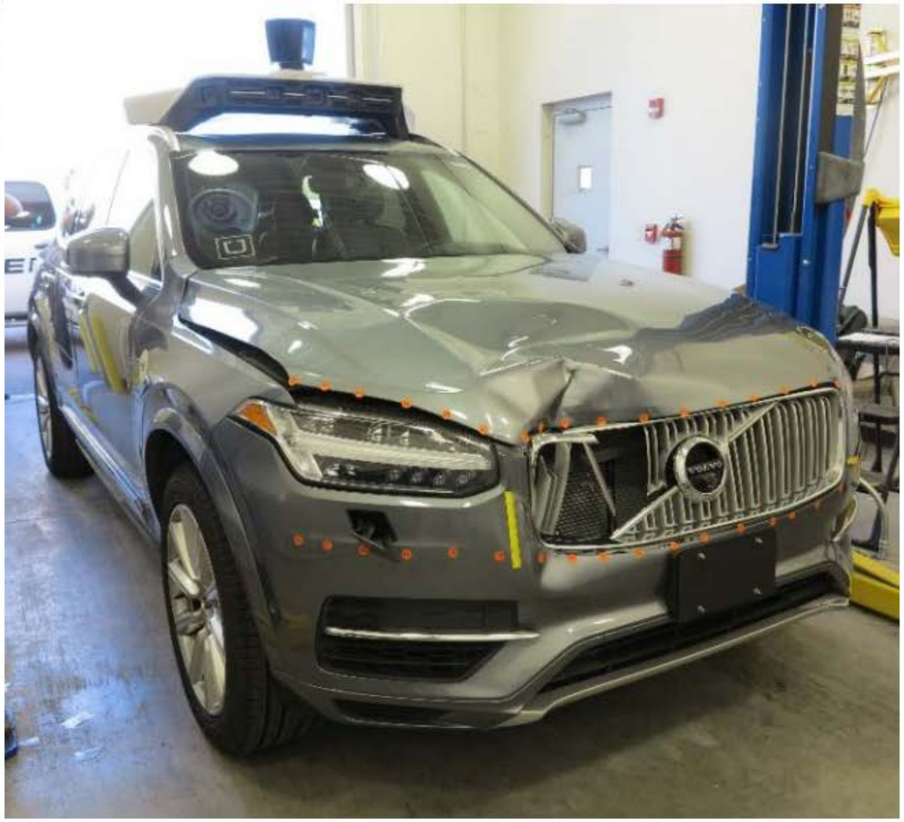
  The incident resulted from **Boeing's failure to incorporate design requirements to mitigate the most severe effects** of an internal short circuit within an APU battery cell **and the Federal Aviation Administration's failure to identify this design deficiency** during the type design certification process."

A different view



- Another way to think about accidents
- Forms foundation for STPA

Uber Crash



Technical factors

- Why didn't autonomy stop?
- Cameras: low light
- Lidar and Radar: should have worked
- Designed to detect pedestrians even without crosswalk
- Uber system automatically disabled Volvo features (City Safety, etc.)
- Automated commands for deceleration greater than $6.5 m/s^2$ are not executed by design (stability)
- Obstacle was detected, filter added
- Uber target: 13 miles/intervention

The Mercury News

Business › Technology

Uber fatal crash: Self-driving software reportedly set to ignore objects on road

By **LEVI SUMAGAYSAY** | lsumagaysay@bayareanewsgroup.com | Bay Area News Group

PUBLISHED: May 8, 2018 at 7:04 am | UPDATED: May 8, 2018 at 4:12 pm

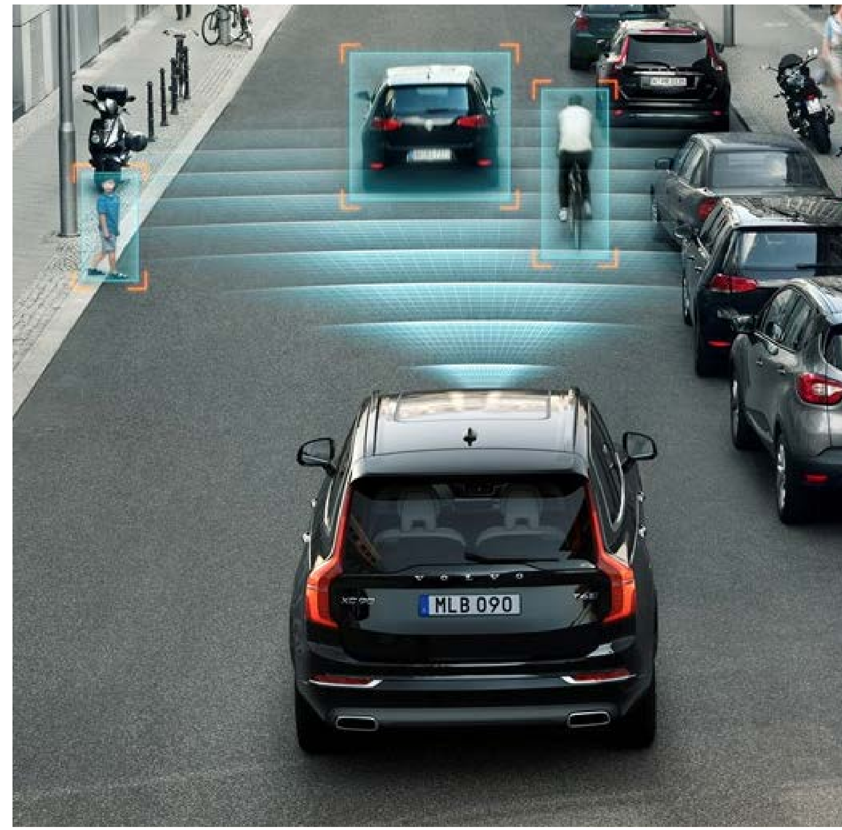
All components operated exactly as designed (intended)

All component requirements met! No failures!

Volvo City Safety System

From Volvo website:

- City Safety is a support system designed to help the driver avoid low speed collisions when driving in slow-moving, stop-and-go traffic.
- City Safety triggers brief, forceful braking if a low-speed collision is imminent.



Volvo City Safety preventing an accident

Accident with City Safety

Volvo response

- "The Volvo XC60 comes with City Safety as a standard feature
- "however this does not include the Pedestrian detection functionality … this is sold as a separate package."
- Optional pedestrian detection functionality costs $3,000

Kashmir Hill, "Volvo says horrible 'self-parking car accident' happened because driver didn't have 'pedestrian detection'", Splinter news, May 2015
https://splinternews.com/volvo-says-horrible-self-parking-car-accident-happened-1793847943

Volvo response

- “The Volvo XC60 comes with City Safety as a standard feature …
- “however this does not include the Pedestrian detection functionality … this is sold as a separate package."
- Optional pedestrian detection functionality costs $3,000
- Even with pedestrian detection, it mostly likely would not have worked because the driver accelerated

Kashmir Hill, “Volvo says horrible 'self-parking car accident' happened because driver didn't have 'pedestrian detection‘”, Splinter news, May 2015
https://splinternews.com/volvo-says-horrible-self-parking-car-accident-happened-1793847943

Volvo City Safety System

From Volvo:

- City Safety is not active if your vehicle's speed is **below approximately 2 mph**. This means that City Safety will not react if your vehicle approaches another vehicle at very low speed, for example, **when parking**.
- The function is active at speeds up to approximately **30 mph**
- However, the system will not intervene in situations where the **driver actively steers** the vehicle or **applies the brakes**, even if a collision cannot be avoided
- City Safety activates in situations where the driver has not applied the brakes in time, which means that the system cannot help the driver in all situations.
- City Safety does not function in all driving situations or in all **traffic**, **weather** or road conditions.
- City Safety only reacts to vehicles traveling in the **same direction** as your vehicle
- City Safety … does not react to **small vehicles** or motorcycles
- City Safety is not activated when your vehicle is **backing up**.
- This system can help prevent a collision if the difference in speed between your vehicle and the vehicle ahead is **less than 9 mph**. If the difference in speed is greater, a collision cannot be avoided but the speed at which the collision occurs can be reduced. **The driver must apply the vehicle's brakes for full braking effect.**

These requirements were met.
All components operated as intended!

City Safety Introduction from https://www.volvocars.com/en-ca/support/cars?pc=y283&my=2015&sw=14w46&tab=ownersmanualonline&category=f13d9e9caab79e66c0a801e80081bf31-om-en-ca-y283-2015-14w46&article=14dd7a22aa482bc2c0a801e800c7463b-om-en-ca-y283-2015-14w46

Barrier: requirements

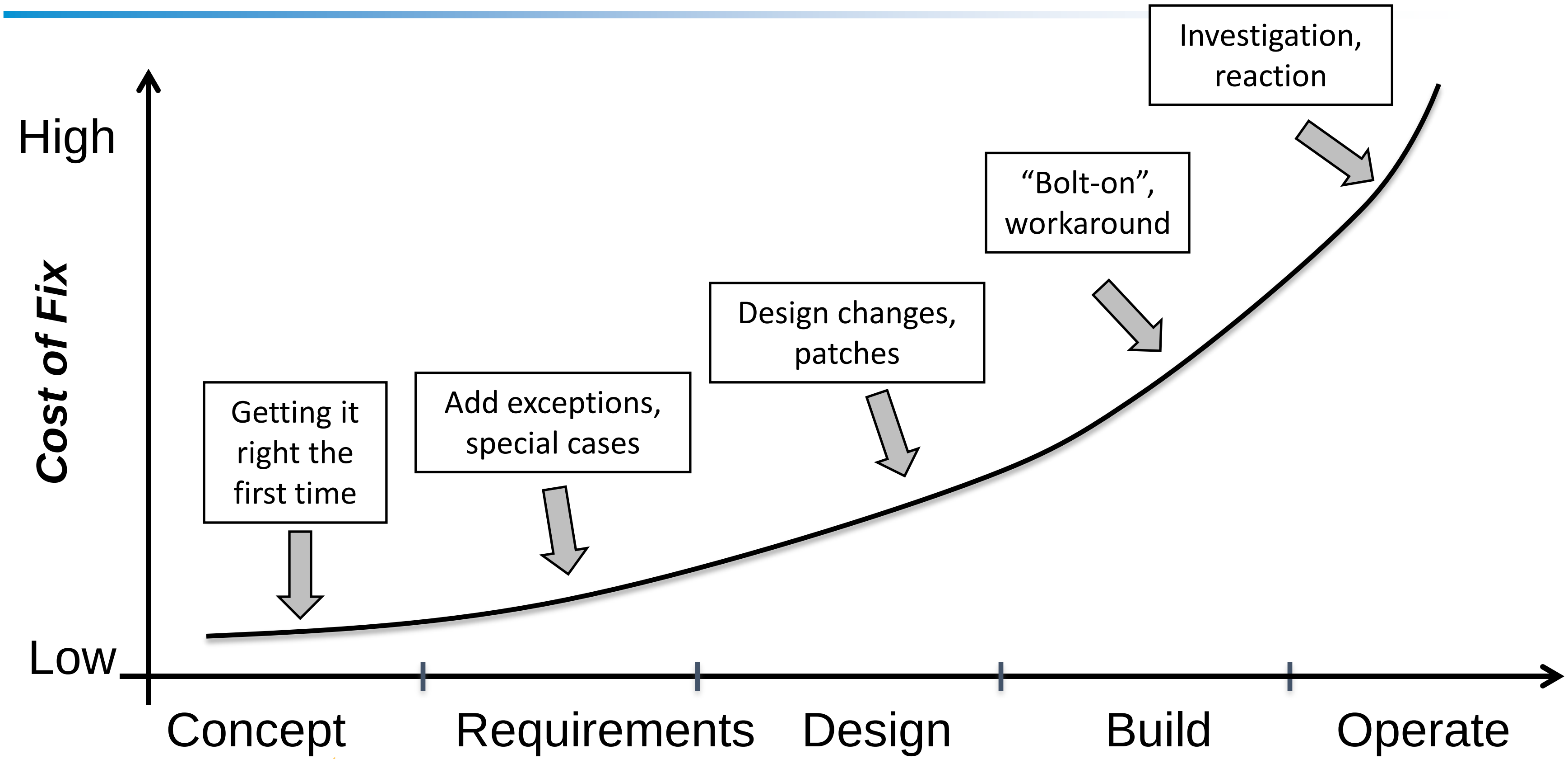
- Most software-related accidents have been traced to flaws in the **requirements**

  (Leveson, 2004) (Endres et al., 2003)(Lutz et al., 1993)

- *"As is well known to software engineers, by far the largest class of problems arises from errors made in the eliciting, recording, and analysis of* ***requirements****"*

  (Jackson et al., 2007)

Addressing SW quality issues



Need to address issues early, don't wait

Early decisions can have biggest impact

Adapted from Young, 2014

What about human interactions?

Monostable shifter design

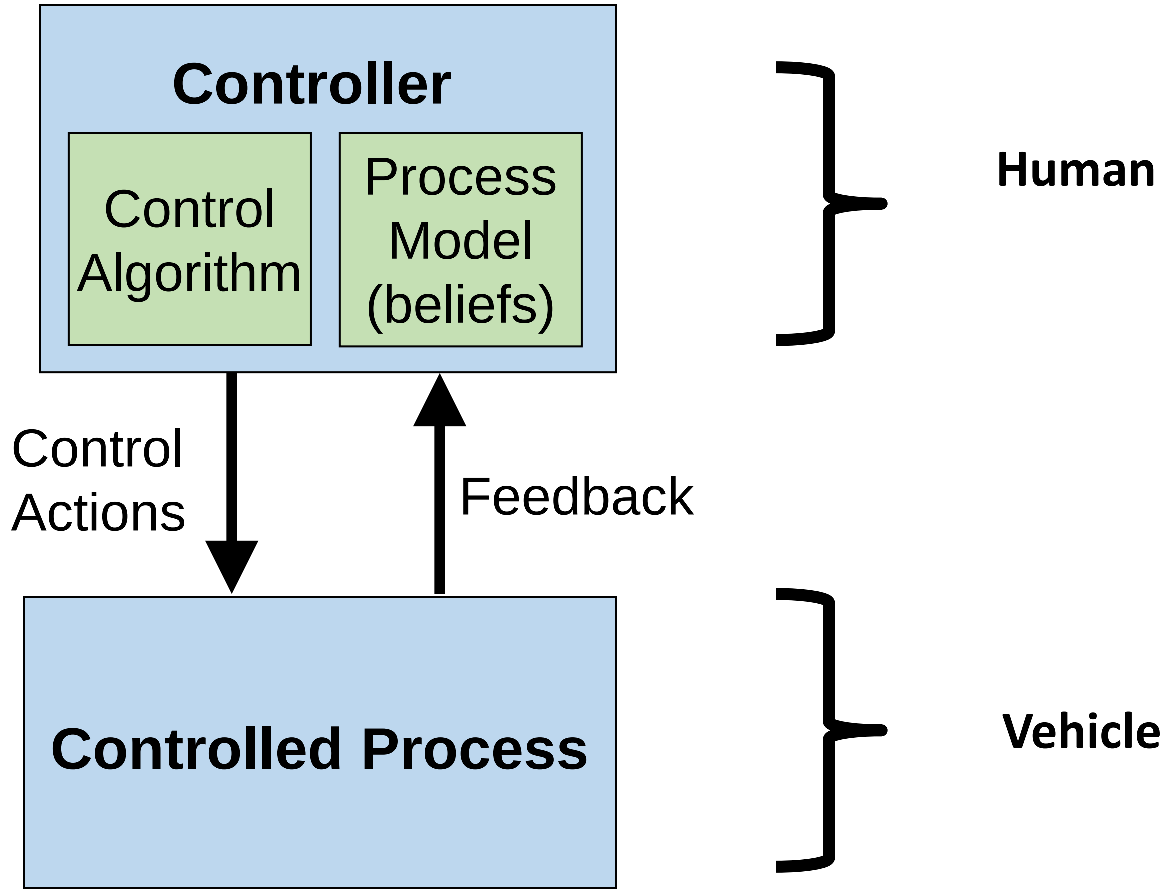
NHTSA: "operation of the Monostable shifter is not intuitive and provides poor tactile and visual feedback to the driver, increasing the potential for unintended gear selection."

Monostable shifter design

Audi A8—Same design, but new SW requirement:

R-1: Computer shall automatically activate the electronic park brake when driver exits

Basic Control Loop



- Another way to think about accidents
- Forms foundation for STAMP/STPA/CAST

Control Structure Modeling

Enabling abstraction

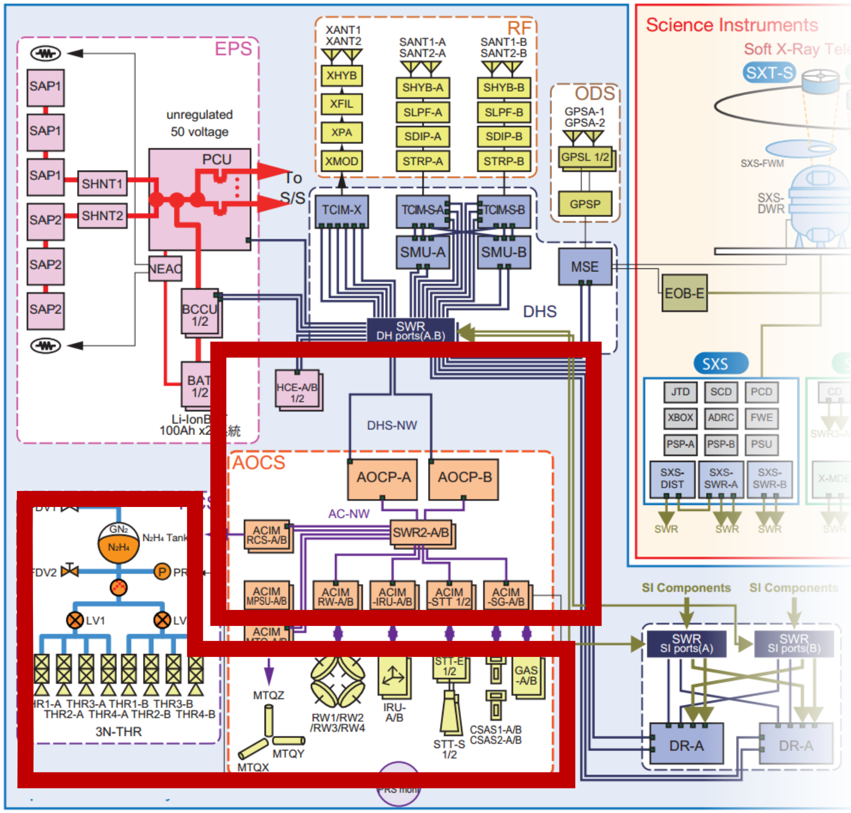
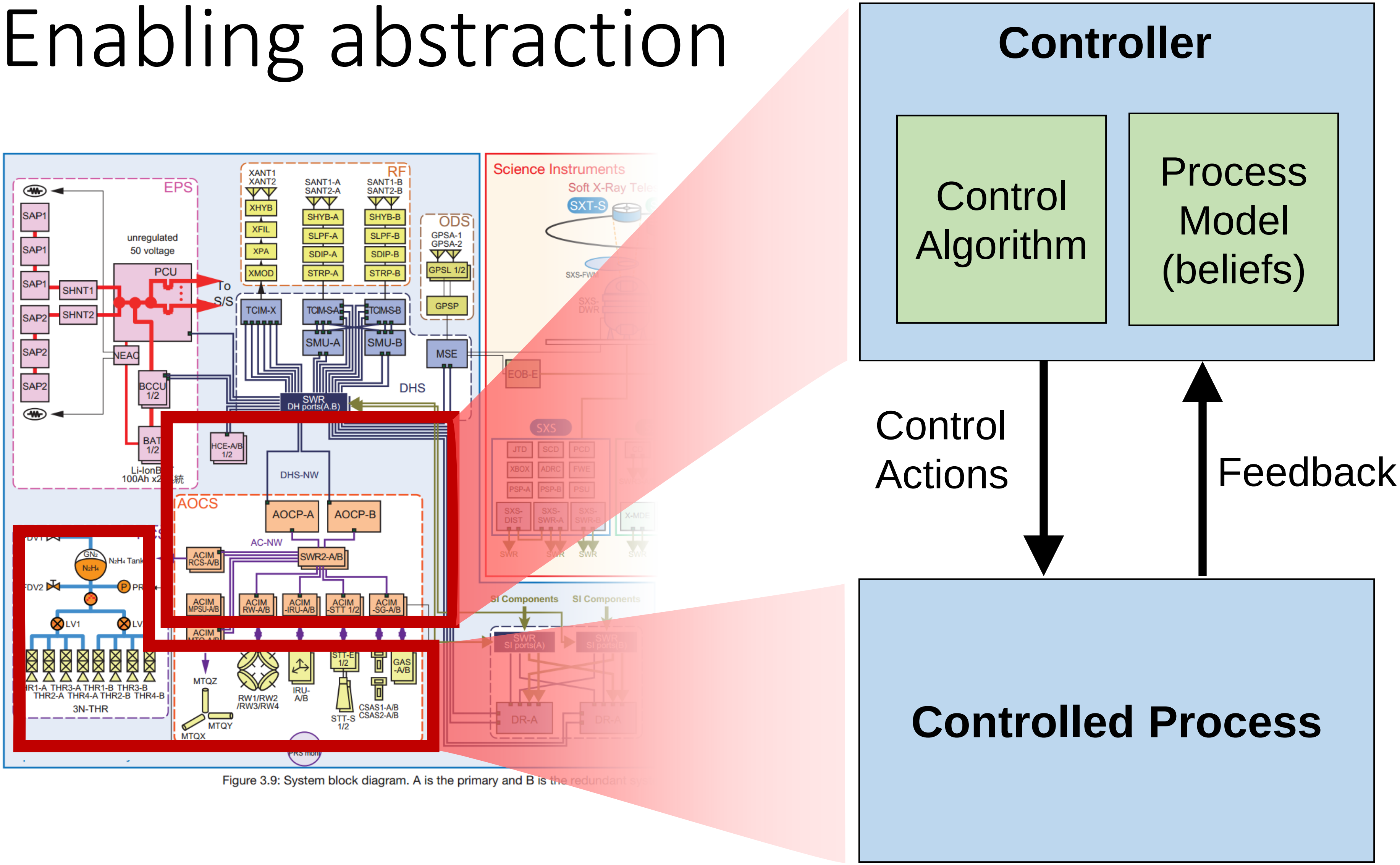


Figure 3.9: System block diagram. A is the primary and B is the redundant syst

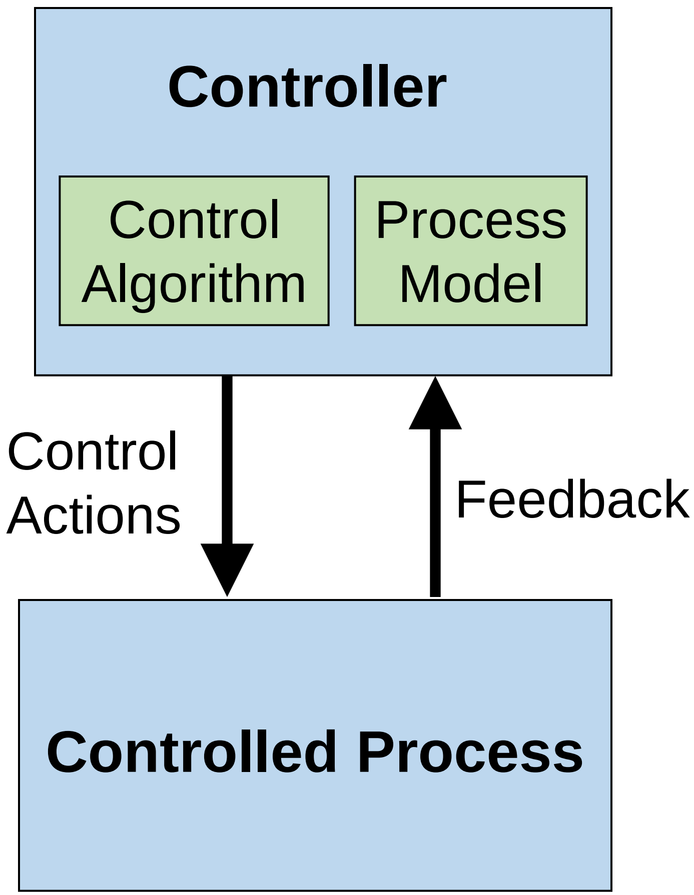
Thomas, 2017

Enabling abstraction



Thomas, 2017

Basic control loop



- **Control actions** are provided to affect a controlled process
- **Feedback** may be used to monitor the process
- **Process model** (beliefs) formed based on feedback and other information
- **Control algorithm** determines appropriate control actions given current beliefs

Enabling abstraction

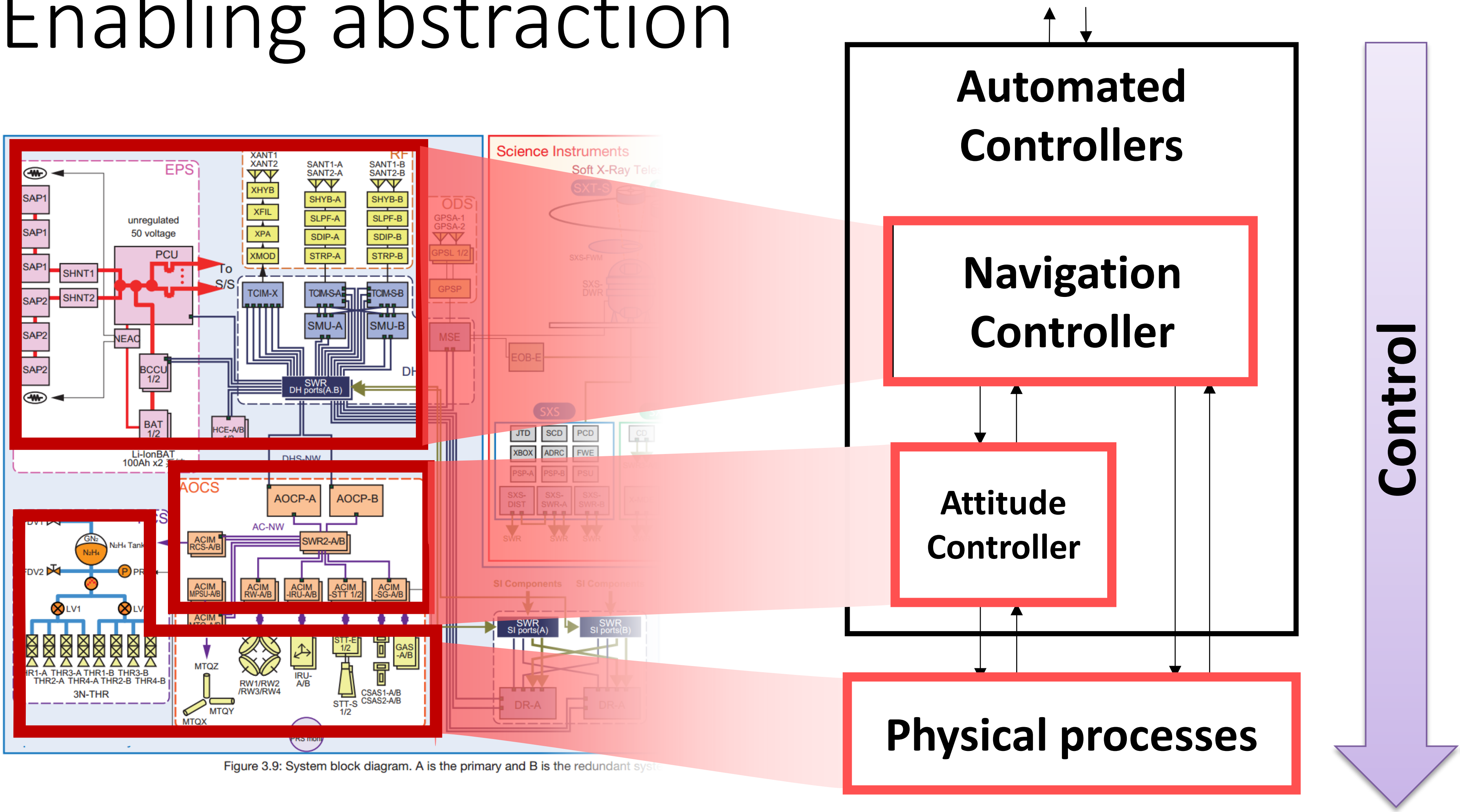
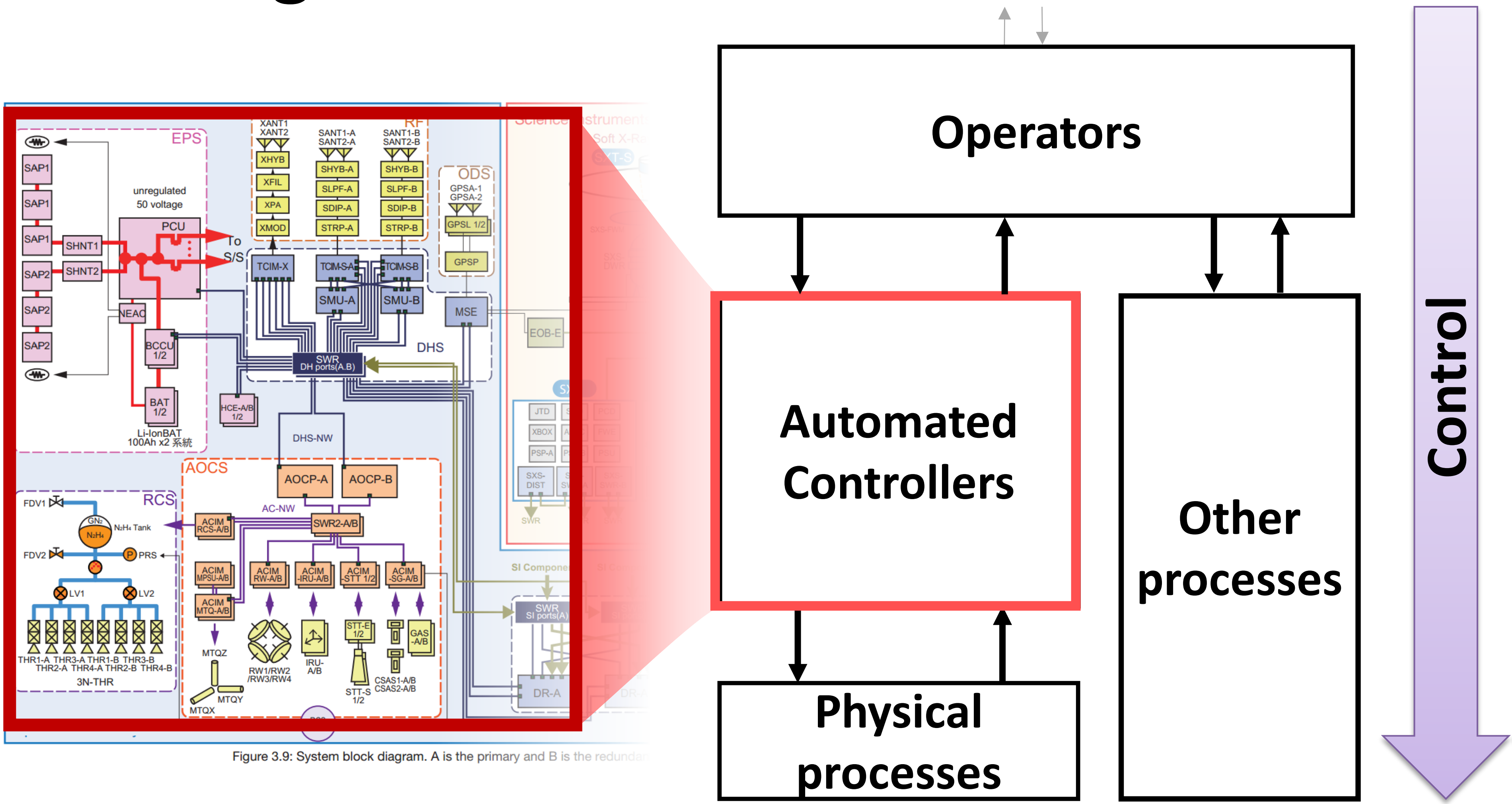


Figure 3.9: System block diagram. A is the primary and B is the redundant sys

Thomas, 2017

Enabling abstraction



Thomas, 2017

Enabling abstraction

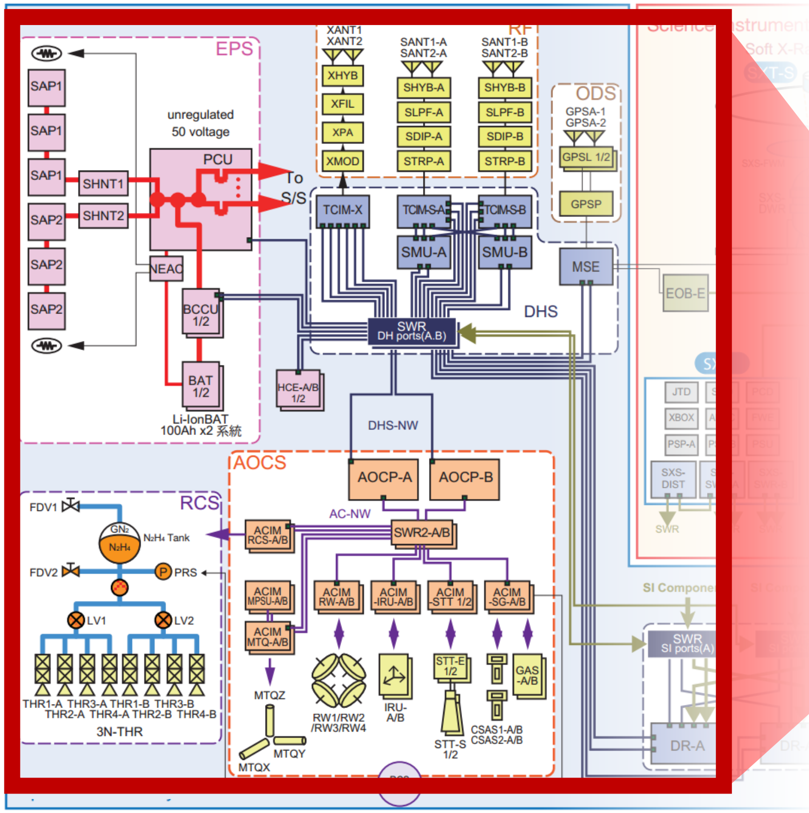
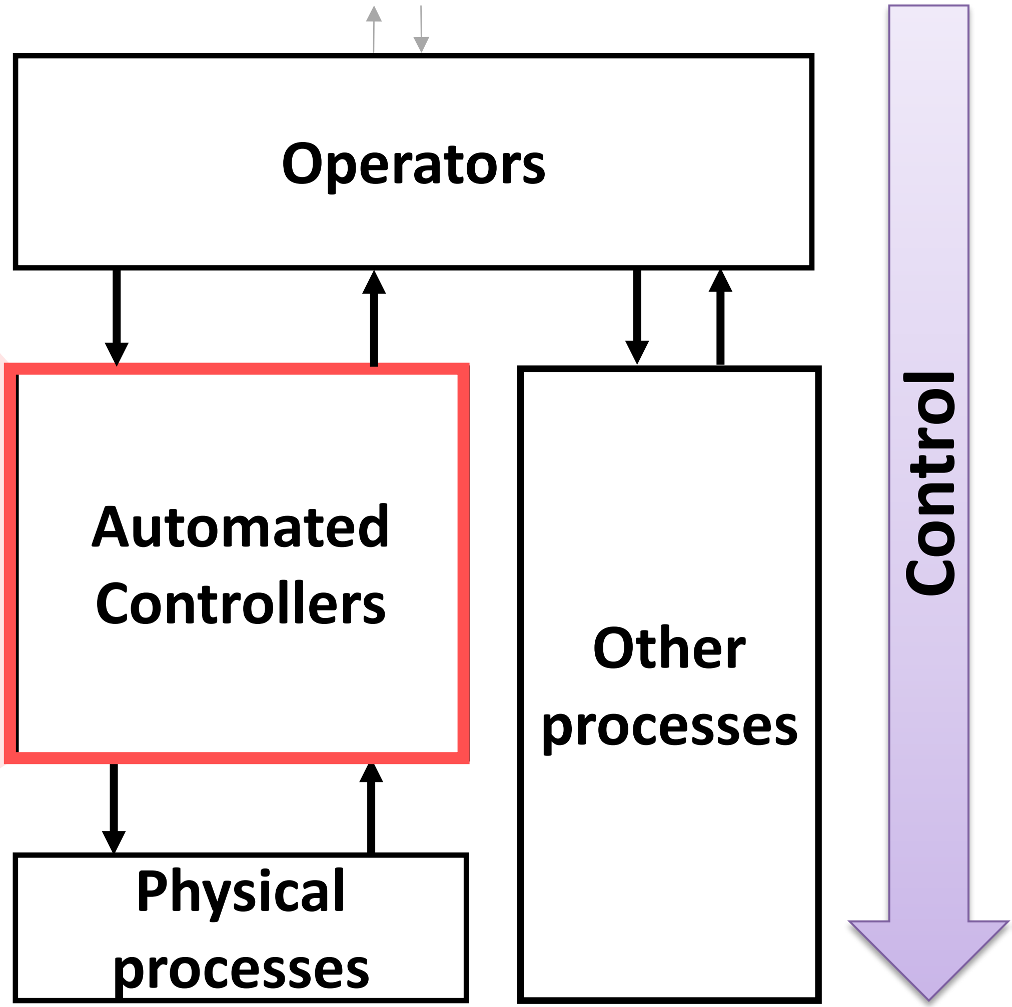


Figure 3.9: System block diagram. A is the primary and B is the redunda

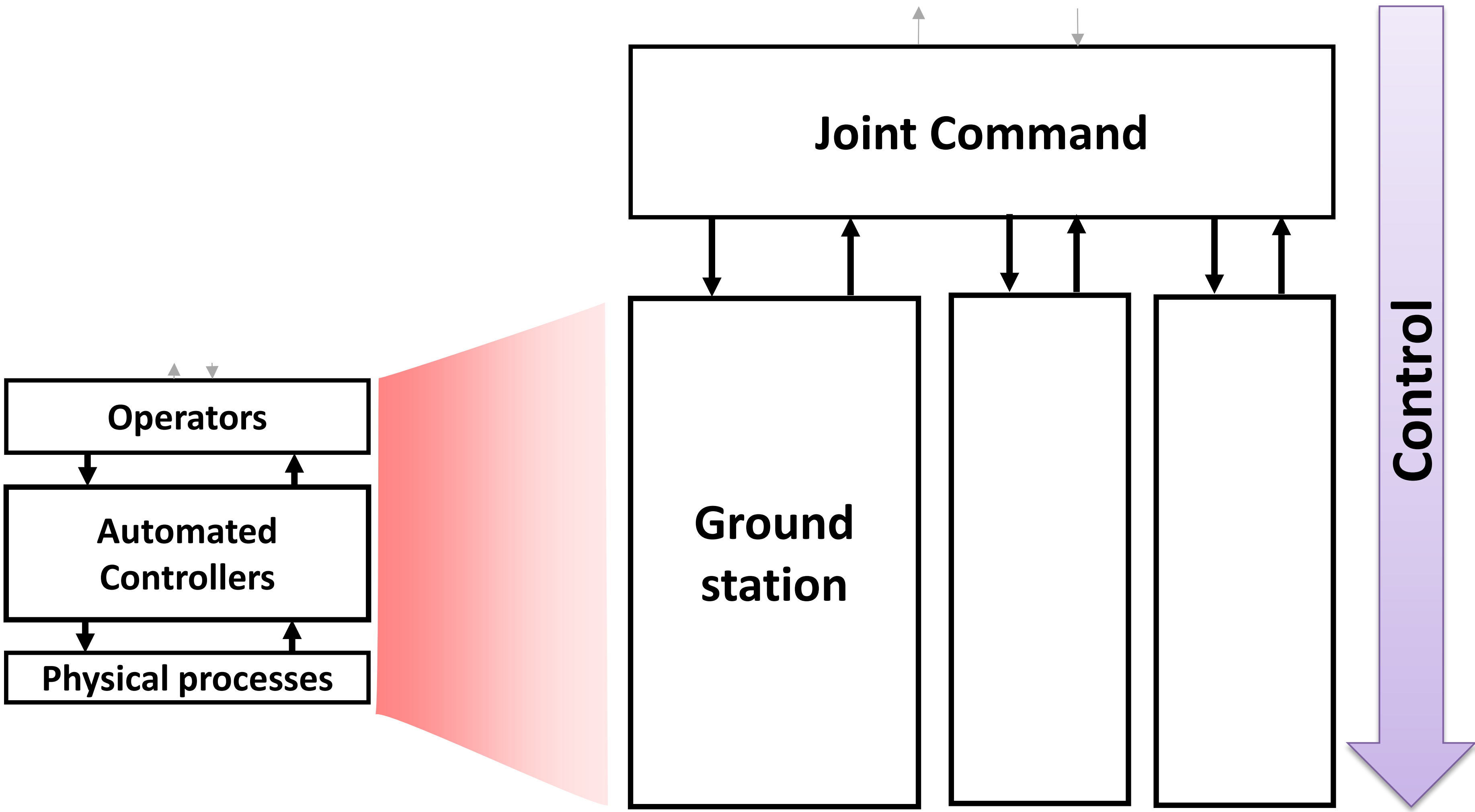


Component view

Systems view

Thomas, 2017

Enabling abstraction



STPA
Systems Theoretic Process Analysis

System-Theoretic Process Analysis (STPA)

1. Identify losses, system hazards — **Losses to prevent**
2. Draw control structure — **Model**
3. Identify unsafe control actions — **Behavior to prevent**
4. Identify loss scenarios — **How could behavior occur**

(Leveson and Thomas, 2018)

System-Theoretic Process Analysis (STPA)

1. Identify losses, system hazards
2. Draw control structure
3. Identify unsafe control actions
4. Identify loss scenarios

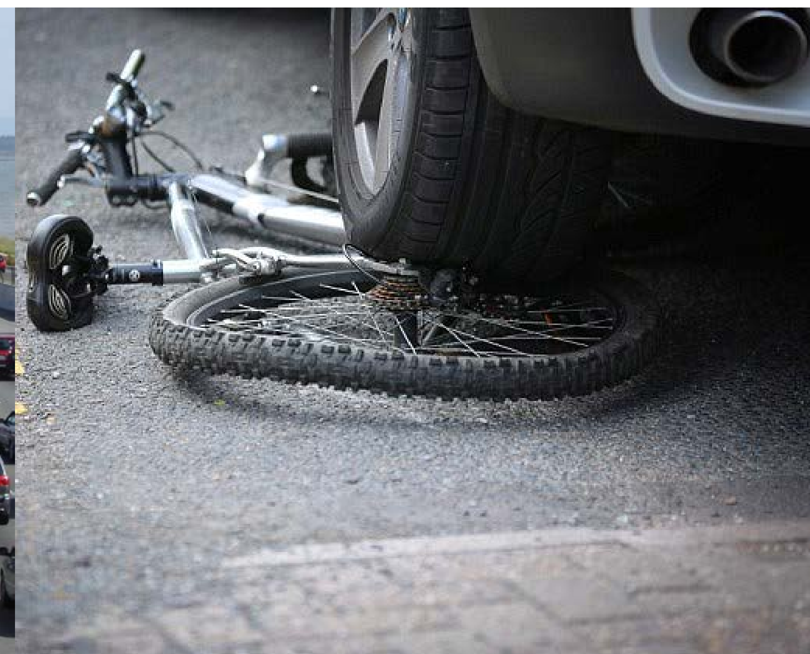
(Leveson and Thomas, 2018)

Aviation Example

- Losses
  - L-1. Loss of life or serious injury to people
  - L-2. Damage to the aircraft or objects outside the aircraft
  - L-3: Loss of mission (transportation)
  - L-4: Loss of performance / efficiency

Automotive Example

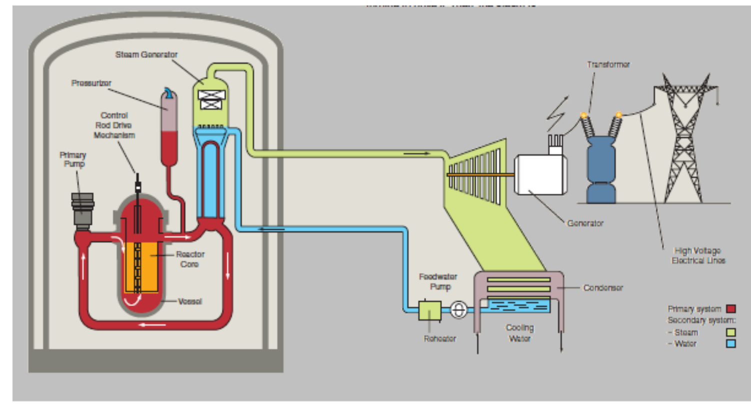
- Losses
  - L-1. Loss of life or serious injury to people
  - L-2. Damage to the vehicle or objects outside the vehicle
  - L-3: Loss of mission (transportation)
  - L-4: Loss of customer satisfaction



Nuclear Power Plant

Define Losses

- L-1: Loss of life or injury
- L-2: Equipment damage
- L-3: Environmental contamination
- L-4: Loss of power generation



Safety or Security?

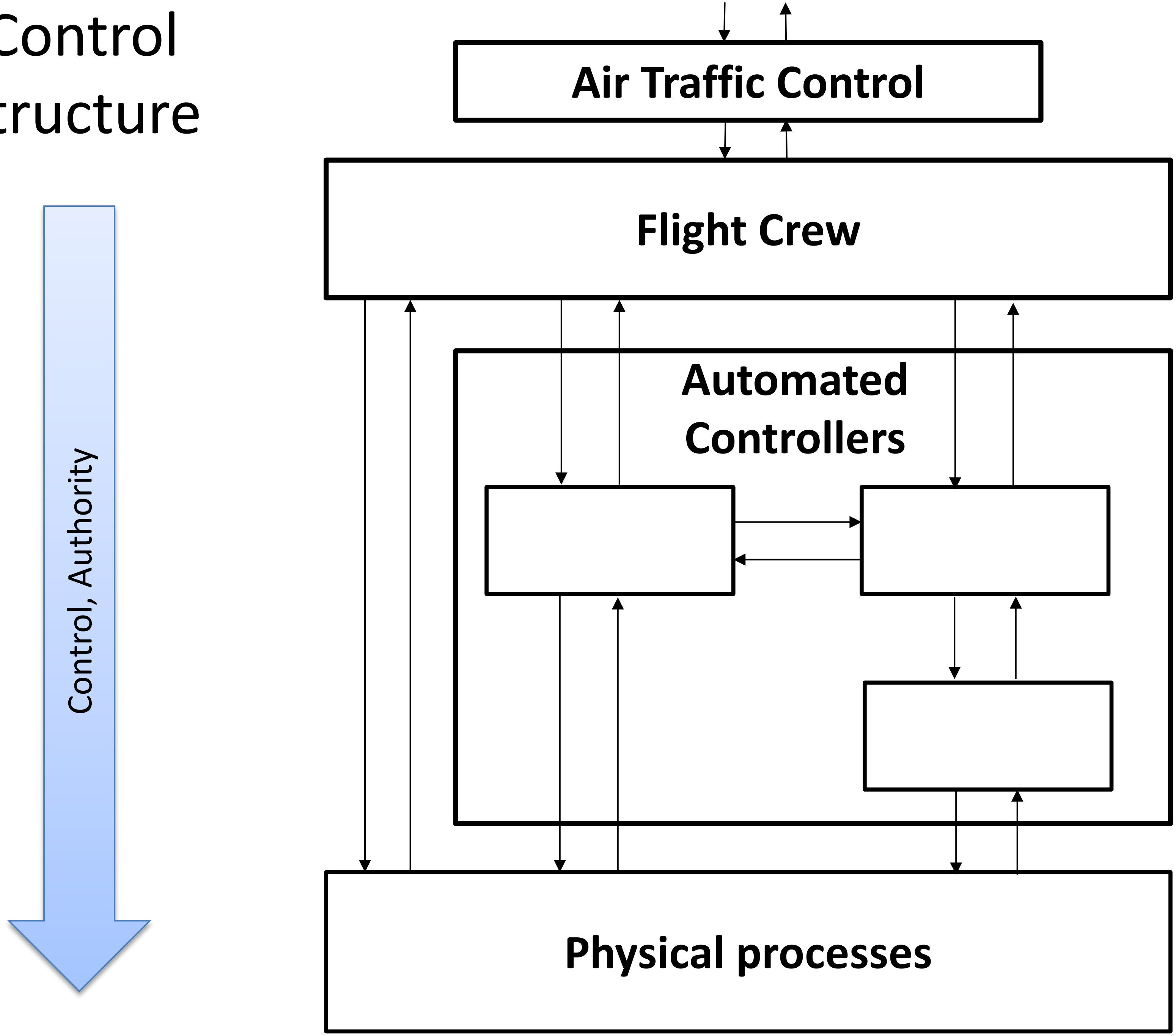
(Thomas, 2014)

System-Theoretic Process Analysis (STPA)

1. Identify losses, system hazards
2. Draw functional control structure
3. Identify unsafe control actions
4. Identify loss scenarios

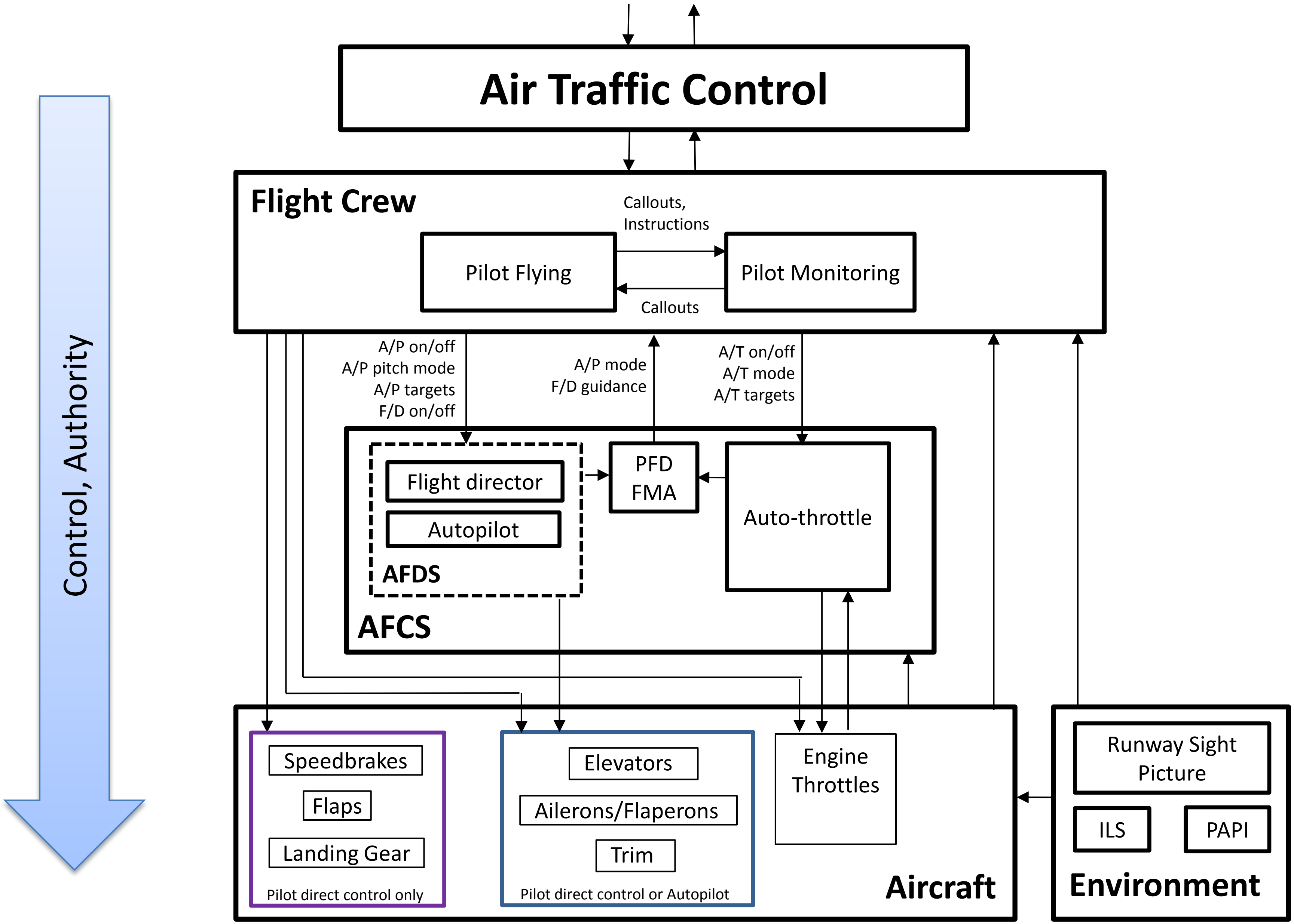
(Leveson and Thomas, 2018)

Control structure



(Thomas, 2017)

Control Structure



(Thomas, 2017)

Adaptive Cruise Control

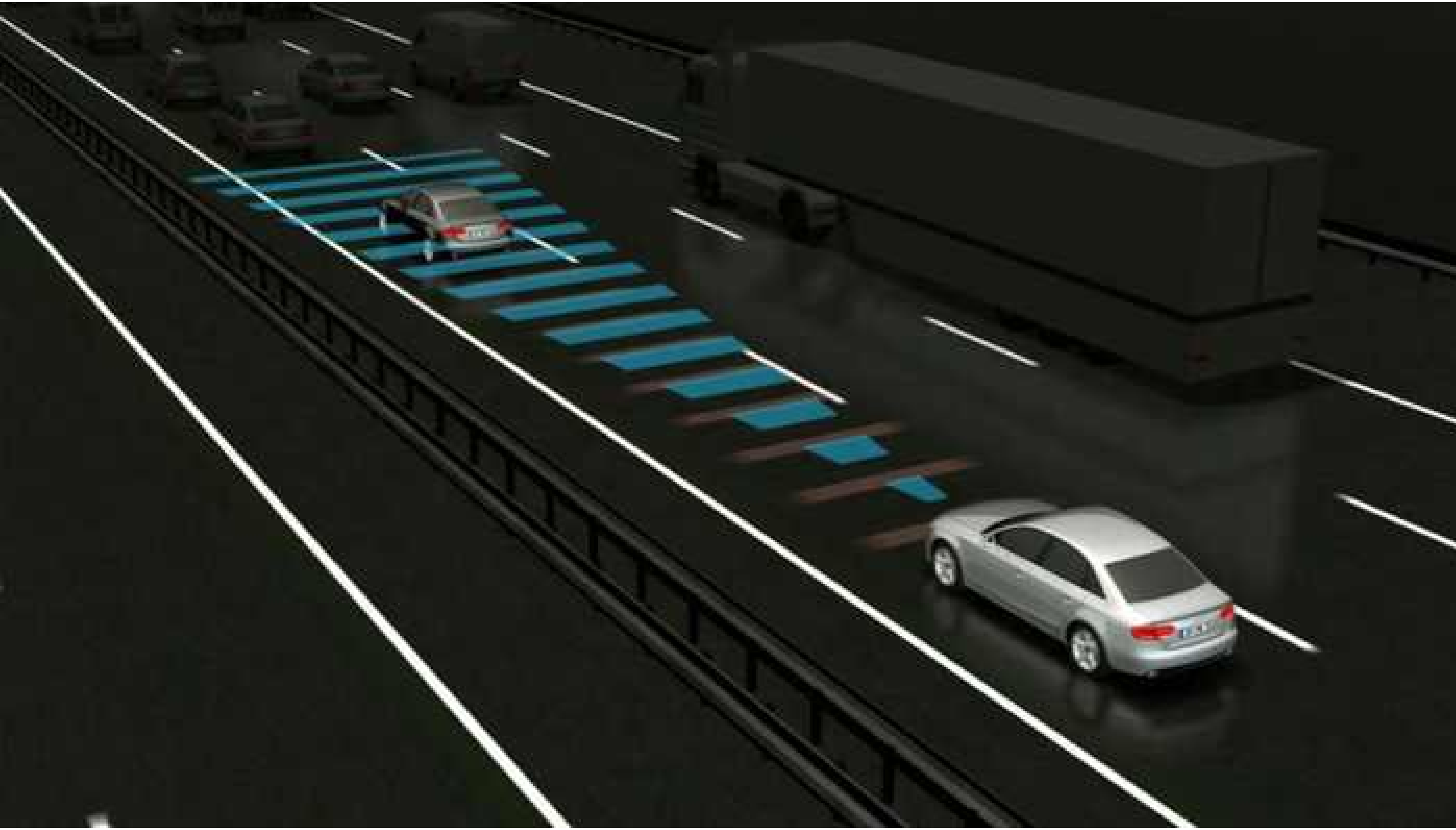
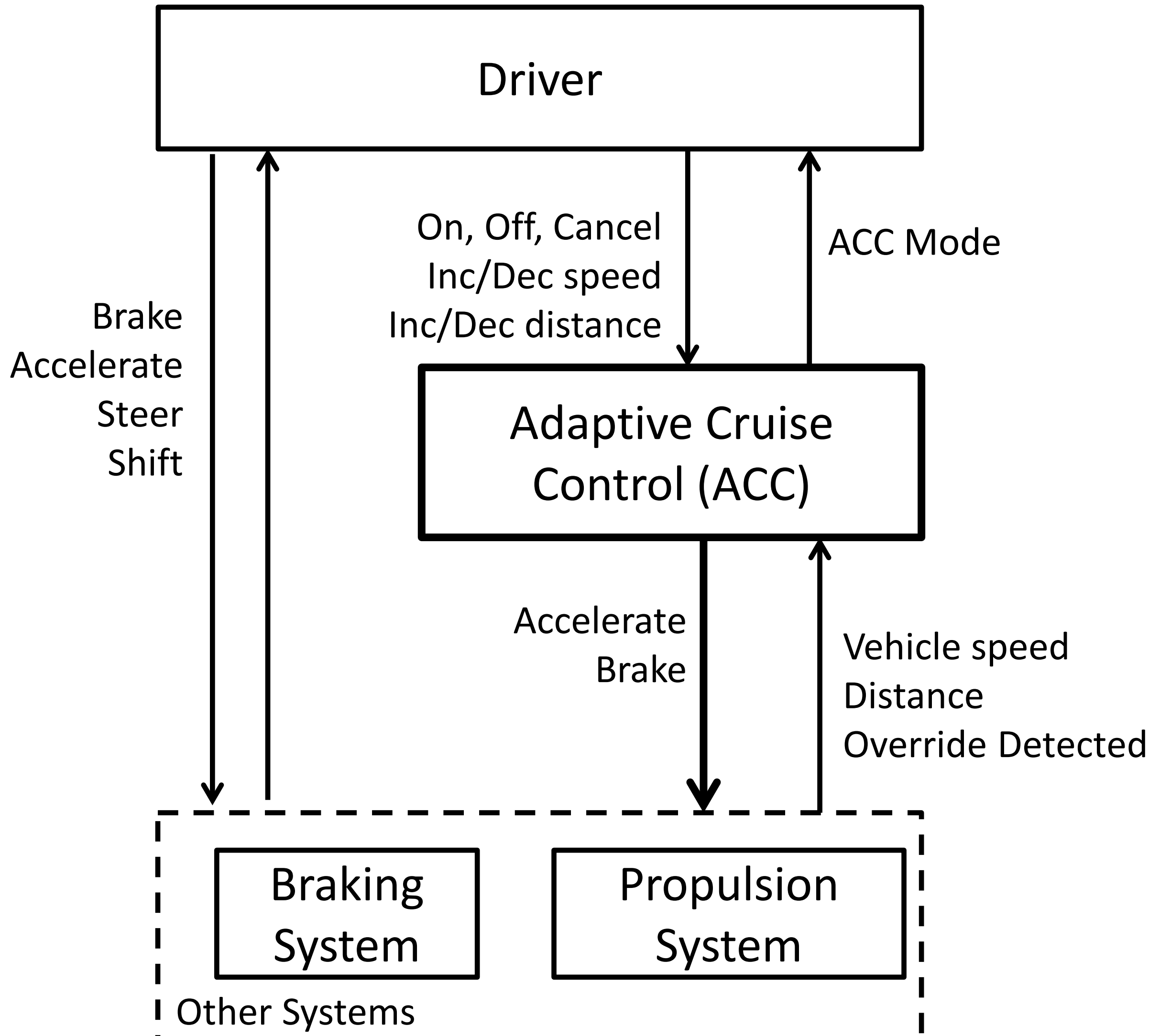


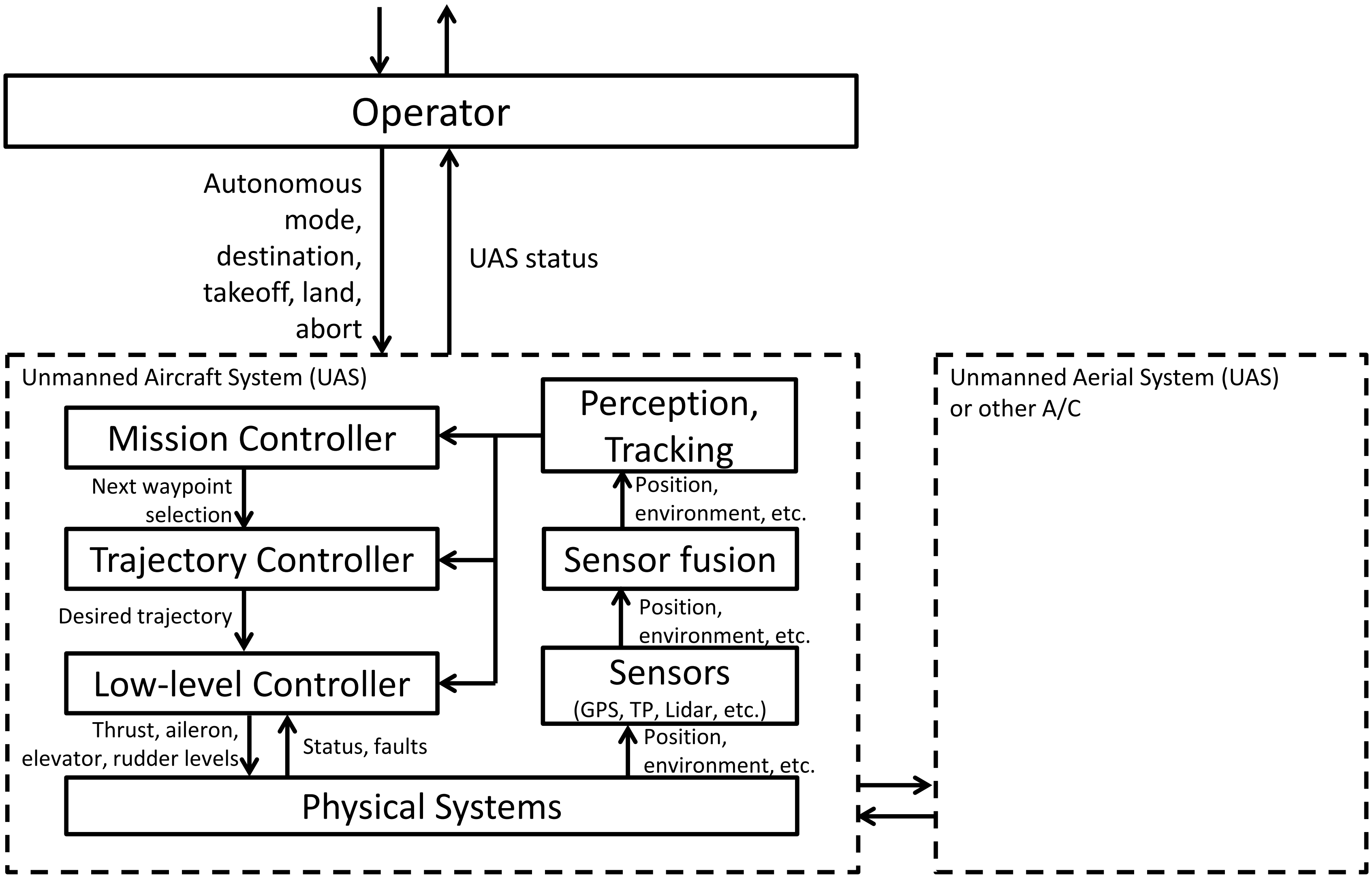
Image from: http://www.audi.com/etc/medialib/ngw/efficiency/video_assets/fallback_videos.Par.0002.Image.jpg

Adaptive Cruise Control (ACC) Control Structure



Thomas, 2017

Example Concept



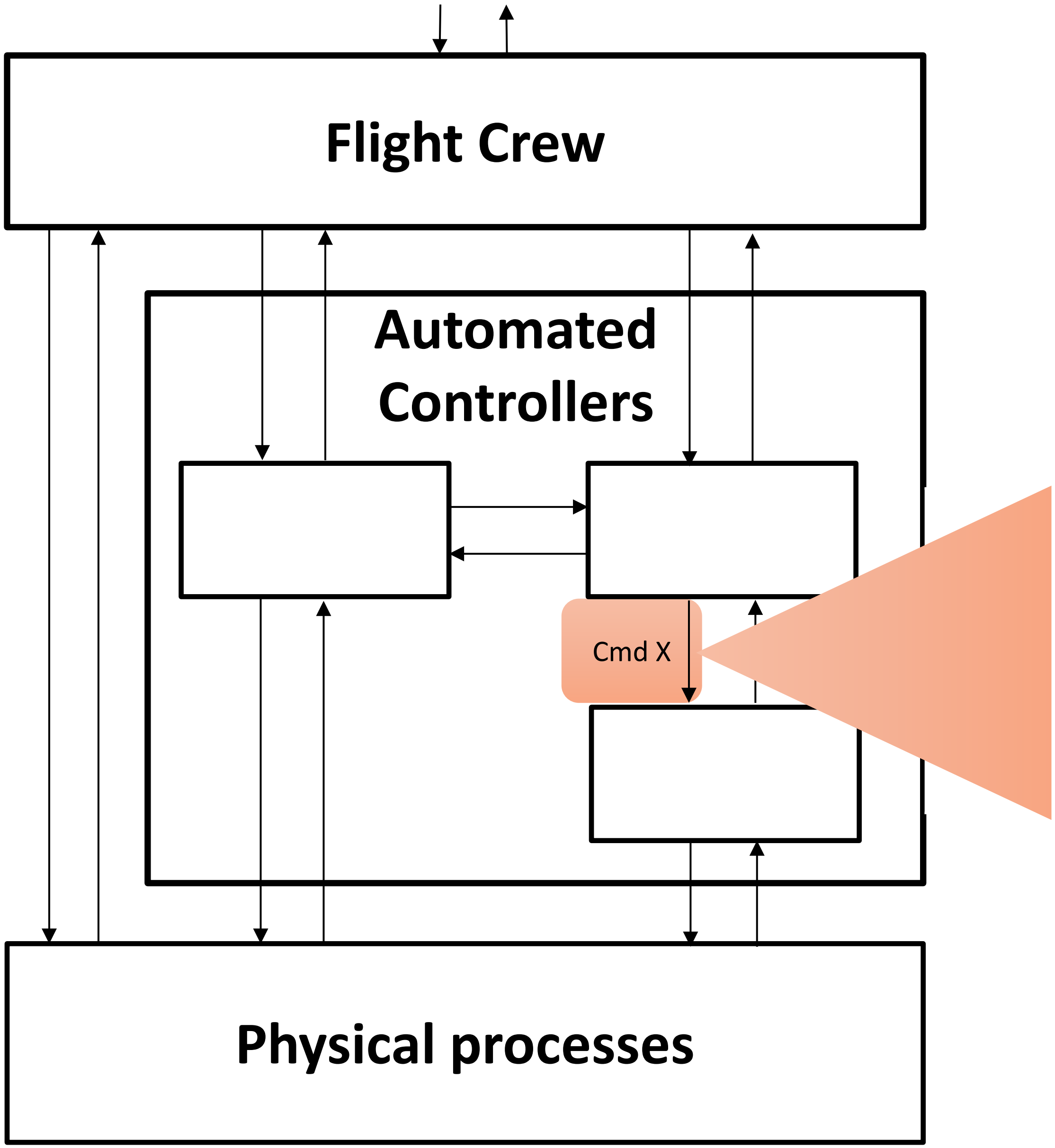
Thomas, 2018

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards
2. Draw functional control structure
3. Identify unsafe control actions
4. Identify loss scenarios

(Leveson and Thomas, 2018)

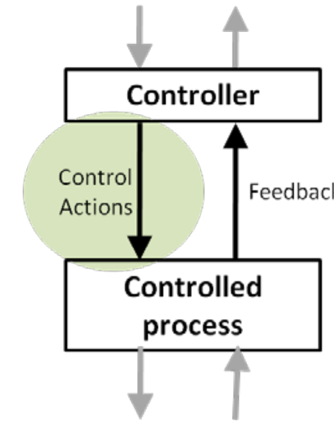
STPA: Identify Unsafe Control Actions (UCA)



Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long

(Thomas, 2017)

Structure of an Unsafe Control Action



Example:

"Computer does not provide Reverse Thrust cmd when aircraft lands"

Source Controller

Type

Control Action

Context

Four parts of an unsafe control action

– Source Controller: the controller that can provide the control action
– Type: whether the control action provided, not provided, etc.
– Control Action: the controller's command that was provided / missing
– Context: conditions for the hazard to occur
    - (system or environmental state in which command is provided)

Thomas, 2017

Component Requirements

Unsafe Control Action	**Component Requirement**
UCA-1: Computer does not provide Reverse-Thrust cmd when aircraft lands [H-3]	R-1: Computer shall provide Reverse-Thrust cmd when aircraft lands and RT armed [UCA-1]

Generating constraints and requirements

	Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
Cmd				

High-level safety constraints
Controller X shall not allow A
Controller X shall enforce B
Etc.

Controller X <u>shall provide</u> CMD when D	Controller X <u>shall not provide</u> CMD when E	Controller X shall provide CMD <u>within Y seconds</u> of F	Controller X shall <u>stop providing</u> CMD within Z seconds of G

Controller functional safety requirements

(Thomas, 2017)

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards

2. Draw functional control structure

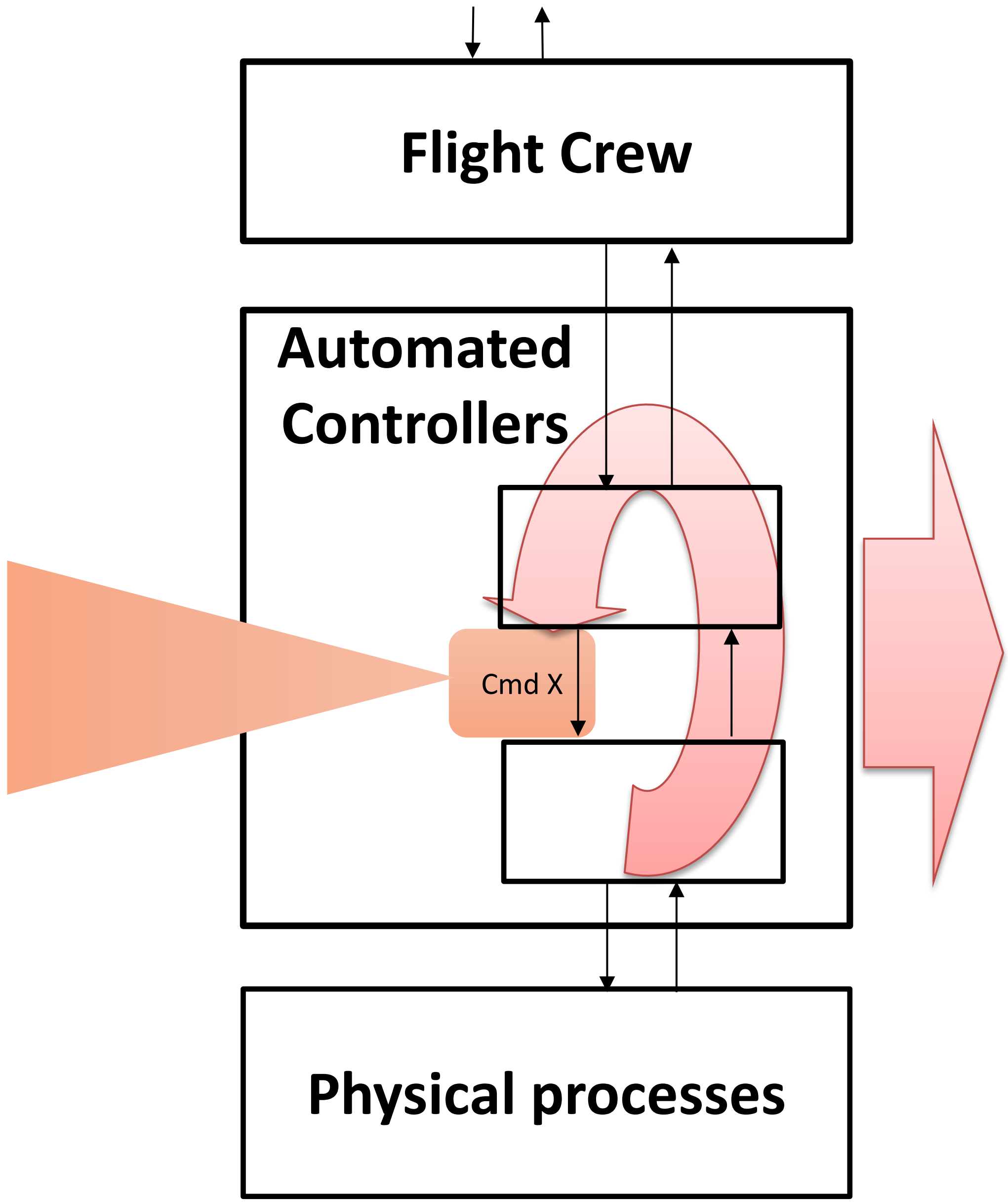
3. Identify unsafe control actions

4. Identify loss scenarios

(Leveson and Thomas, 2018)

Identify loss scenarios

What could cause Unsafe Control Actions?



Scenarios
Controller incorrectly believes X because …
Controller control algorithm does not enforce Y because …
Incorrect feedback Z received because …
Sensor failure causes…
Etc.

(Thomas, 2017)

Design recommendations and component requirements

Scenarios

Design recommendations
Component A should be able to respond within B seconds to avoid C
Controller X should take into consideration D to prevent E
Etc.

Rationale and assumptions identified

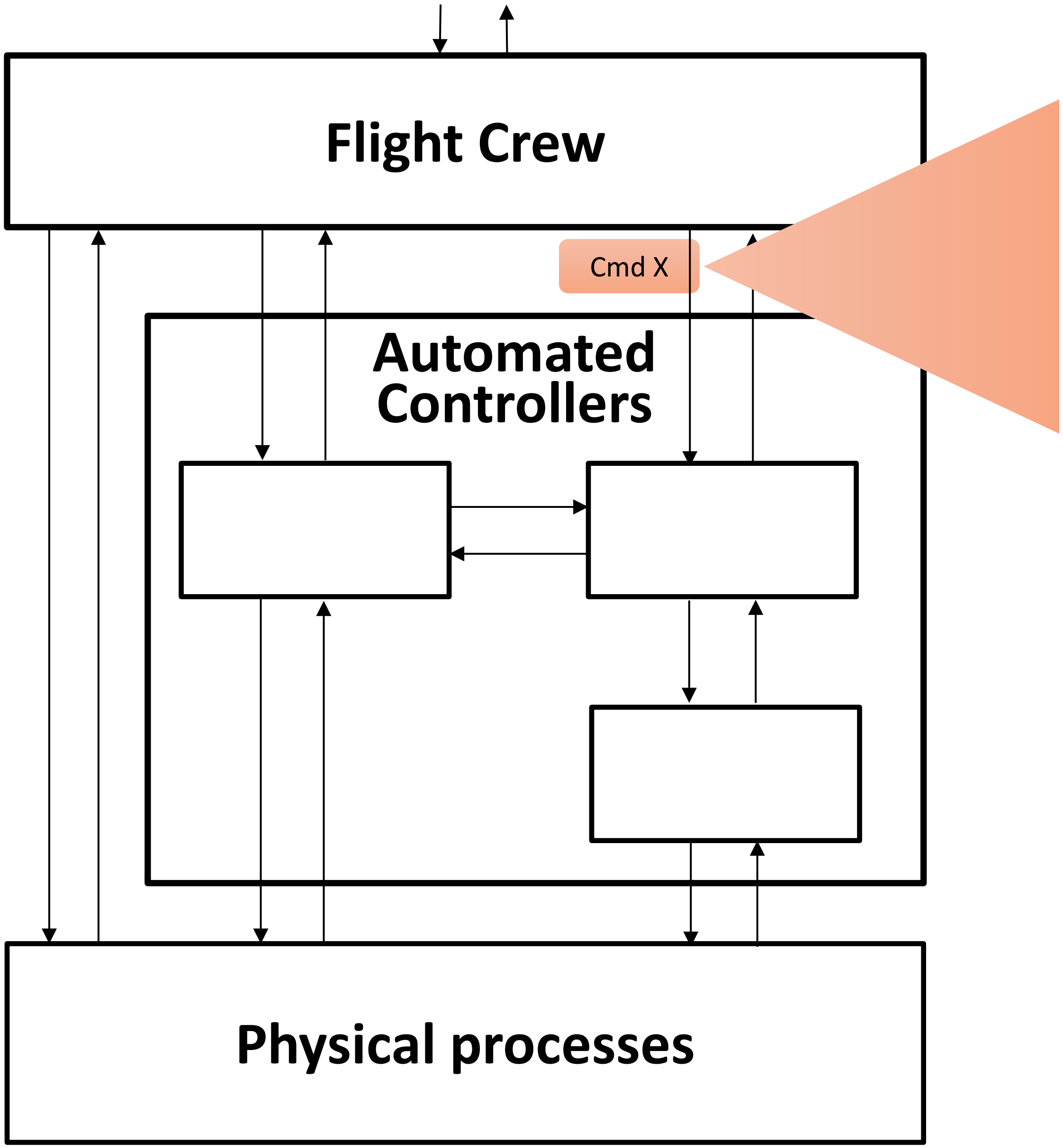
Component requirements
Component F shall automatically operate within G seconds when H
Component I and J shall be operated at the same time to prevent K
Etc.

Every recommendation and requirement is traceable

(Thomas, 2017)

What about human interactions?

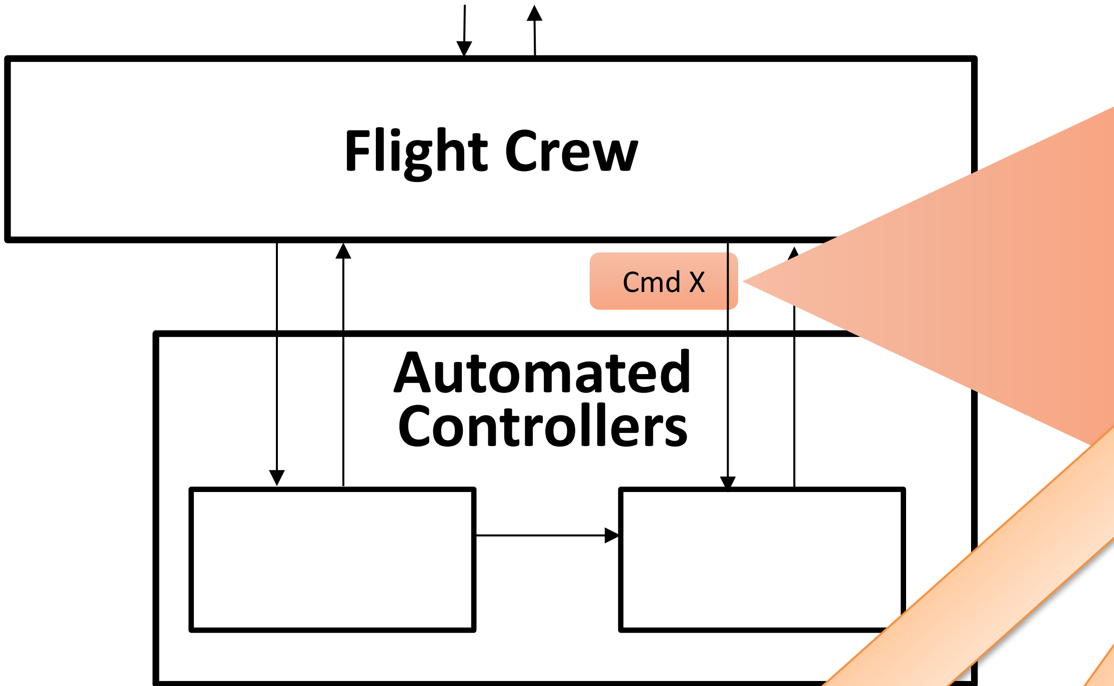
Unsafe Control Actions (UCA)



Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long

(Thomas, 2017)

Unsafe Control Actions (UCA)



Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long

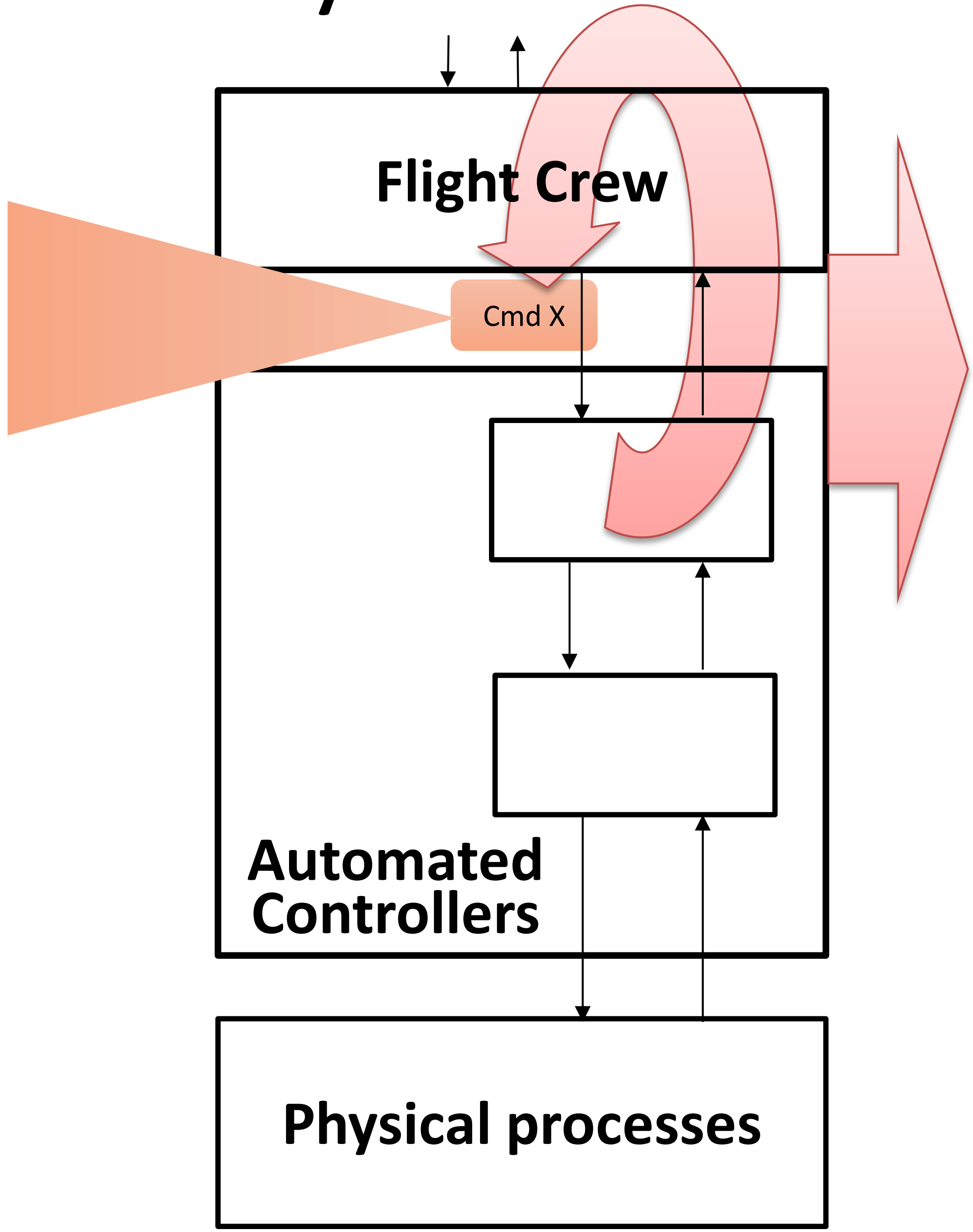
Crew shall provide CMD when D	Crew shall not provide CMD when E	Crew shall provide CMD within Y seconds of F	Crew shall stop providing CMD within Z seconds of G

Crew procedures

(Thomas, 2017)

Identify loss scenarios

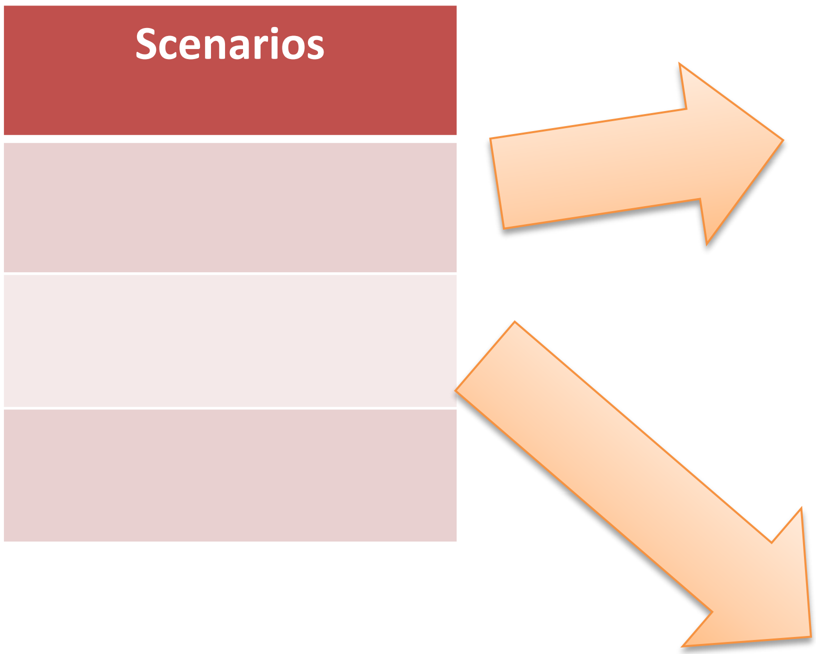
What could cause Unsafe Control Actions?



Scenarios
Crew incorrectly believes X because …
Crew does not perform Y because …
Crew received incorrect feedback Z because …
Etc.

(Thomas, 2017)

Design decisions and recommendations



Design decisions
Crew must be notified of A within B seconds to avoid C
Component F should operate automatically when H
Etc.

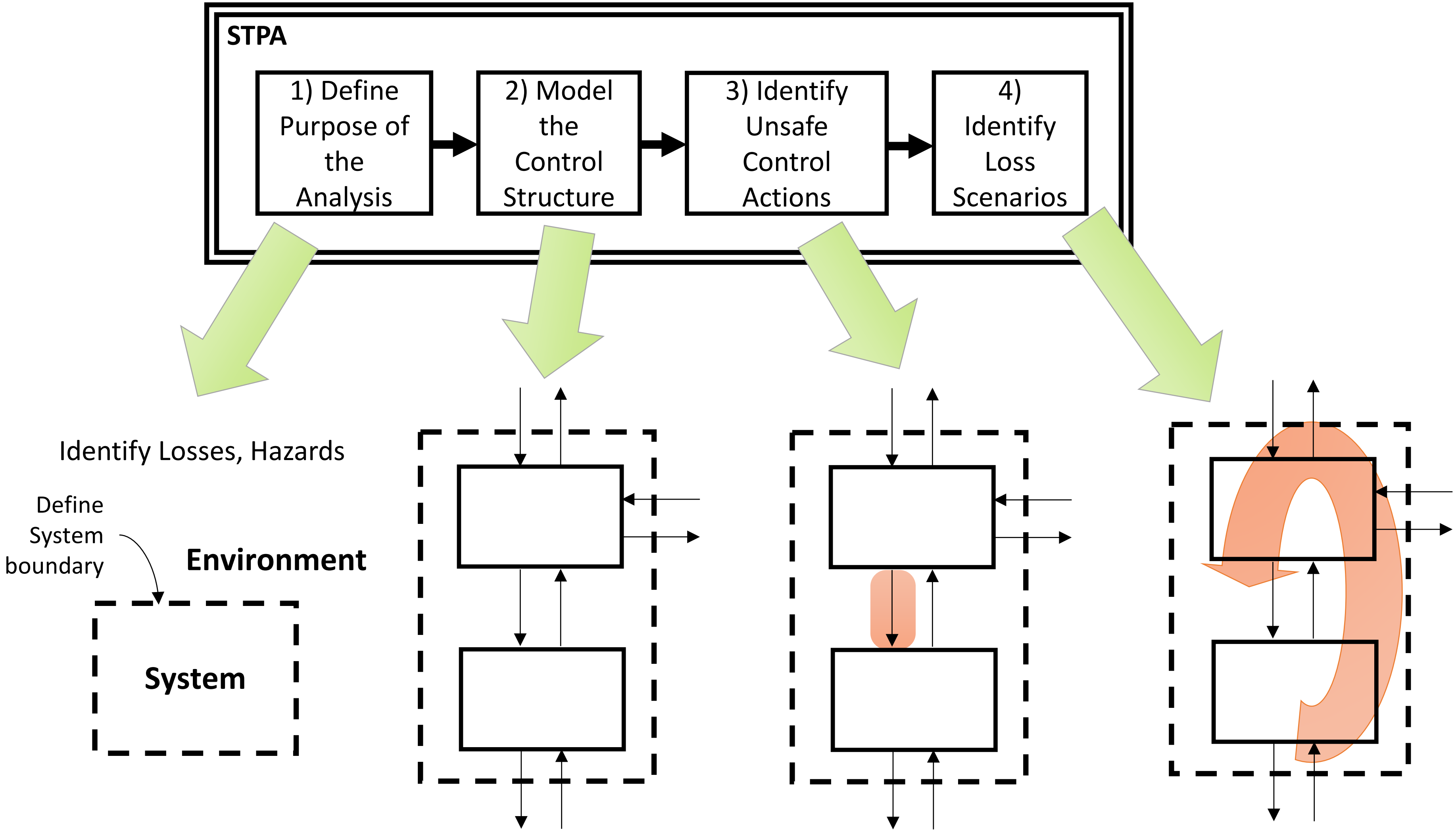
Rationale and assumptions identified

Recommendations
Crew X should take into consideration D to prevent E
Crew should operate I and J at the same time to prevent K
Etc.

Every recommendation and decision is traceable

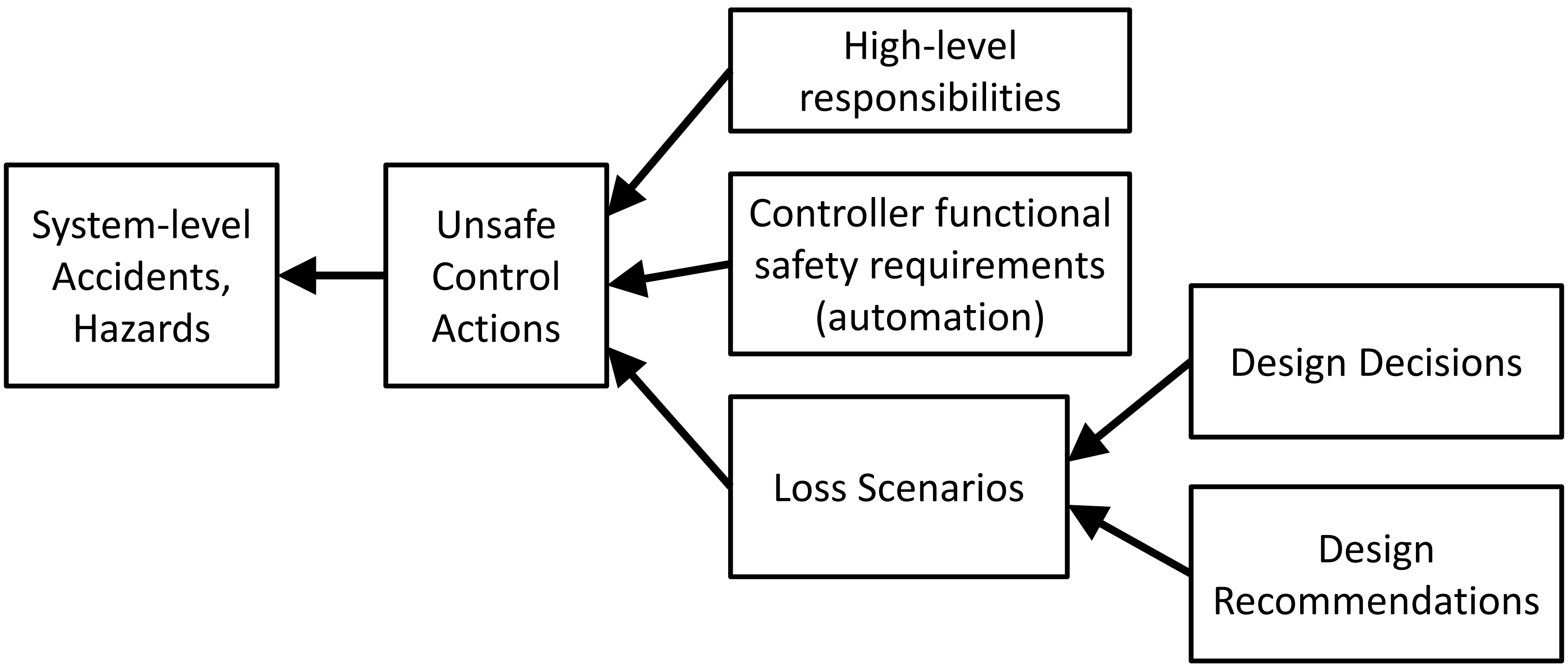
(Thomas, 2017)

STPA Overview



(Leveson and Thomas, 2018)

Traceability is maintained throughout



Thomas, 2017

Short STPA example

Google's self-driving car

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards
2. Draw functional control structure
3. Identify unsafe control actions
4. Identify loss scenarios

(Leveson and Thomas, 2018)

Losses

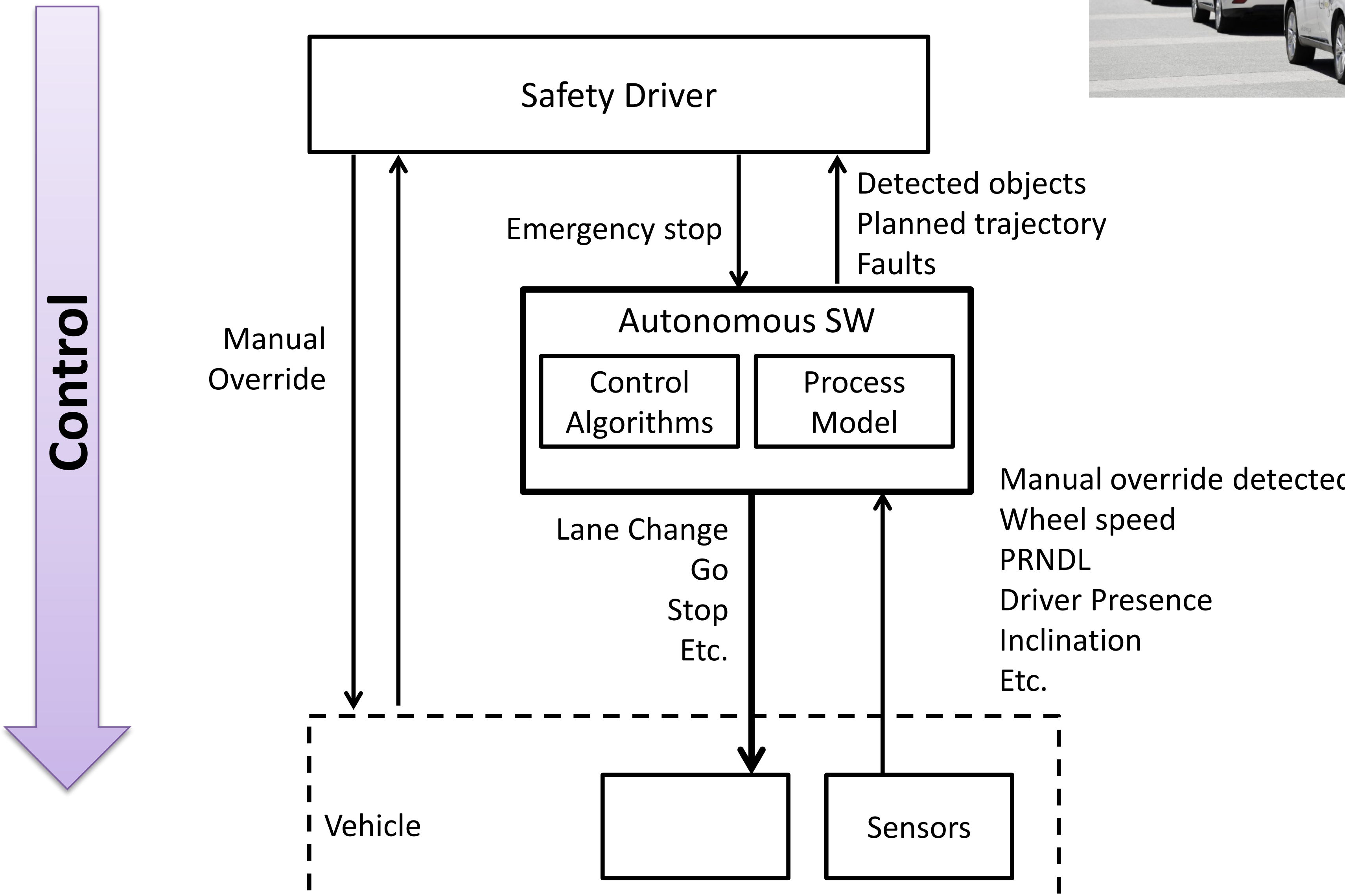
- Losses
  - L-1. Loss of life or serious injury to people
  - L-2. Damage to the vehicle or objects outside the vehicle

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards
2. **Draw functional control structure**
3. Identify unsafe control actions
4. Identify loss scenarios

(Leveson and Thomas, 2018)

High-level Control Structure



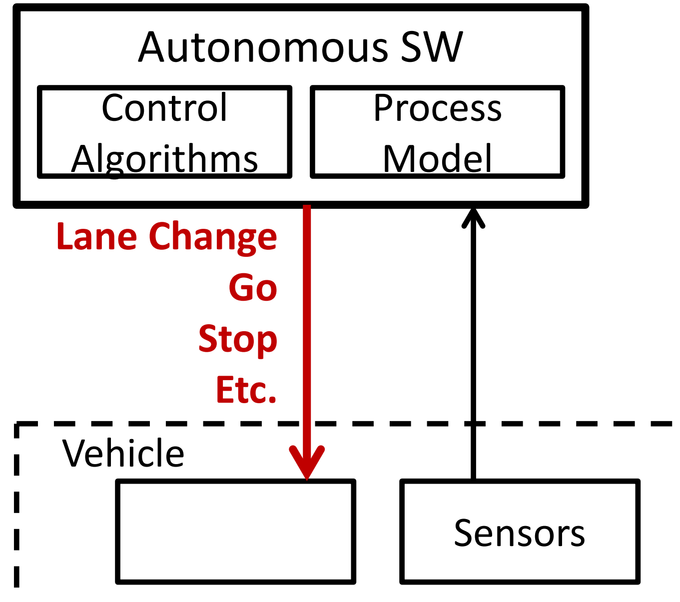
Thomas, 2018

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards
2. Draw functional control structure
3. Identify unsafe control actions
4. Identify loss scenarios

(Leveson and Thomas, 2018)

STPA: Unsafe Control Actions (UCA)



Example:

"Driver provides Park cmd when vehicle is moving"

Source Controller

Type

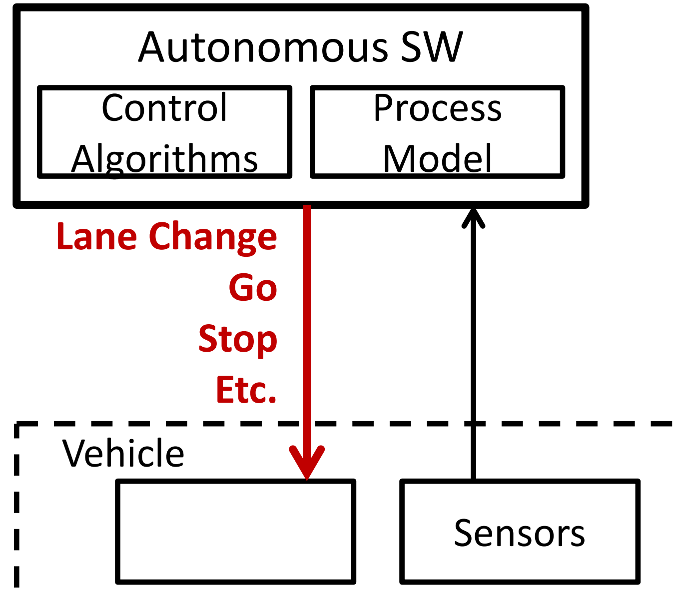
Control Action

Context

	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped Too Soon / Applied too long
Lane Change Command		?		

Thomas, 2018

STPA: Unsafe Control Actions (UCA)



Example:

"Driver provides Park cmd when vehicle is moving"

Source Controller — Driver

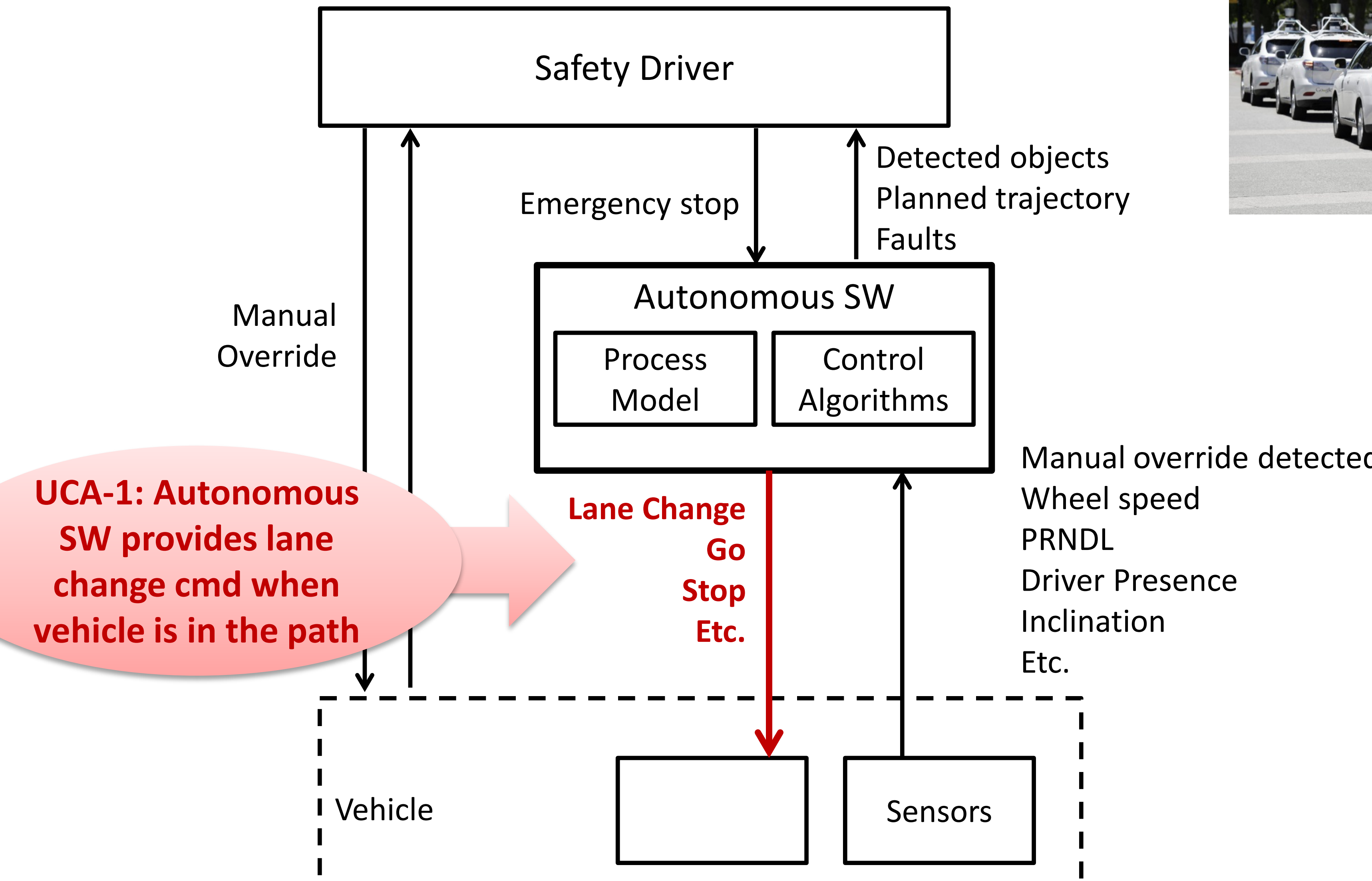
Type — provides

Control Action — Park cmd

Context — vehicle is moving

	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped Too Soon / Applied too long
Lane Change Command		Autonomous SW provides lane change cmd when object is in the path		

Unsafe Control Actions (UCA)



Thomas, 2018

System-Theoretic Process Analysis (STPA)

1. Identify losses (accidents), system hazards

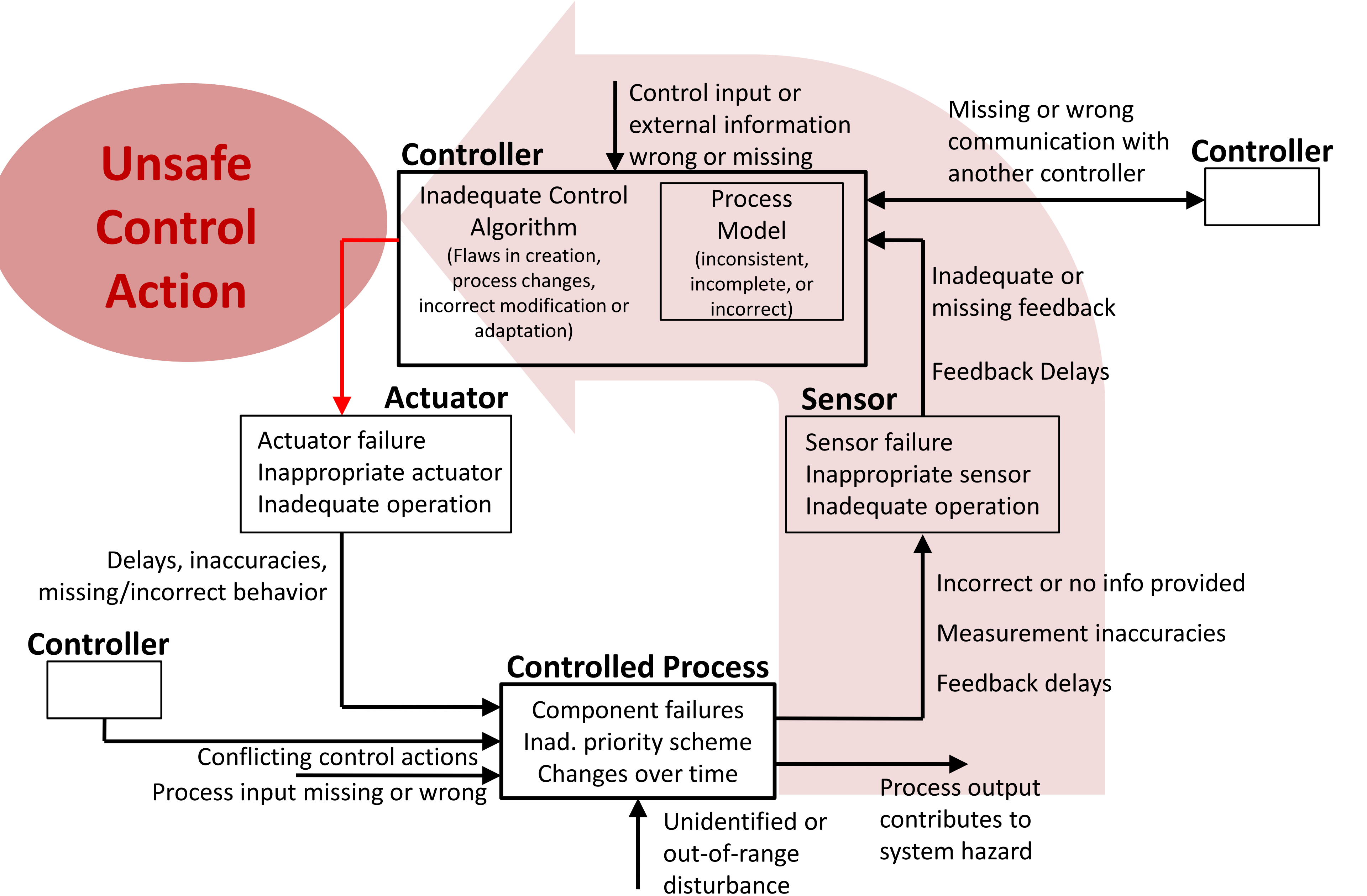
2. Draw functional control structure

3. Identify unsafe control actions

4. Identify loss scenarios

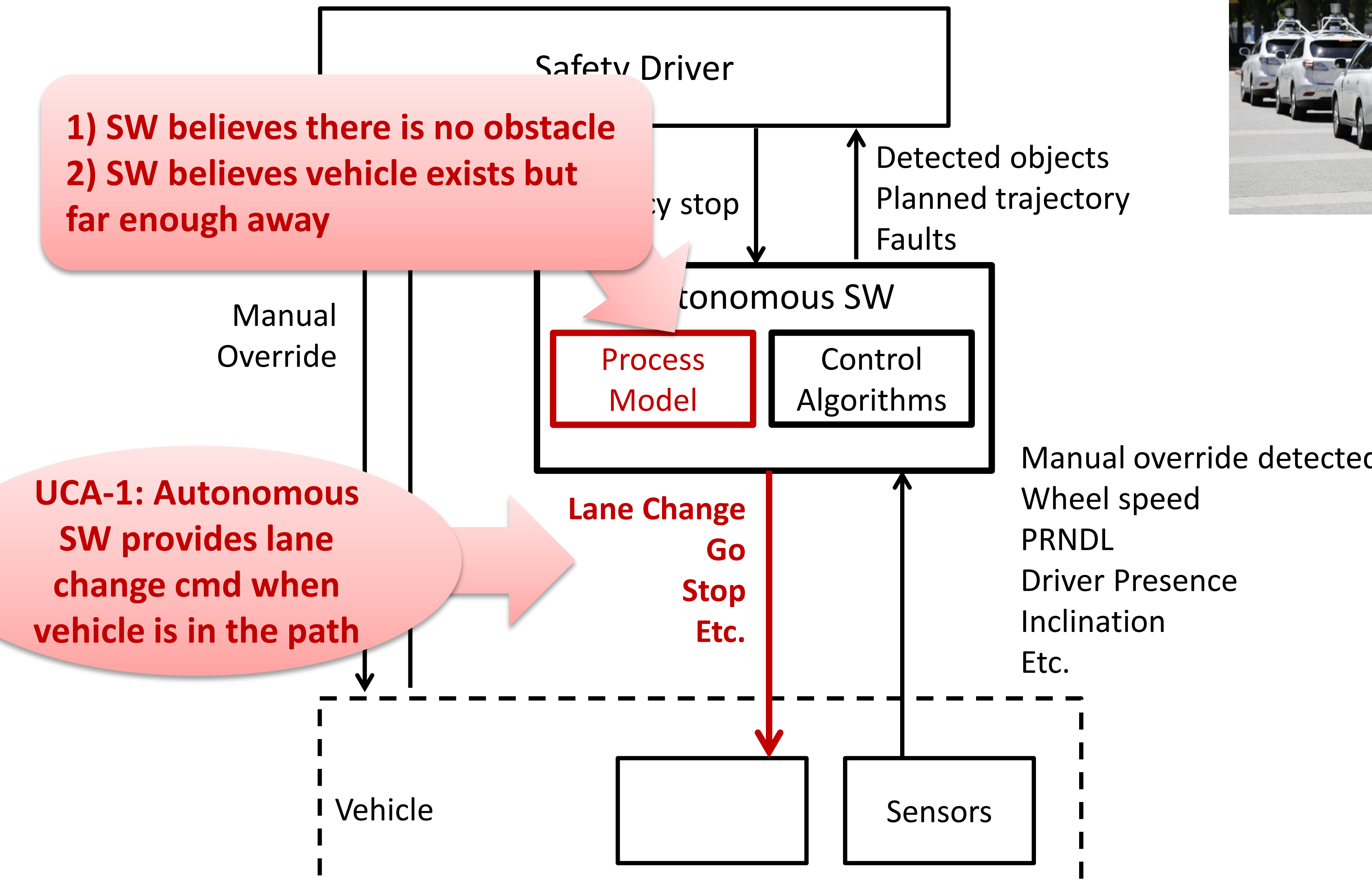
(Leveson and Thomas, 2018)

Potential control flaws

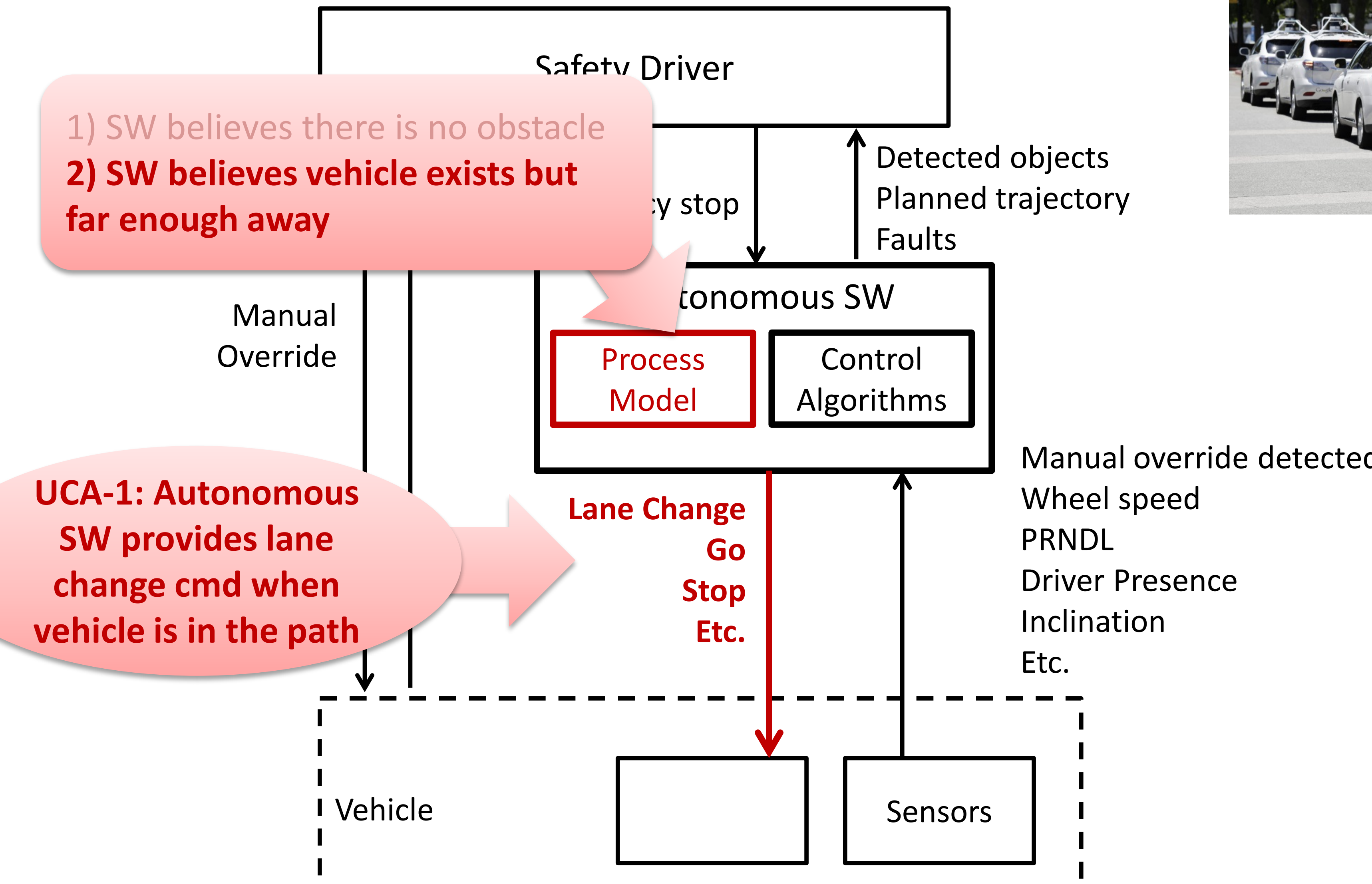


(Leveson, 2012)

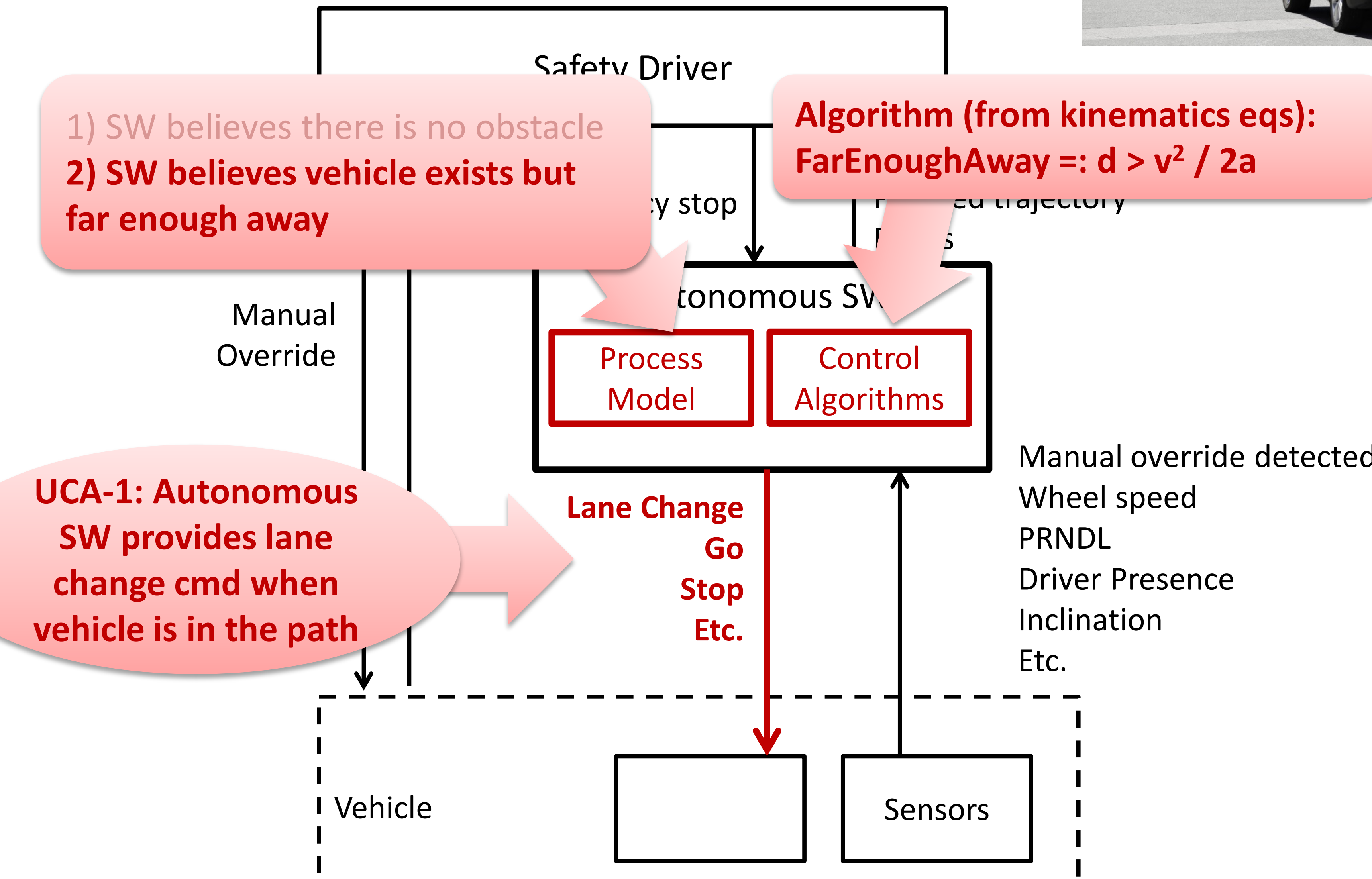
Why would SW provide UCA-1?



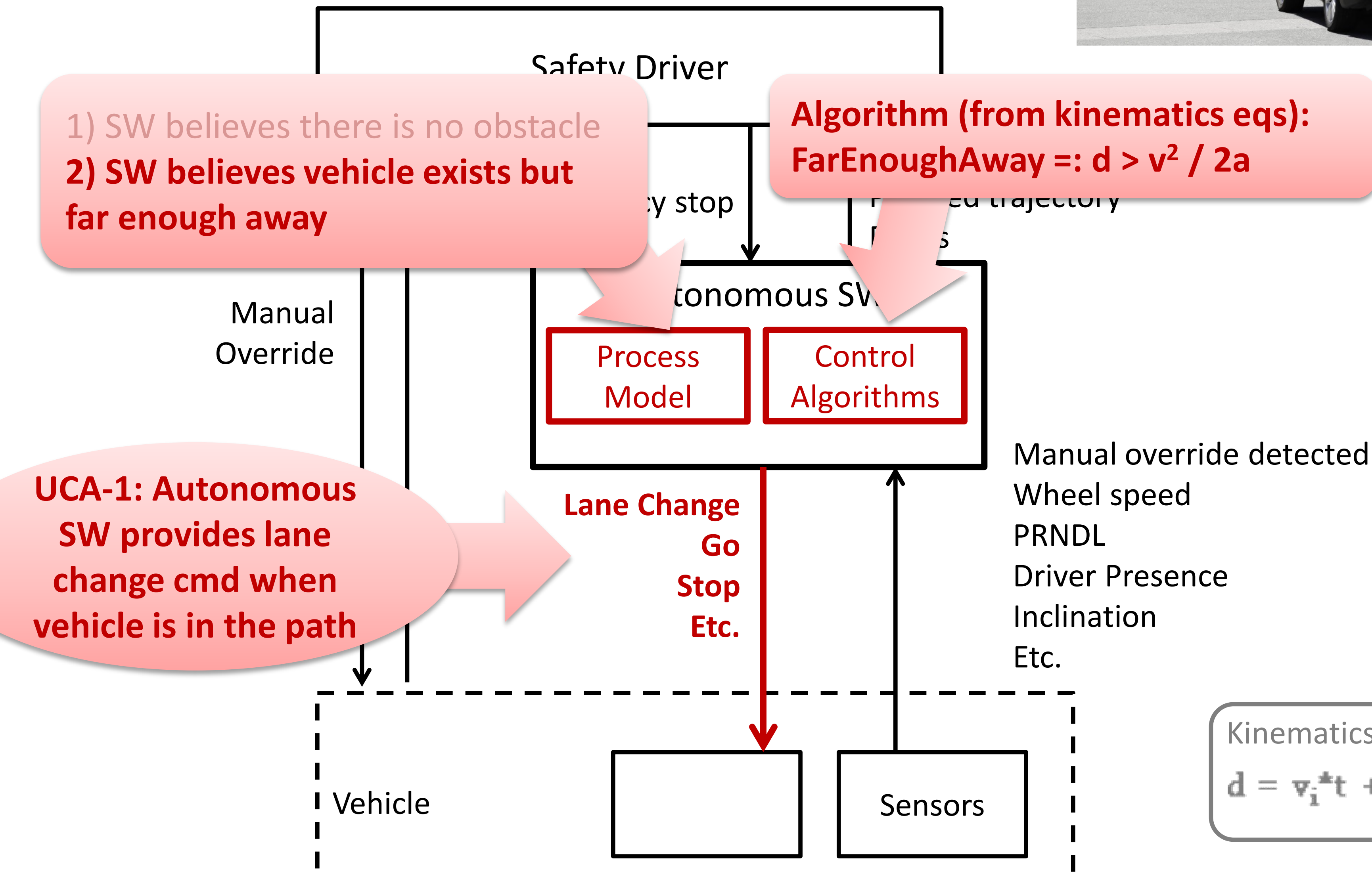
Why would SW provide UCA-1?



STPA scenario building



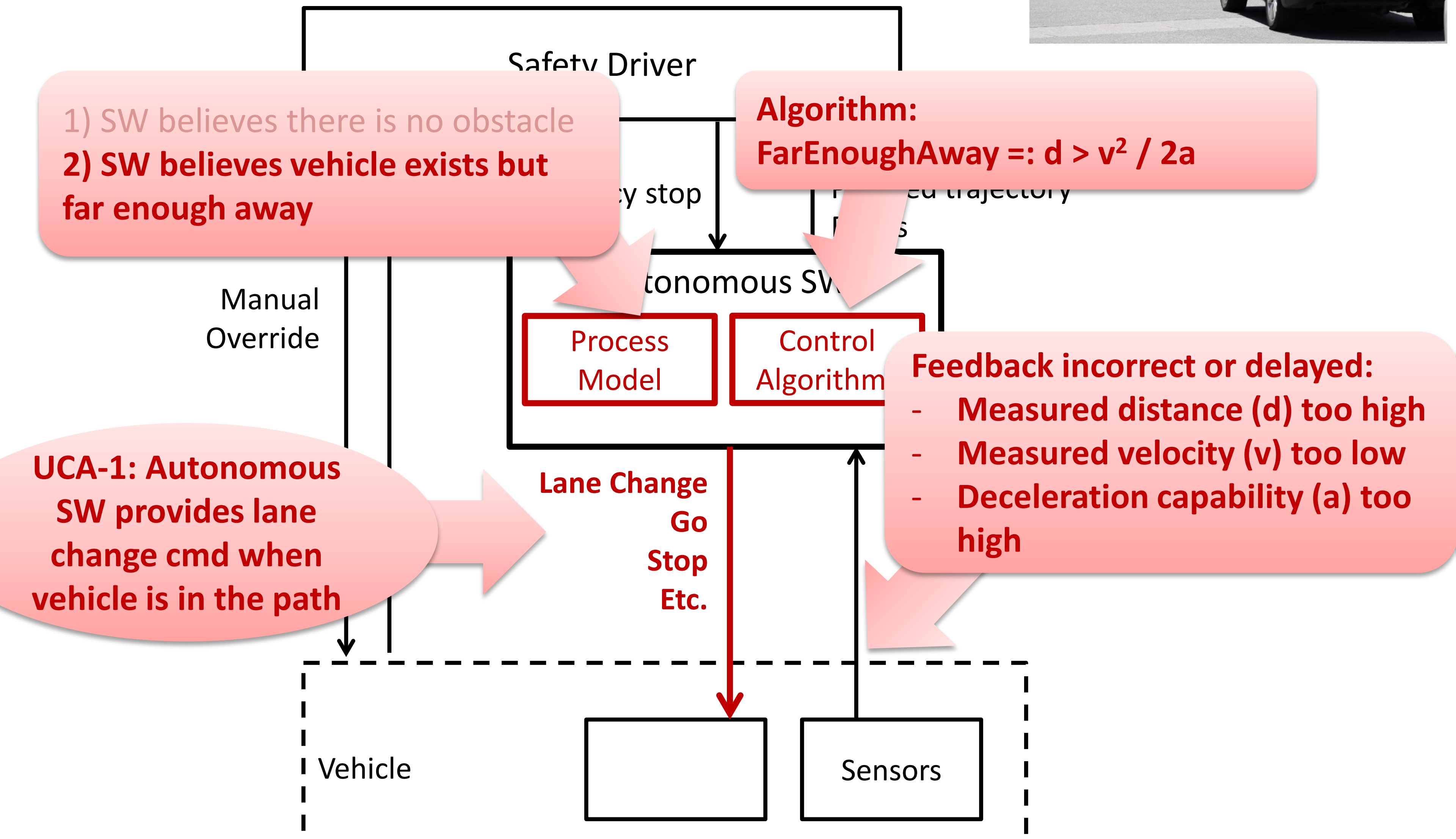
STPA scenario building



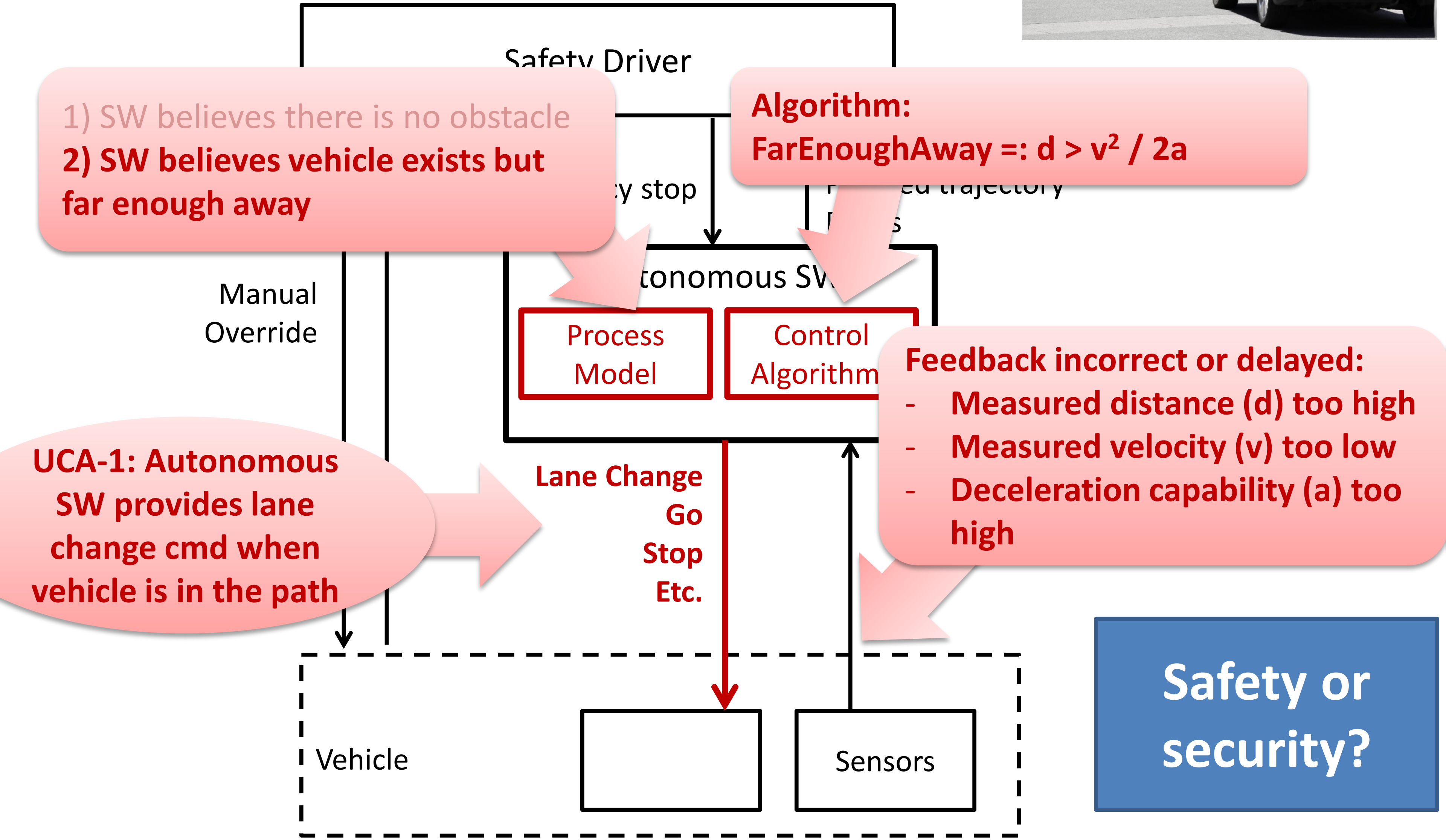
Kinematics Eqn:

$$d = v_i * t + \frac{1}{2} * a * t^2$$

STPA scenario building



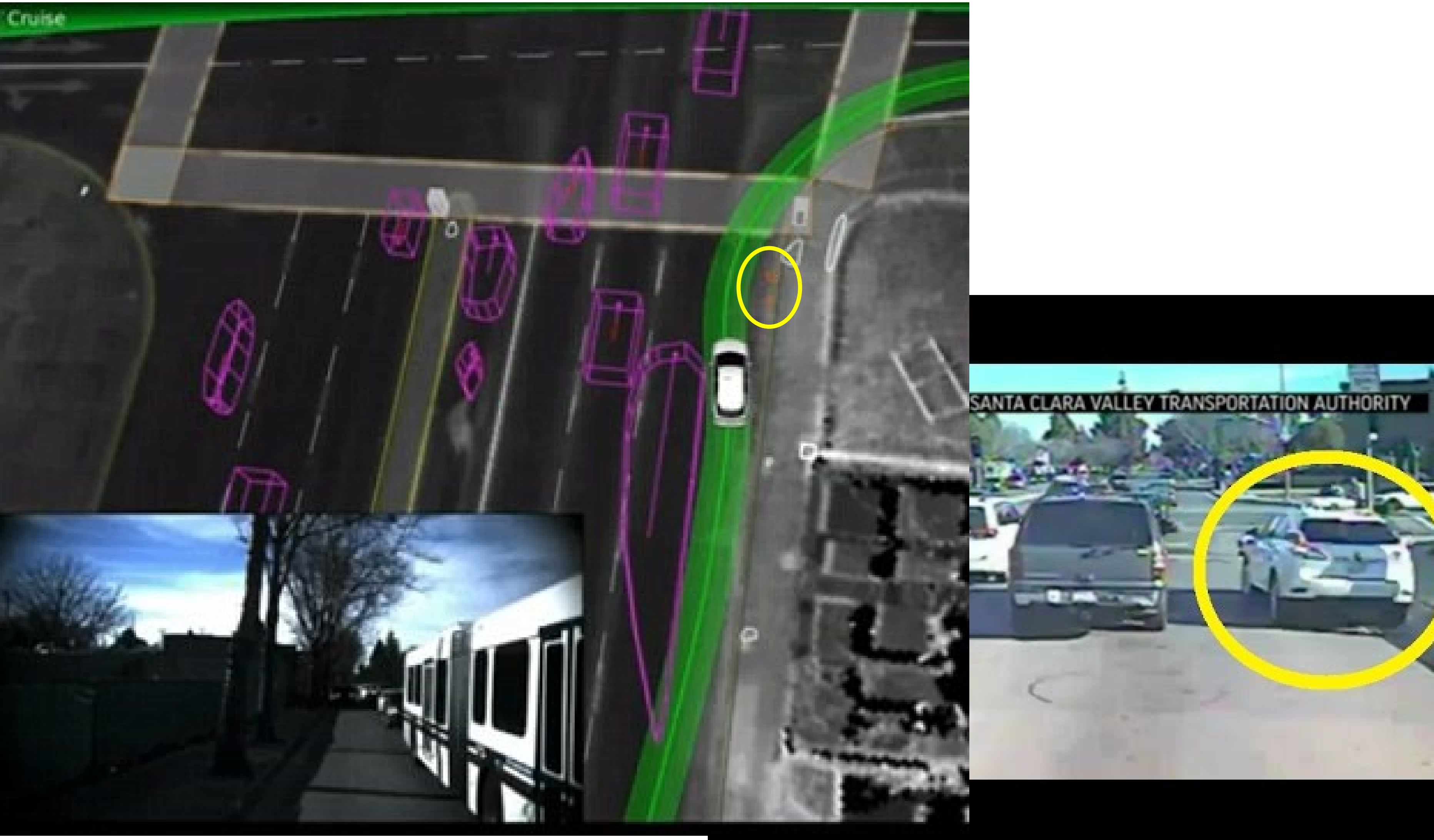
STPA scenario building



Safety or security?

Thomas, 2018

Google Self-Driving Car

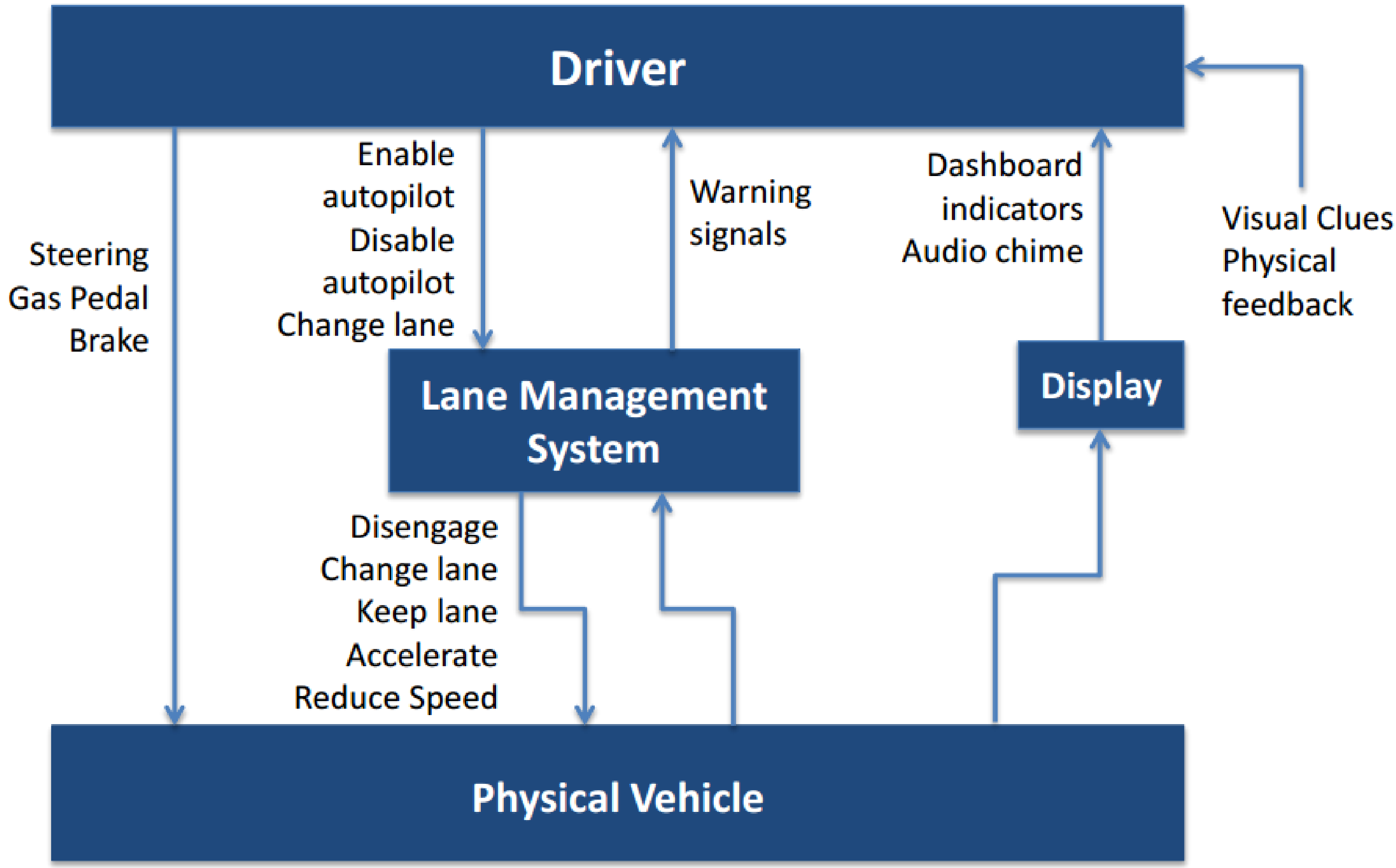


Short STPA example

Tesla Autopilot

Tesla Autopilot example

Tesla Autopilot

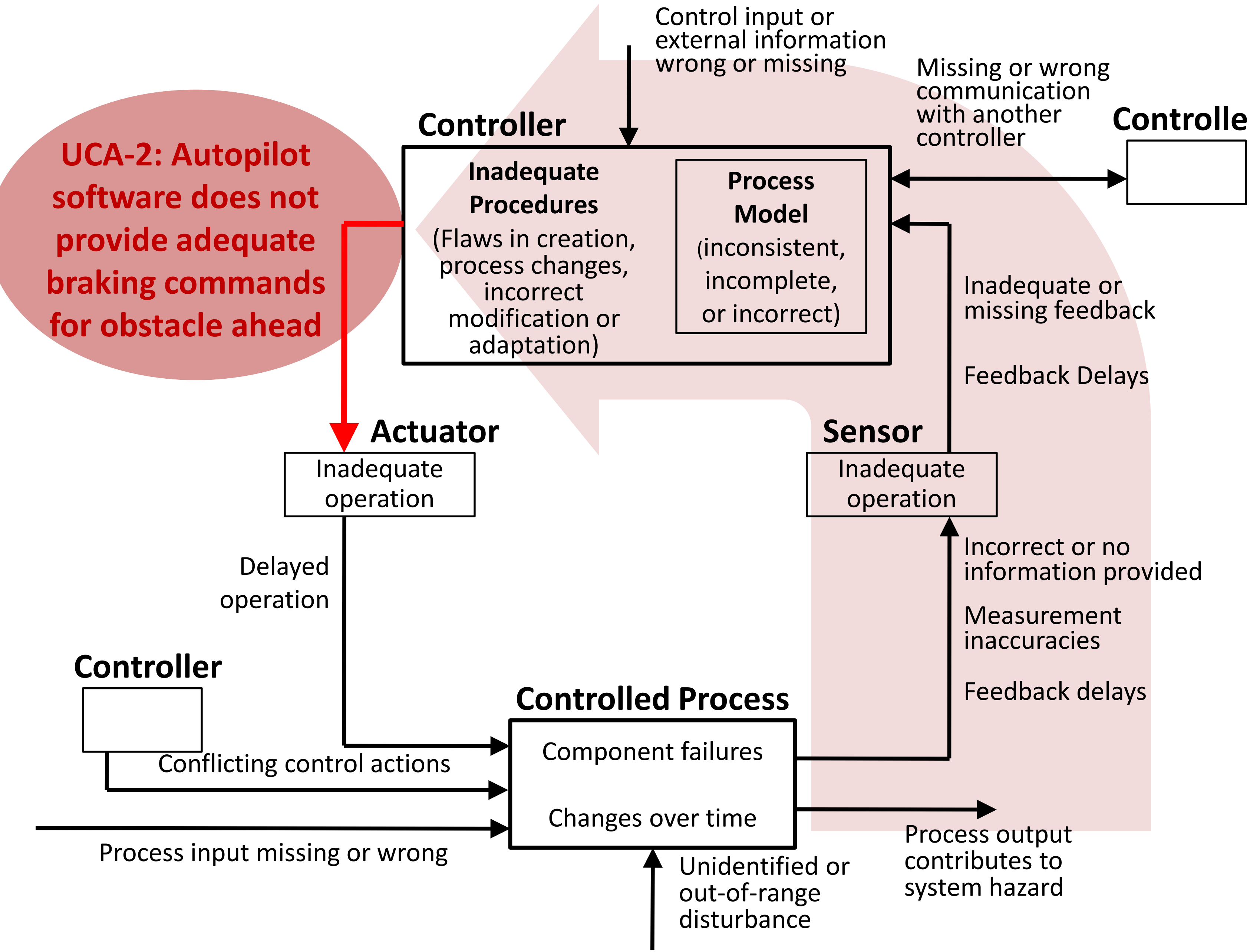


Tesla Autopilot

Controller	Control Action	Not providing causes hazards	Providing causes hazards	Incorrect Timing / Order	Stopped too soon / Applied too long
Driver	**Steering**	-	UCA Driver provides steering can cause hazards if autopilot is changing the lane to the opposite direction	-	-
Driver	**Steering**	UCA Driver does not provide steering to avoid obstacles when autopilot does not react	-	-	-
Auto-Pilot	**Lane changing**	UCA Auto-pilot Not providing lane changing automatically causes hazards	-	-	-
Auto-Pilot	**Reduce Speed**	UCA Auto-pilot does not provide reducing speed can cause hazards if range and range rate of current vehicle is above the limit	-	-	-

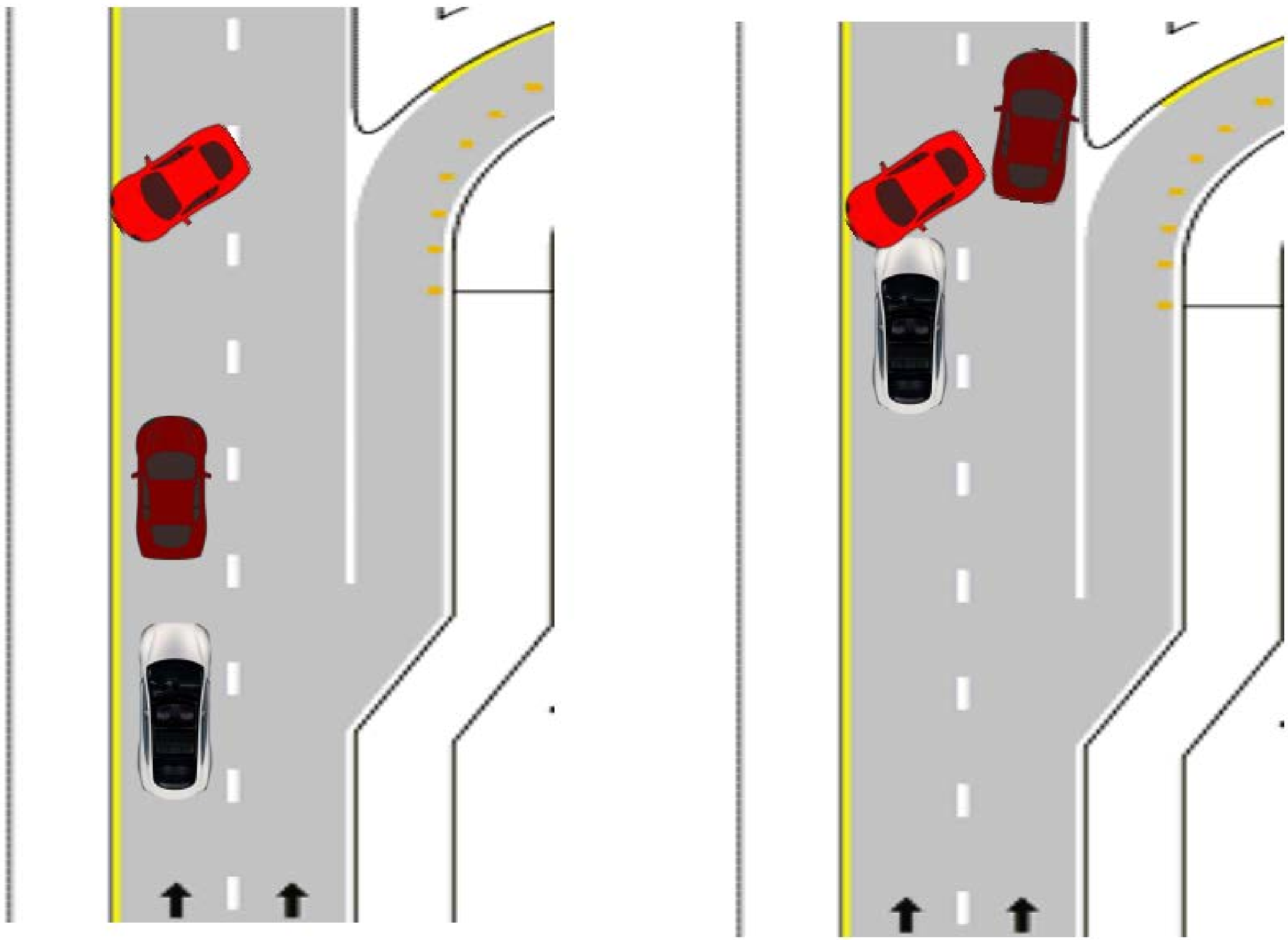
Spring 2016 Student project: Diogo Castilho, Megan France

Step 4: Potential causes of UCAs



Tesla Autopilot

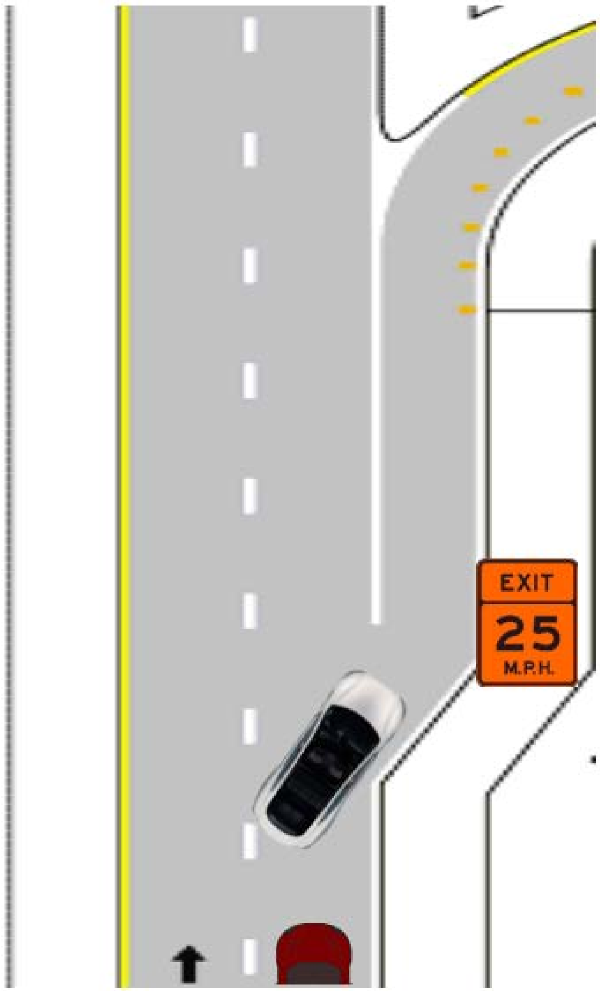
UCA-2: Autopilot does not provide adequate braking commands for obstacle ahead

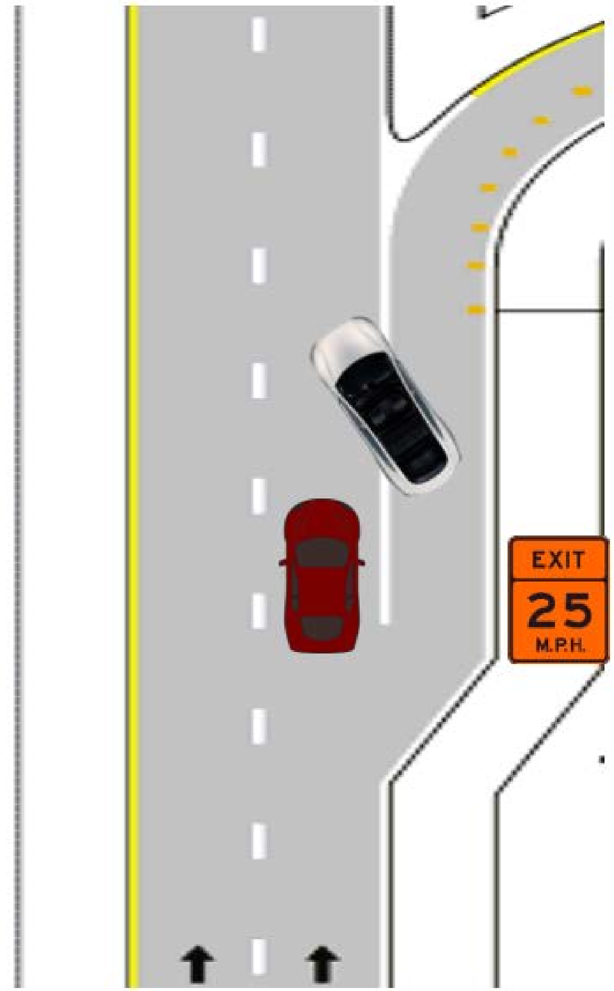


Spring 2016 Student project: Diogo Castilho, Megan France

Tesla Autopilot

UCA-1: Driver provides unsafe steering override commands when autopilot is engaged





Spring 2016 Student project: Diogo Castilho, Megan France

Tesla Autopilot

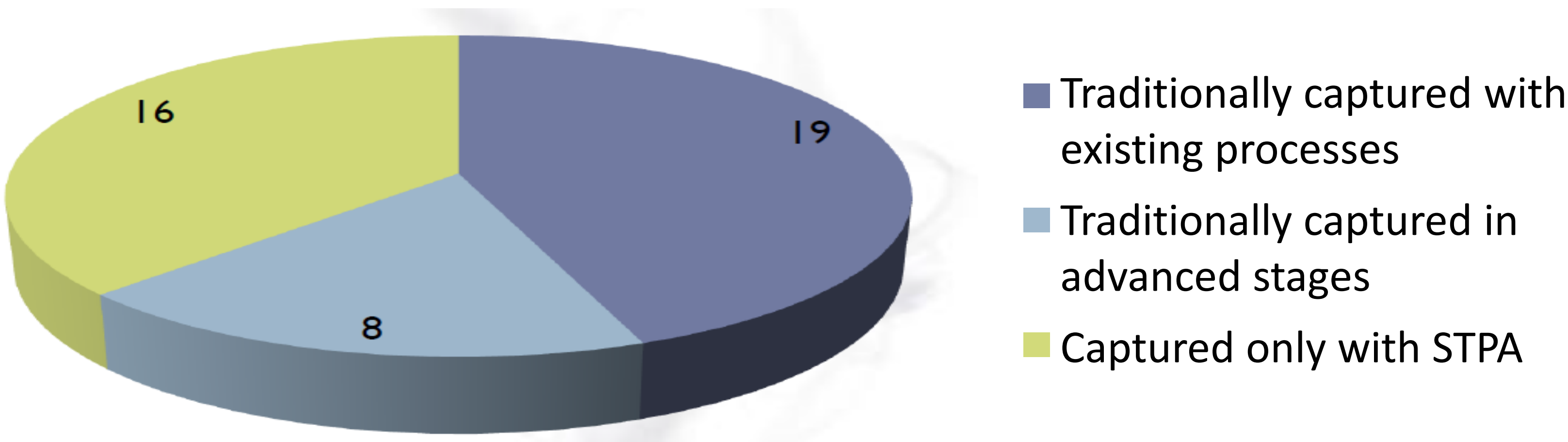
Tesla Autopilot

STPA Adoption

Embraer STPA application

- 2016: Air Management System
  - Identified 200+ safety constraints (requirements) and 700+ design recommendations to eliminate or mitigate hazards (satisfy the safety constraints).

Embraer Aircraft Smoke Control System analysis



Embraer Conclusions

- STPA is a systematic methodology to support **safety assessment and product development** in hazard scenarios identification
- Powerful methodology for highly integrated system based on **software**
- Provide design recommendations for the product development to define the system **requirements**
- **Broadly** applicable: safety, operational, human factors, design etc.
  - Some activities might be used/performed by development engineers
  - Analysis can be done across different abstraction levels
  - Keep good traceability of all results, UCAs, hazards and accidents
- STPA takes in consideration **human-machine interface** during entire system development process
- Improves the design of the **system interfaces**
- Application to aircraft
  - Some overlaps and **terminologies** to be aligned
  - Could be used as a method to assist in **early development and engineering**

GM STPA adoption

Human Machine Interaction and Requirements Definition at General Motors

- **Implementation of ETRS driver interaction device designed with human machine interaction requirements defined by STPA analysis**

Mark Phelan ,
Detroit Free Press
Auto Critic
Published 10:38 p.m. ET July 1, 2017
Updated 4:24 p.m. ET July 2, 2017

http://www.freep.com/story/money/cars/mark-phelan/2017/07/02/gmc-2018-terrain-suv/441807001/
for video and article

Boeing STPA adoption

- Future Vertical Lift (FVL) Mission system and Flight control system
- V-22 STPA requirements generation/validation
- 777X St. Louis factory Automate Ground Vehicle (AGV) system
- 777 Wing body join STPA analysis
- 777 Robotic system STPA
- Auburn Composite FAB center
- Boeing Radiation Effects Lab (BREL)
- Everett Delivery Center (control of aircraft hazardous energy (LOTO))
- BDS Commercial Crew (CCTS) Service Module Hot Fire Test
- Other development and cyber security projects with military customers
- Operational STPA analysis with Cathay Pacific for flight deck development

Summary

- Role of air/ground switch failure states was not fully recognized during the original design process
  - Inputs protecting against inadvertent activation had a common mode failure case
- Changed environment during flight at altitude allows Thrust Control Malfunction (TCM) detection
- STPA analysis identified
  - The inadequate operation of the air-ground switch
  - The TCM protection process output contributing the unsafe control action of inadvertent engine shutdown
  - Relative to the original design work STPA identified approximately 30 additional items that required review including several design changes
- Although a “novel” approach (STPA) applied techniques slightly different from the examples, the ability to explain the approach and understand the results drove consensus for the solutions
- Improved software now in customer’s flight tests with no TCM functional issues. Aircraft level approval for both engines in 2014.

Rolls-Royce

Automotive companies using STAMP/STPA

Other large silicon valley companies*

STPA in Industry Standards

- ISO/PAS 21448: Safety of the Intended Functionality (SOTIF)
  - STPA used assess safety of digital systems
- ASTM WK60748
  - “Standard Guide for Application of STPA to Aircraft”
- SAE AIR6913
  - “Using STPA during Development and Safety Assessment of Civil Aircraft”
- RTCA DO-356A
  - “Airworthiness Security Methods and Considerations”
  - STPA-sec used for cybersecurity of digital systems
- SAE JXXXX
  - “Recommended Practice for STPA in Automotive Safety Critical Systems”
- EPRI/Sandia
  - Recommending to use STPA for digital I&C

STPA Adoption

- Automotive (Ford, GM, Nissan, Toyota, others)
  - Adaptive Cruise Control
  - Engine Stop Start
  - Auto Hold
  - Shift By Wire
  - Keyless Ignition
  - Other automated systems and human-computer interfaces
- Aviation (Boeing, Embraer, FAA, INTA, EASA, etc)
- Medical devices
  - Proton therapy machine, PCA, etc.
- Defense
  - New missile defense system
  - Other systems
- Space
  - NASA Safety-driven design of new JPL outer planets explorer
  - Safety analysis of the JAXA HTV (unmanned cargo spacecraft to ISS)
  - Incorporating risk into early trade studies (NASA Constellation)
  - Orion (Space Shuttle replacement)
- Nuclear
  - NRC, EPRI, Palo Verde, other large nuclear utilities
- Rail
  - Maglev train control systems (Japan Central Railway)

For more information

- Google: “STPA Handbook”
  - How-to guide for practitioners applying STPA
- MIT STAMP Conference (March 25-28, 2019)
- Website: mit.edu/psas
- Training classes
- Send me questions/comments! JThomas4@mit.edu

LinkedIn

Search: “John Thomas MIT”