Artificial Intelligence for Cyber-Physical Systems

Laura Nenzi

Università degli Studi di Trieste
I Semestre 2024

Lecture 1: Course Logistic and Introduction

Who I am

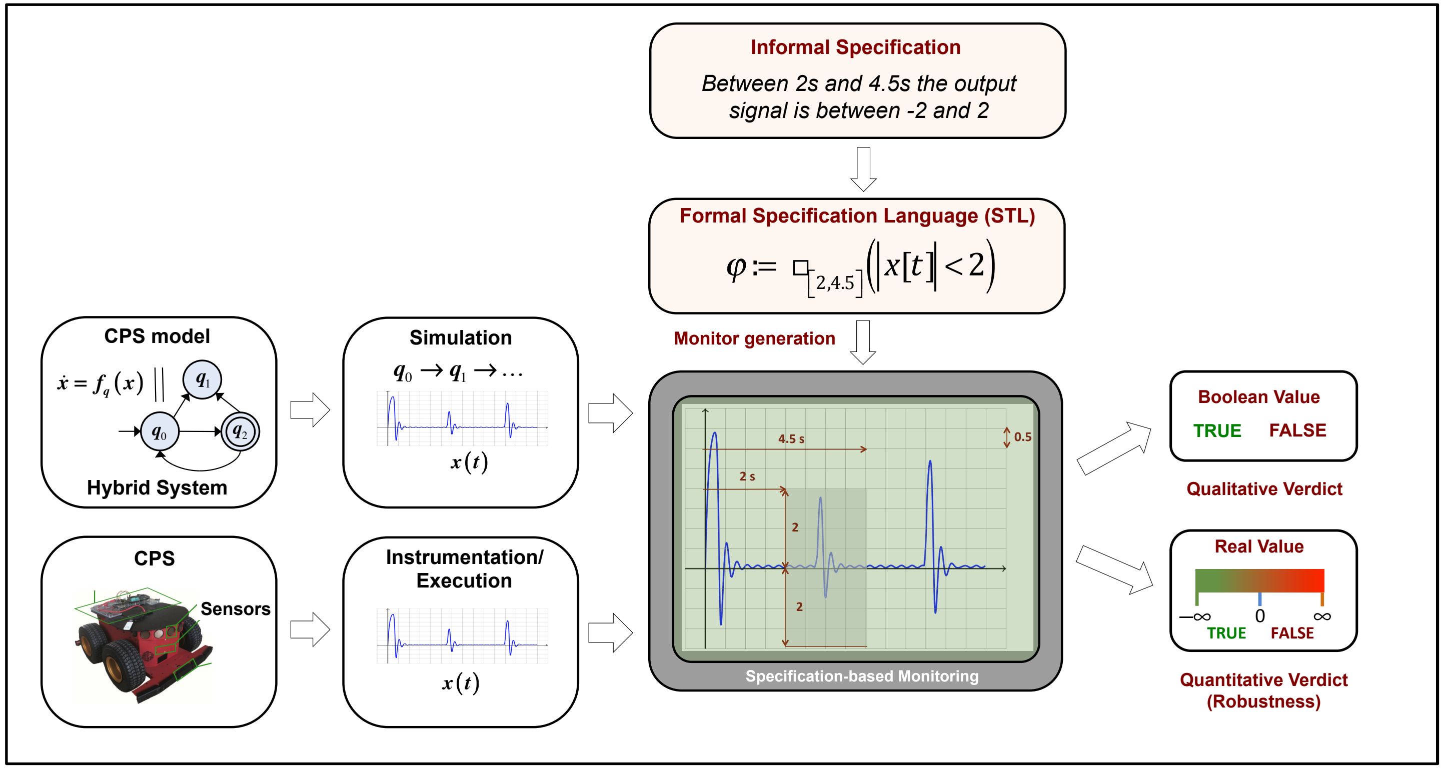
Assistant-professor (tenure-track)
DIA, Università degli Studi di Trieste

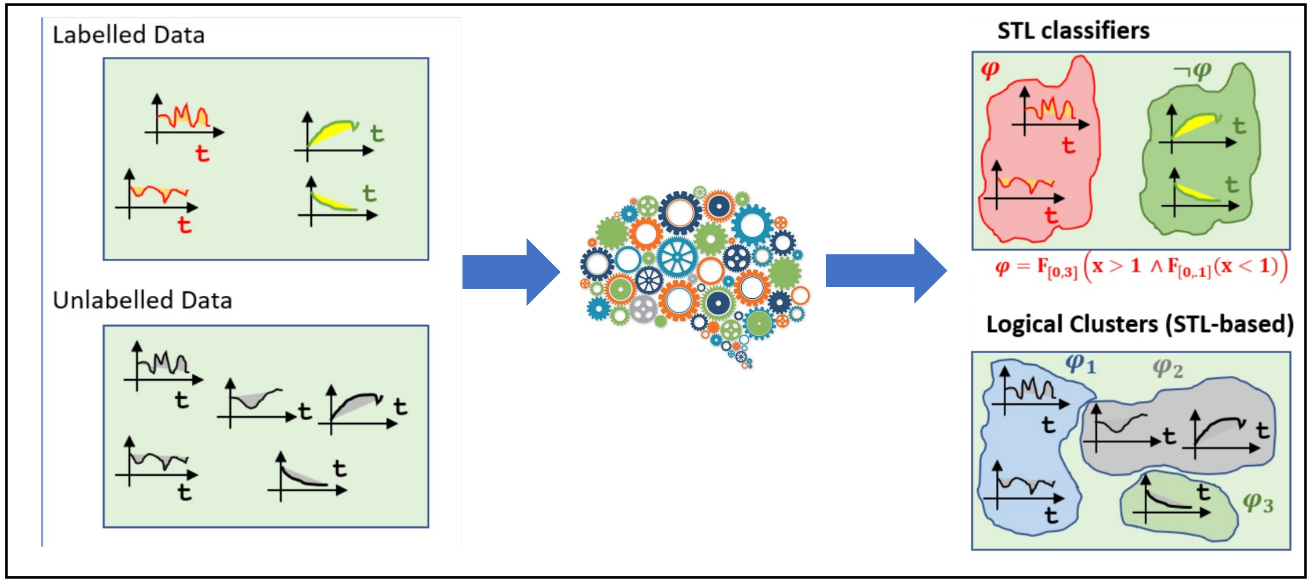
Master in Mathematics, Phd in Computer Science

Office c3 2.55
Office Hours: by appointment
Mail: lnenzi@units.it

Who do I do..

Runtime Verification applied to AI





Explainable AI

Course Logistics

Timing

- Tuesday 14:15-16:00 Edificio C7, Aula A
- Thursday 11:00-13:30 (Break 12:10-12:25), Edificio H2Bis, 3B

- Some seminars

Course Website

Moodle

Teams

What is a Cyber-Physical System?

What is a Cyber-Physical System?

A CPS is a **mechanism** that is controlled or monitored by **computer-based algorithms**, tightly integrated with the Internet and its users.

Physical = physical device or system + environment

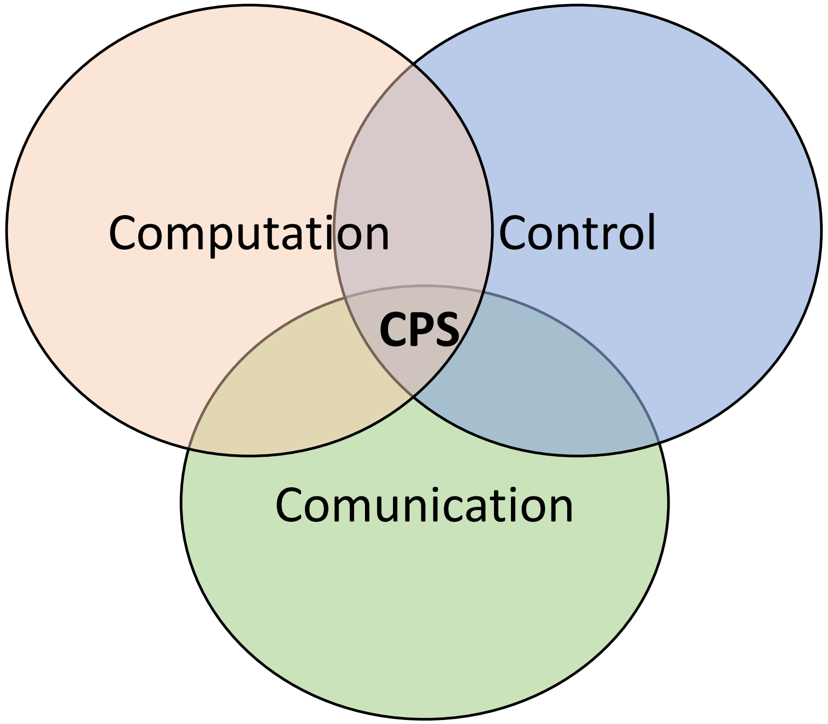
Cyber = computational + communicational

Coined in 2006 by Helen Gill (National Science Foundation)

The important part in CPS is the conjunction/intersection between the computing part and physical dynamics

What is a Cyber-Physical System?

- Systems where the behavior of the physical components is strongly influenced by the software components
- Systems where there the **communication** between the physical component and the software component may be direct or through a network
- Systems in which the primary role played by software is ***control*** (in contrast to passive monitoring)



What is a Cyber-Physical System?

In cyber-physical systems, physical and software components are:

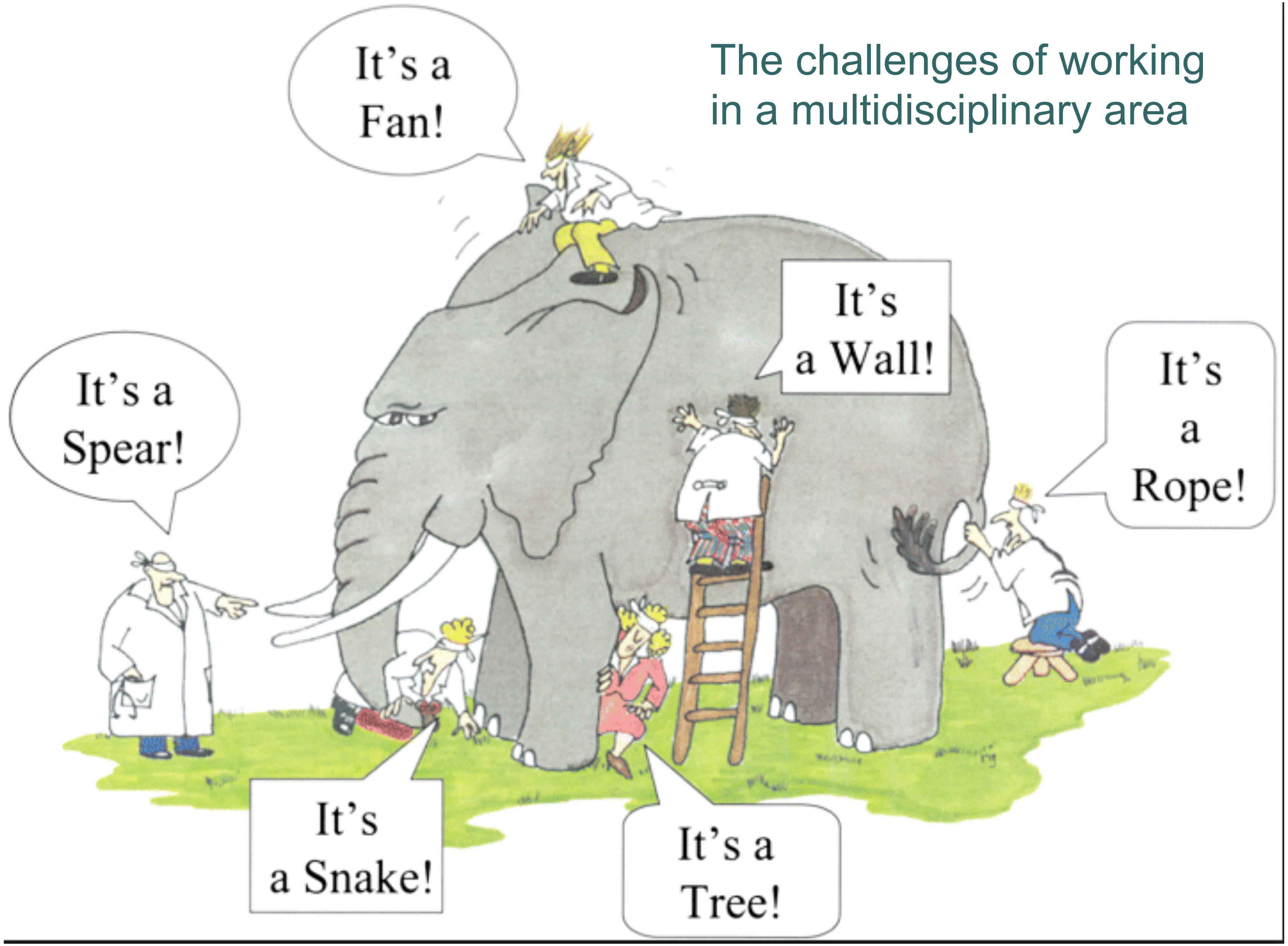
- **deeply intertwined**
- each operating on **different spatial and temporal scale**
- exhibiting **multiple and distinct behavioral modalities**
- interacting with each **other in a lot of ways** that change with context.

CPS combines elements of cybernetics, mechatronics, control theory, process science, embedded systems, distributed control, and more recently communication.

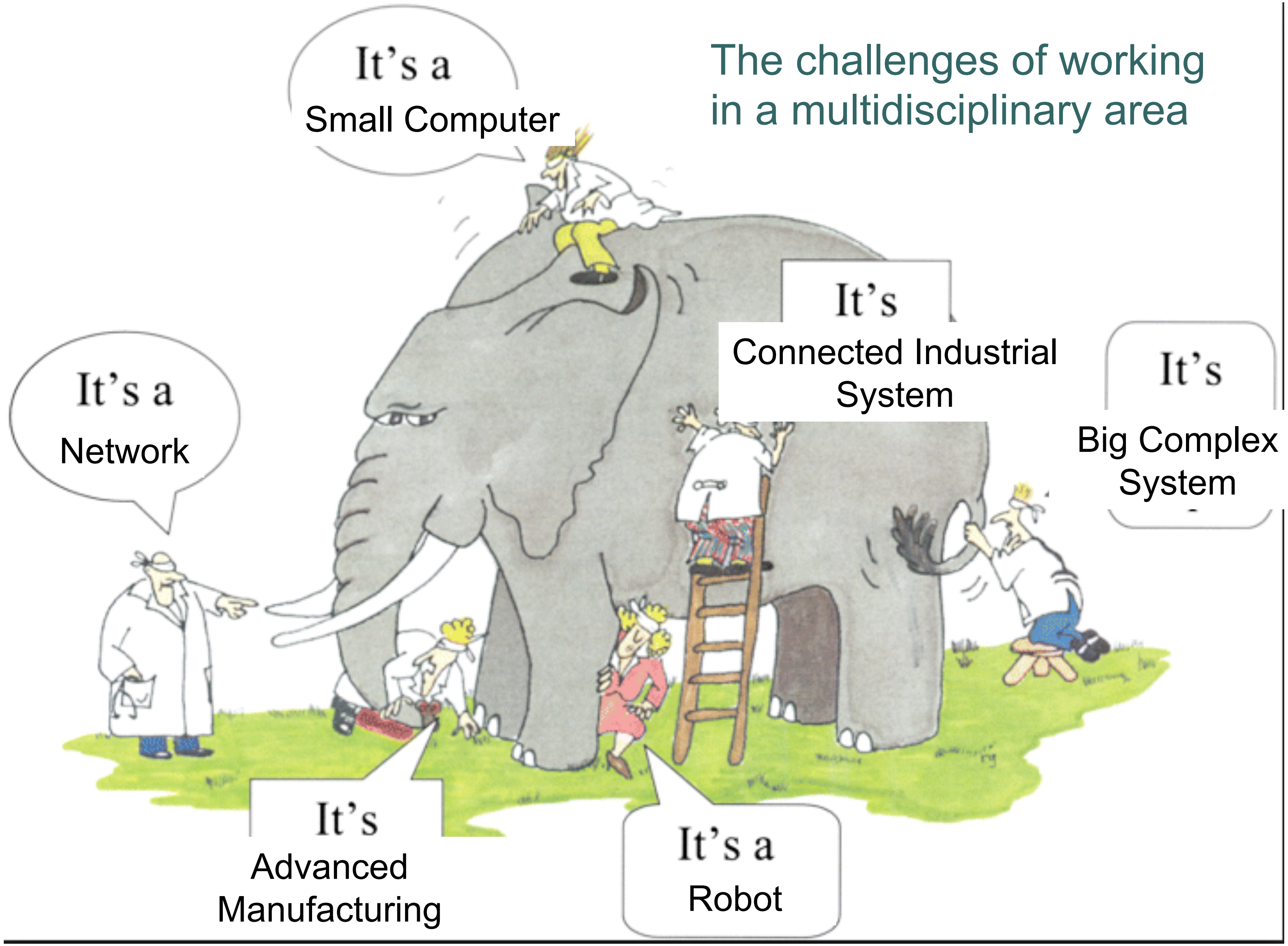
Is the Field of Cyber-Physical Systems New?

- **Hybrid Systems**: are a mathematical abstraction, CPS are real-world objects.
- **Embedded Systems**: are computational system embedded in a physical system. Any CPS contains an embedded system.
- **Real-time Systems**: must respond to external changes within certain timing constraints. Control systems can have or not real-time constraints.
- Other related disciplines: multi-agent system, mechanotronics, control theory, robotics, Internet of Things (IoT), formal methods for specification and verification.

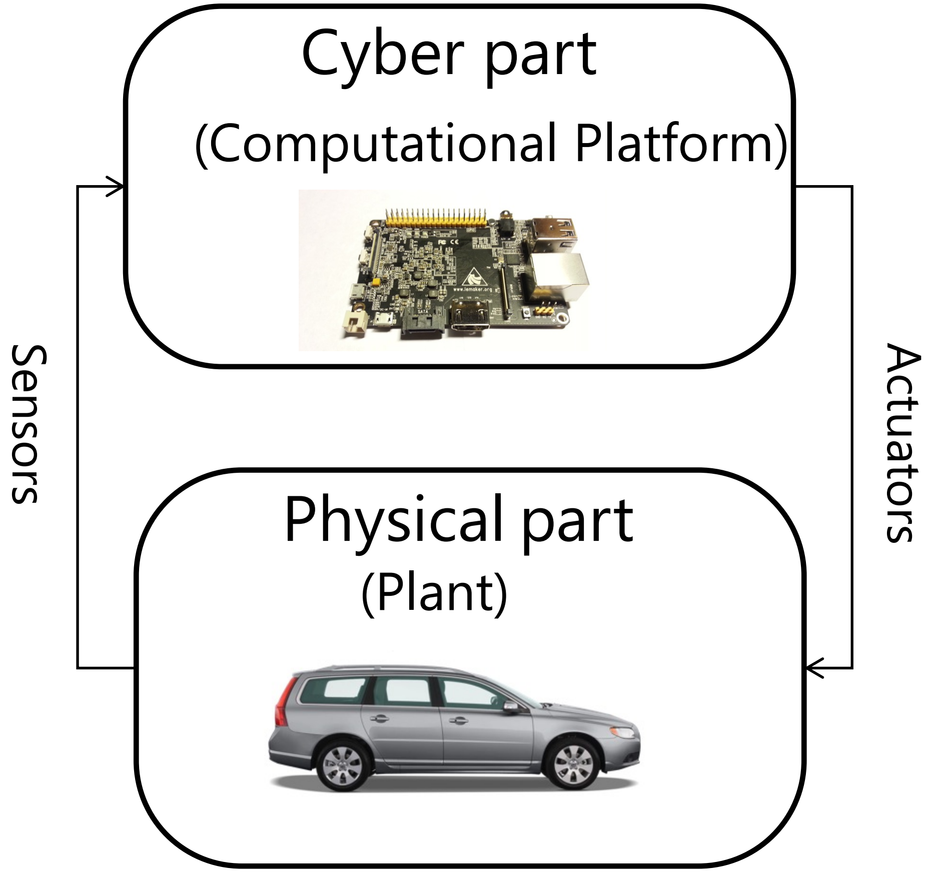
The challenges of working in a multidisciplinary area



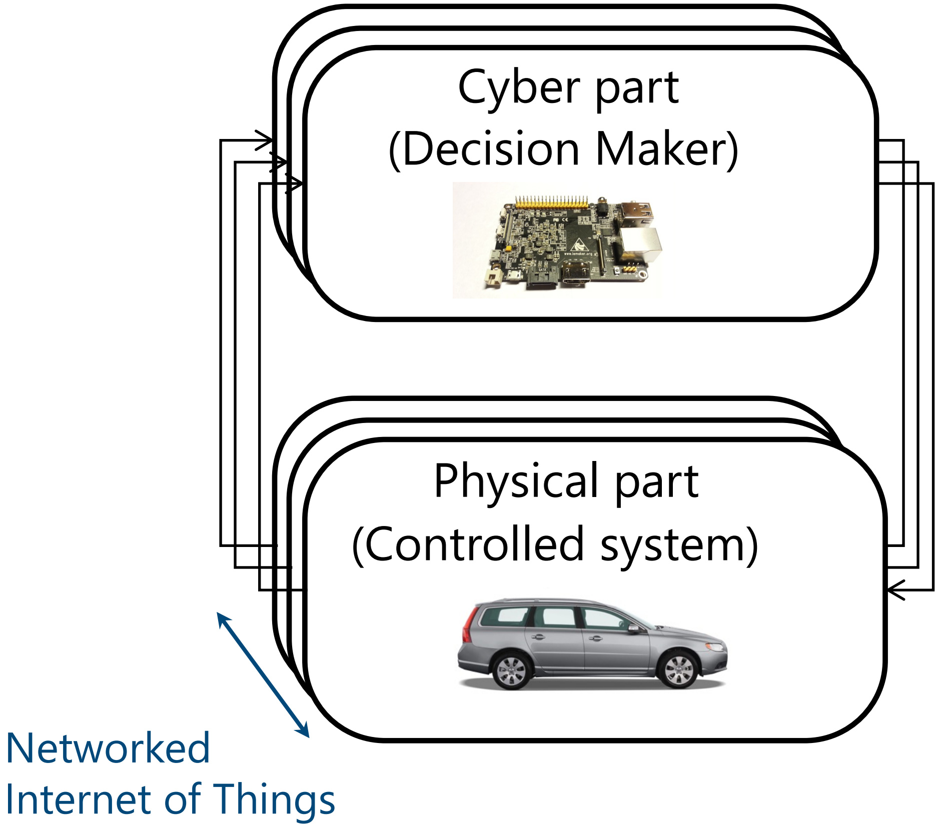
The challenges of working in a multidisciplinary area



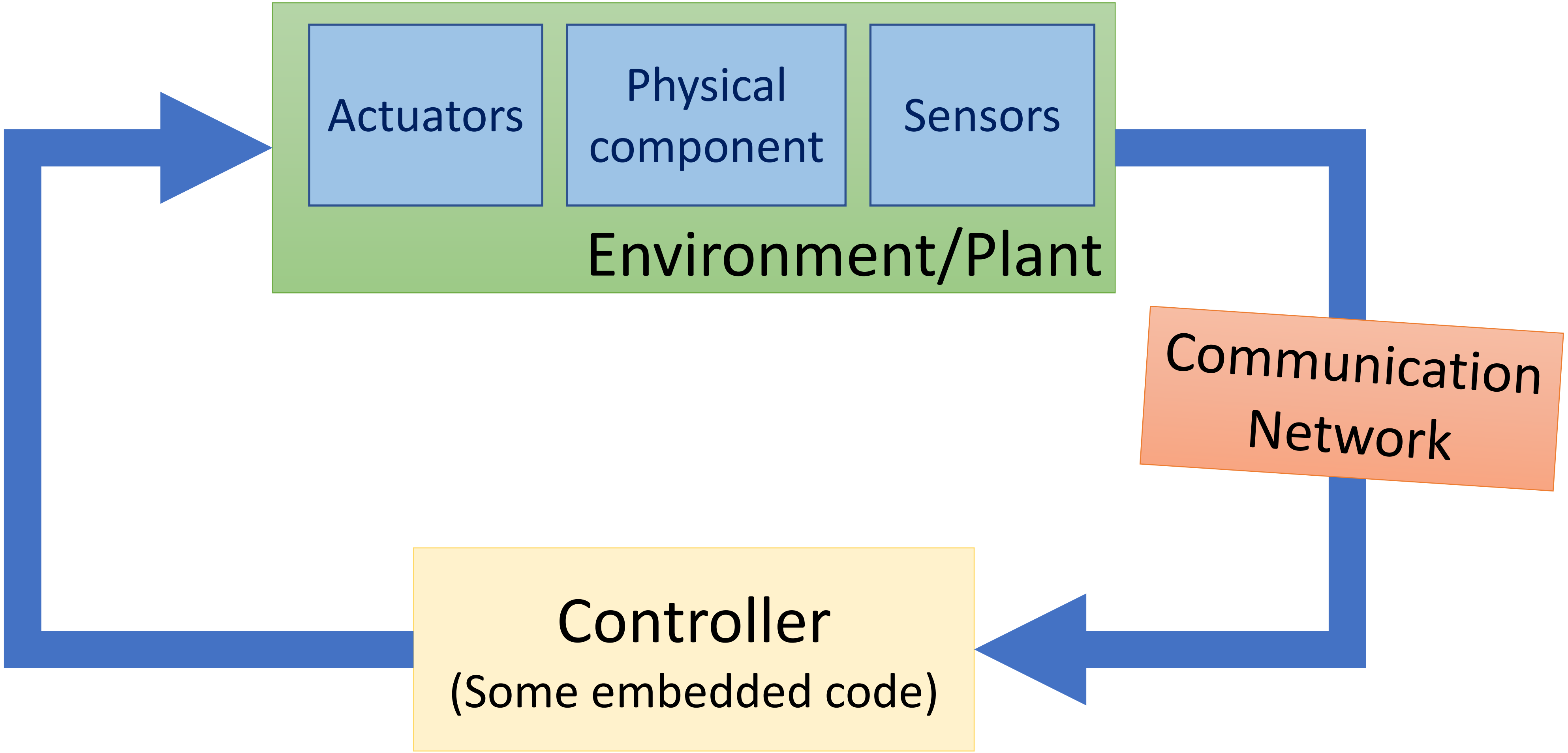
Example Structure of a CPS



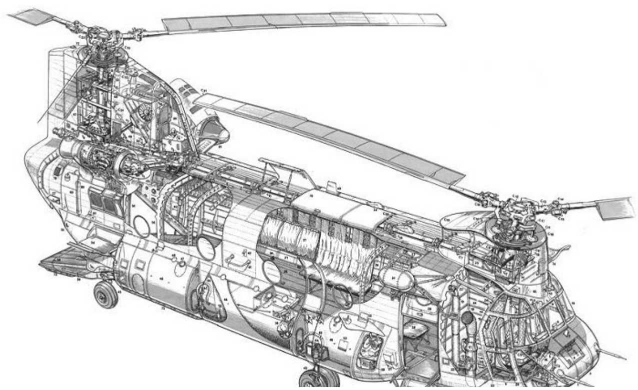
Example Structure of a CPS

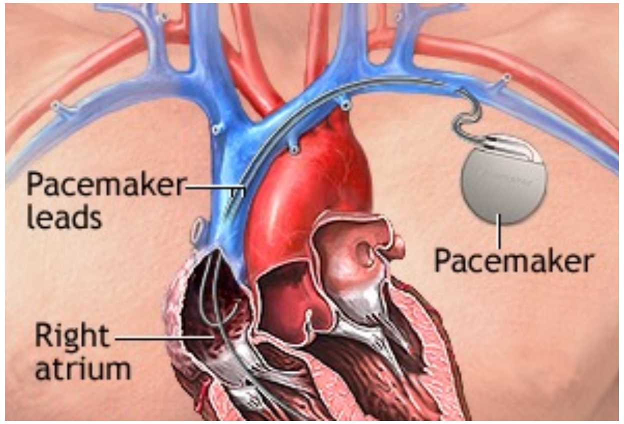


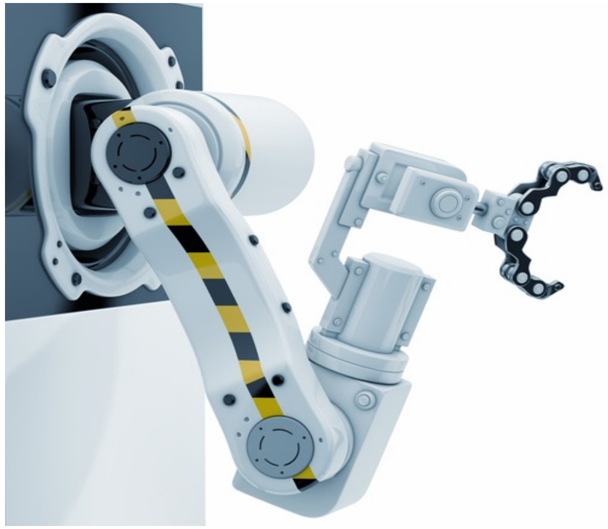
Our view of a CPS



Examples

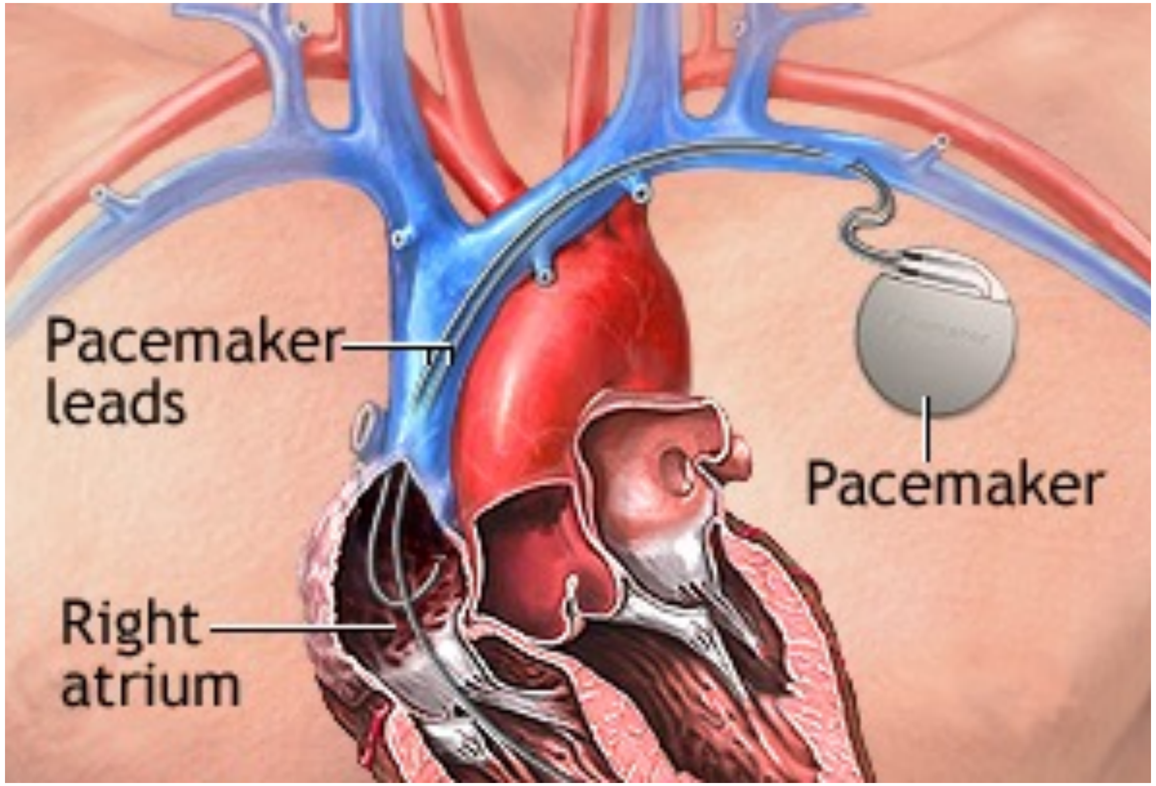


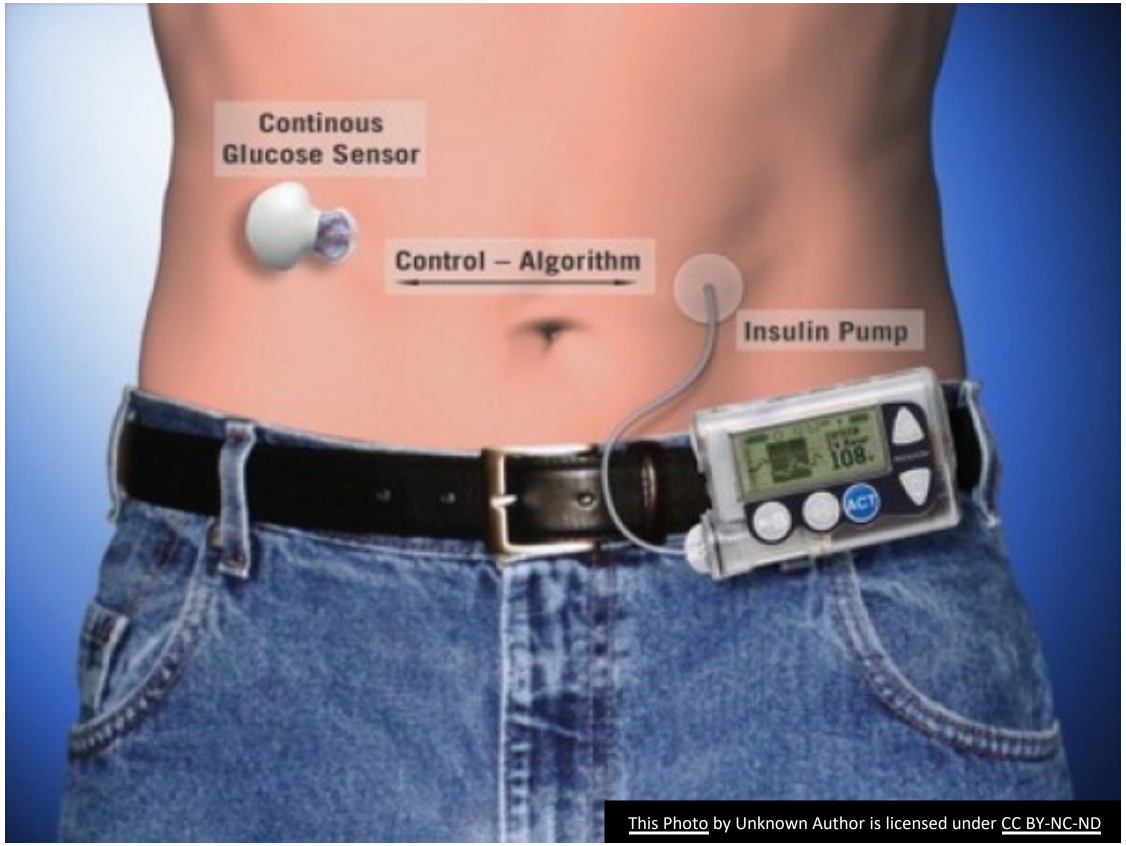




[All images from Google image search]

Application domains: Medical Device





Application domains: Transportation

Application domains: Energy

And many other applications…

- Robotics
- Critical Infrastructures
- Industrial Control
- Manufactering
- Agricolture

Autonomous CPS

- Autonomous: without the need for human intervention or control
- Autonomous CPS = CPS with no human operator!
- Semi-autonomous: CPS with autonomy under specific conditions but requiring a human operator otherwise.
- Today, several CPS examples are semi-autonomous, and getting to fully autonomous

Are we safe ?

The Washington Post
Democracy Dies in Darkness

Tech | Help Desk | Artificial Intelligence | Internet Culture | Space | Tech Policy

17 fatalities, 736 crashes: The shocking toll of Tesla's Autopilot

Tesla's driver-assistance system, known as Autopilot, has been involved in far more crashes than previously reporte

By Faiz Siddiqui and Jeremy B. Merrill
June 10, 2023 at 7:00 a.m. EDT

NBC BAY AREA | LOCAL | WEATHER | INVESTIGATIONS | VIDEO | SPORTS | NEWSLETTERS

Hurriquake | SJC luggage theft | Former SF teacher lawsuit | How to help Maui

SAN FRANCISCO

Cruise faces backlash after self-driving car appears to block crews responding to SF's Mission District shooting

By NBC Bay Area staff • Published June 10, 2023 • Updated on June 10, 2023 at 6:05 pm

Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk

The automated car lacked "the capability to classify an object as a pedestrian unless that object was near a crosswalk," an NTSB report said.

Milbank Q. 2017 Sep; 95(3): 535–553.
Published online 2017 Sep 12. doi: 10.1111/1468-0009.12278
PMCID: PMC5594275
PMID: 28895231

Software-Related Recalls of Health Information Technology and Other Medical Devices: Implications for FDA Regulation of Digital Health

JAY G. RONQUILLO[1,2] and DIANA M. ZUCKERMAN[2]

MOTORTREND | News | Reviews | Buyer's Guide | Watch | MotorTrend+ | The Future | FREE TV | SEARCH

NHTSA Finds Teslas Deactivated Autopilot Seconds Before Crashes

The finding is raising more questions than answers, but don't jump to any conclusions yet.

Alexander Stoklosa - Writer; Getty Images - Photographer | Jun 15, 2022

Bell APT Autonomous Cargo Drone Crashes in Texas

A Bell APT 70 UAV cargo drone being developed for civil and military missions crashed during flight testing last week in Texas.

ACPS are safety-critical, and/or mission-critical with huge implications on human health, well-being, economy, etc.

Safety-Critical Systems

Systems where failure could lead to severe consequences, such as loss of life, injury, environmental damage, or property destruction.

Examples: Aircraft control systems, medical devices, nuclear reactors

Mission-Critical Systems

Systems essential for the successful completion of a mission or operation. Failure leads to significant financial or operational impact but not necessarily loss of life.

Examples: Banking systems, communication networks, logistics systems.

Some tragic accidents

Tesla driver dies in first fatal crash while using autopilot mode

The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky

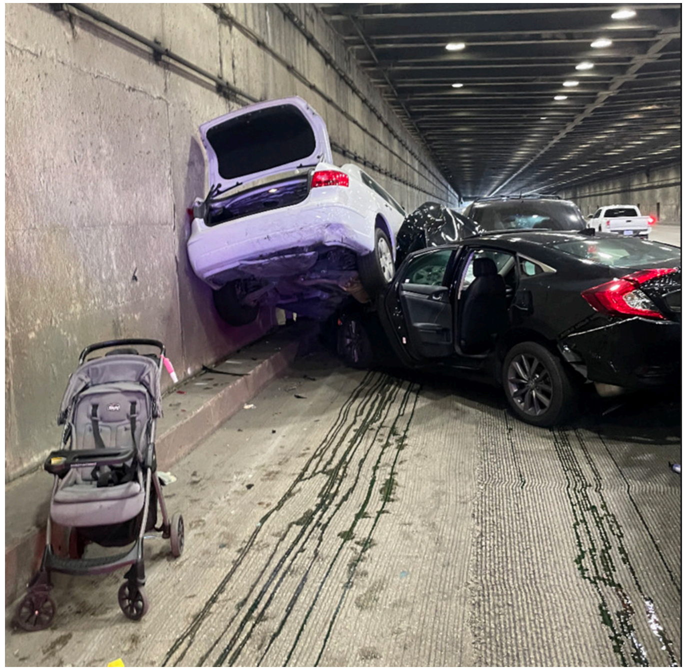
The first known death caused by a self-driving car was disclosed by Tesla Motors on Thursday, a development that is sure to cause consumers to second-guess the trust they put in the booming autonomous vehicle industry.

The 7 May accident occurred in Williston, Florida, after the driver, Joshua Brown, 40, of Ohio put his Model S into Tesla's autopilot mode, which is able to control the car during highway driving.

Against a bright spring sky, the car's sensors system failed to distinguish a large white 18-wheel truck and trailer crossing the highway, Tesla said. The car attempted to drive full speed under the trailer, "with the bottom of the trailer impacting the windshield of the Model S", Tesla said in a blogpost.

https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot

EXCLUSIVE: SURVEILLANCE FOOTAGE OF TESLA CRASH ON SF'S BAY BRIDGE HOURS AFTER ELON MUSK ANNOUNCES "SELF-DRIVING" FEATURE



Highway surveillance footage from November 24 shows a Tesla Model S vehicle changing lanes and then abruptly braking in the far-left lane of the San Francisco Bay Bridge, resulting in an eight-vehicle crash.

https://theintercept.com/2023/01/10/tesla-crash-footage-autopilot/

Are we safe ?

IOT SECURITY

Medtronic Recalls Medical Devices Due to Security Risks That Can Lead to Injury, Death

Medical device maker Medtronic is recalling remote controllers used with some of its insulin pumps due to cybersecurity risks that could lead to injury and even death.

By Eduard Kovacs
October 7, 2021

Medical device maker Medtronic is recalling remote controllers used with some of its insulin pumps due to cybersecurity risks that could lead to injury and even death.

The recall is related to a series of vulnerabilities discovered by a team of cybersecurity researchers in 2018. In June 2019, the U.S. Food and Drug Administration (FDA) and Medtronic informed the public of a recall of MiniMed 508 and Paradigm series insulin pumps due to vulnerabilities that could allow an attacker to remotely hack the devices.

The FDA and Medtronic said that some affected users — whose devices were under warranty — were notified as early as August 2018.

That recall is now being expanded by Medtronic to the optional remote controllers associated with the affected insulin pumps. Users of these devices have been sent updated instructions, including for stopping the use of impacted controllers and returning them.

The FDA said more than 31,000 devices have been recalled in the United States. The agency and Medtronic noted that the affected MiniMed MMT-500 and MMT-503 controllers are no longer manufactured or distributed.

MiniMed™ remote controller
MMT-500

The model # is behind the remote under the barcode

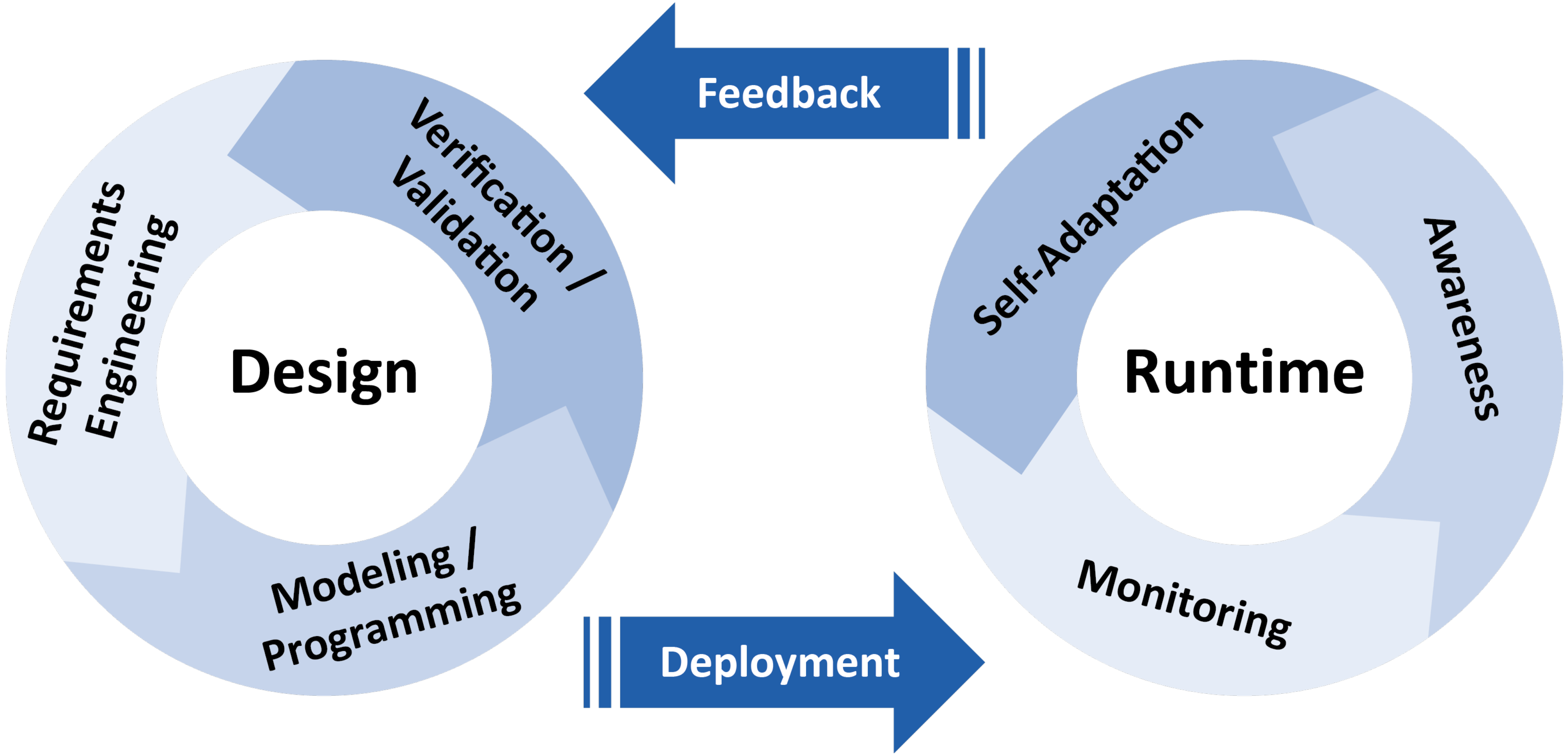
On Monday 19th of September, the US National Highway Safety Administration (NHTSA) issued a safety recall for Tesla vehicles addressing an issue with the automatic window reversal function that can pinch the fingers of the users (Recall #: 22V-702 / PDF below).

The frameless windows in Teslas work in a unique fashion, as soon as a person opens the door the glass slightly slides down automatically to avoid hitting the body frame. As the door closes, the window slides up automatically to seal the cabin of the car.

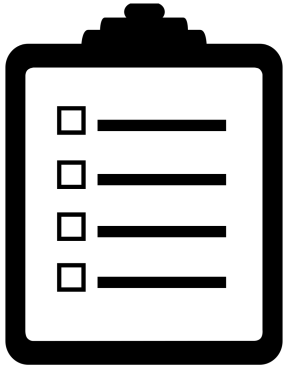
NHTSA is concerned about people's fingers getting pinched when the windows of a Tesla vehicle drop down automatically. Around 1.1 million (1,096,762) Tesla vehicles are affected by this issue and have been recalled by the automaker.

https://www.teslaoracle.com/2022/09/24/tesla-fix-pinching-finger-window-nhtsa-recall-fmvss-118-22v-702/

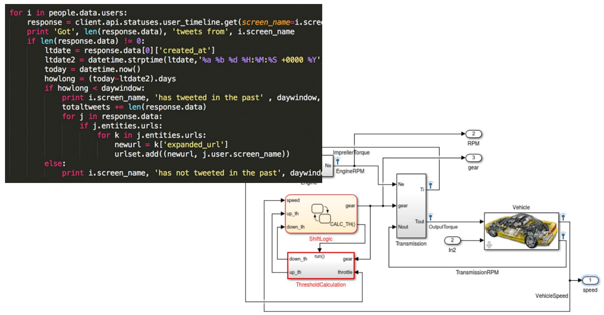
Rigorous Engineering of CPS



Model-Based Design (MBD)



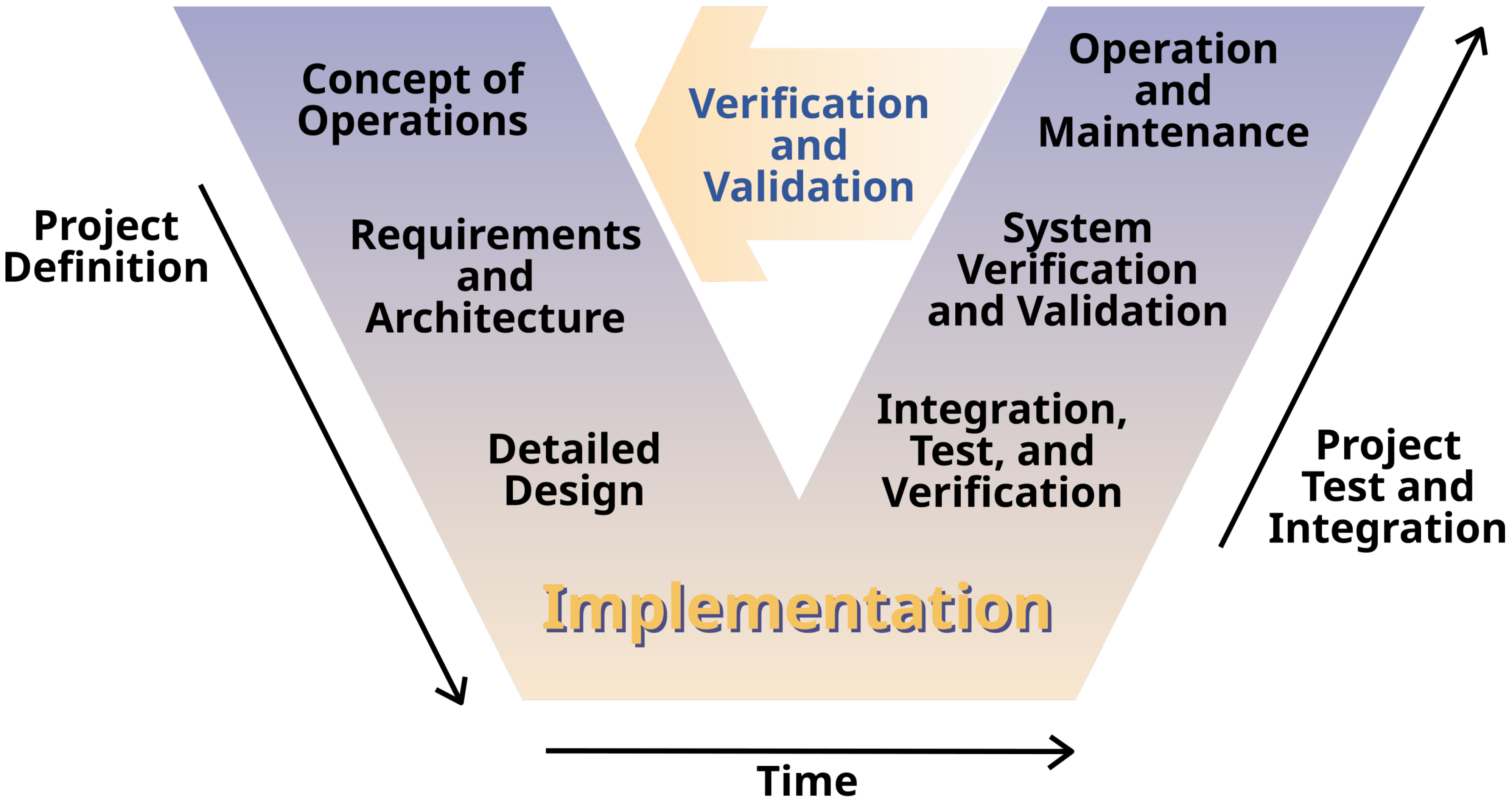
Requirement Definition



Design/Implementation

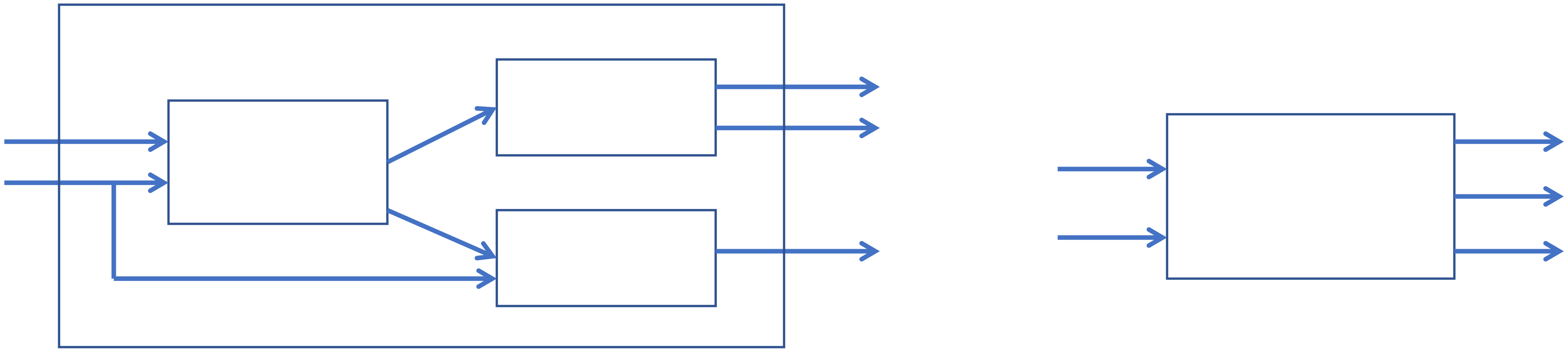
Verification

The V-Model


Concept of Operations
Requirements and Architecture
Detailed Design
Verification and Validation
Operation and Maintenance
System Verification and Validation
Integration, Test, and Verification
Implementation
Project Definition
Project Test and Integration
Time

Model-based Design Approach

- MBD when used for designing embedded software[1] has 4 main steps
  1. Model the physical components/environment (also known as a plant model)
  2. Analyze the plant, and synthesize/design the control-software at a high-level
  3. Co-Simulate the plant and control-software
  4. Automatically generate code from the control-software model for deployment
- MBD languages are often visual and block-diagram based, e.g. Simulink



- Why?

[1] Nicolescu, Gabriela; Mosterman, Pieter J., eds. (2010). Model-Based Design for Embedded Systems. Computational Analysis, Synthesis, and Design of Dynamic Systems. 1. Boca Raton: CRC Press.

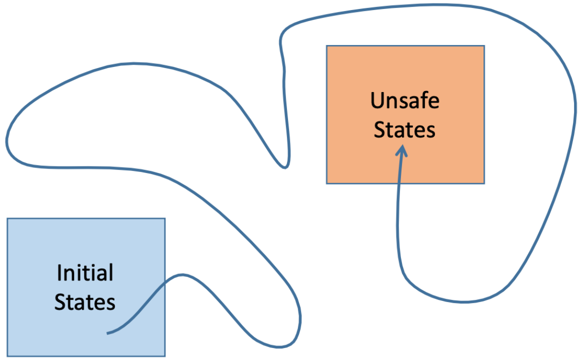
Requirement Definition

- E.g. Functional Requirements:
  - "The system should control vehicle speed based on driver input."
  - "The drone should reach position X."
- E.g. Non-Functional Requirements:
  - "The system should process sensor inputs in under 5 milliseconds."
  - "The system must be available 99.9% of the time, 24/7."
- To be precise we need to be FORMAL

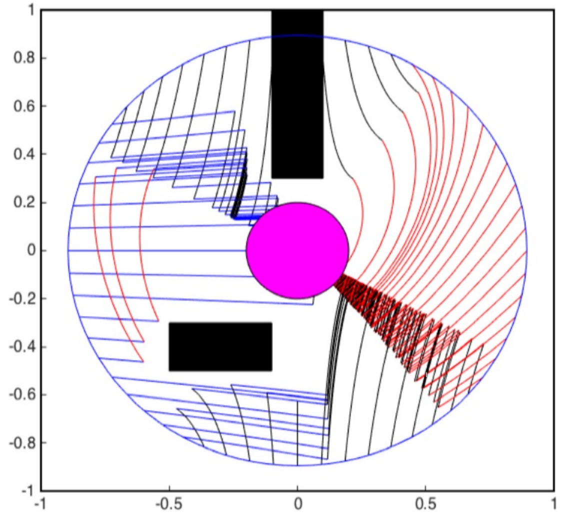
Formal Reasoning

- Writing requirement clearly, precisely, and unambiguously
- Often using structured or mathematical language.
- This ensures that the requirements are testable, verifiable, and consistent.
- Formality reduces misinterpretation, facilitates verification, and allows for automated analysis, making sure the system behaves as expected based on well-defined criteria.

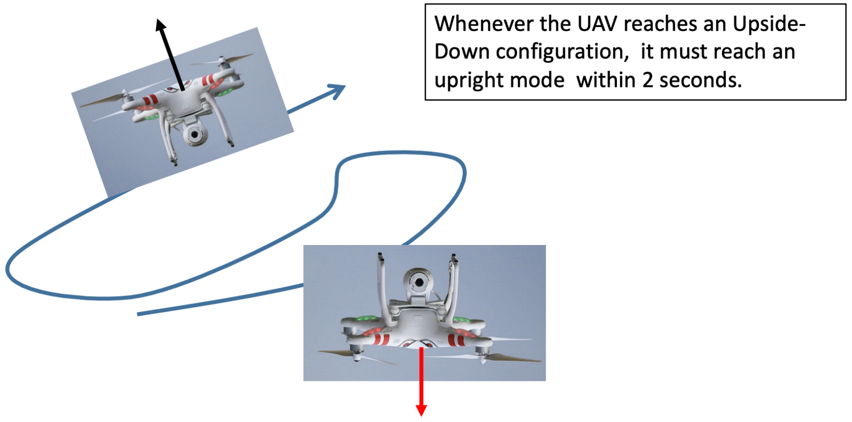
Reachability



Stability



Real-Time Temporal Properties



Formal Reasoning

Formal Methods

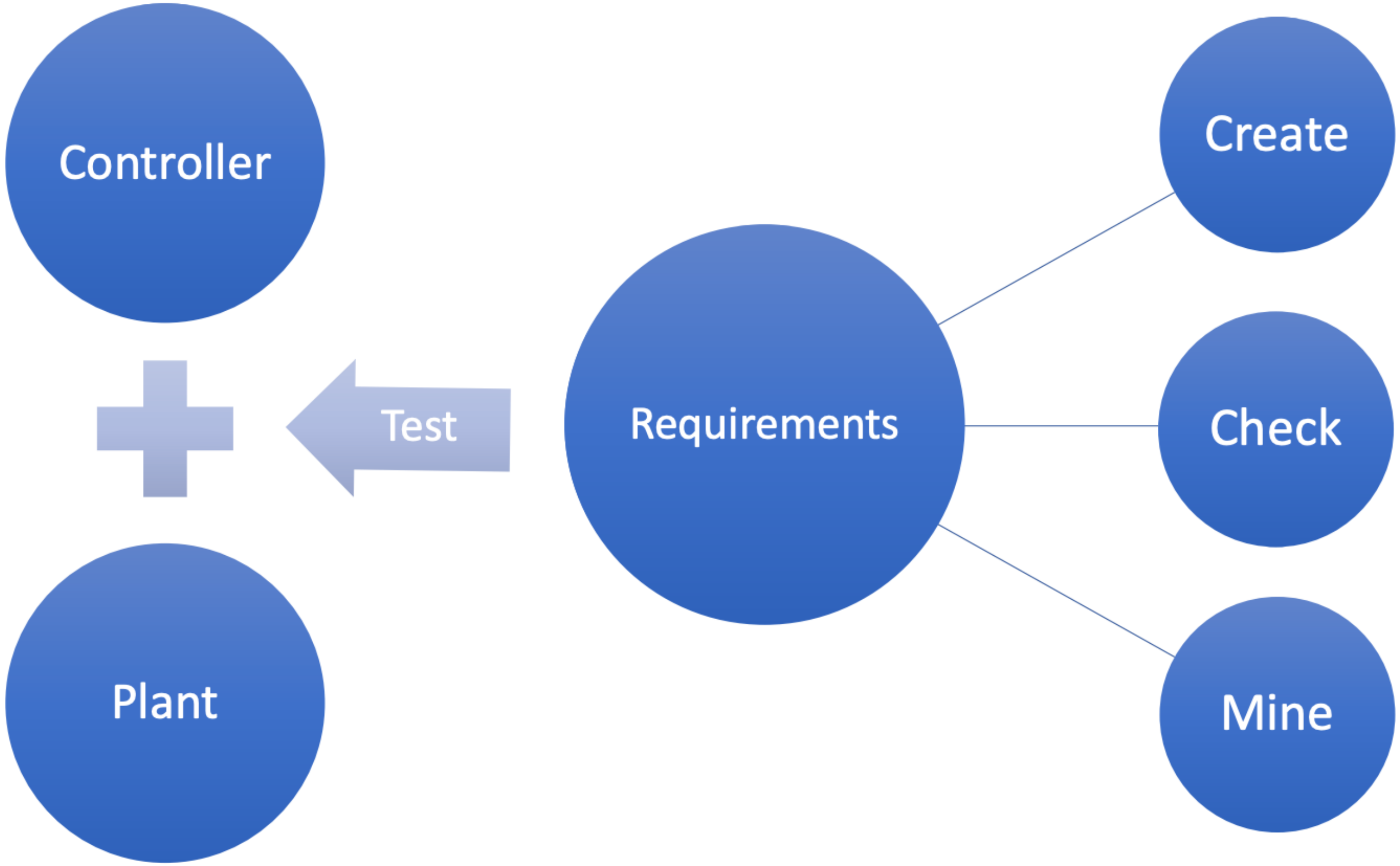
Mathematical, Algorithmic techniques for modeling, design, analysis

– **Specification**: WHAT the system must/must not do

– **Verification**: WHY it meets the spec (or not)

– **Synthesis**: HOW it meets the spec (correct-by-construction design)

Requirement-Driven Design



Requirements formally capture what it means for a system to operate correctly in its operating environment

Requirement-Driven Design

Exhaustive verification of CPS is increasingly intractable:

- Openness, environmental change
- Uncertainty, spatial distribution
- Emergent behaviors resulting from the local interactions are not predictable by the analysis of system’s individual parts
- Classic state-space explosion problem

How to ensure safety-critical requirements in CPS ?

Course Objectives

- Gain familiarity with CPS topics

    Challenge Problems/Case studies

- Model-Based Software Development Paradigm for CPS

    Developing models for physical components + software (+ communication)

- Write formal requirements for CPS models and perform testing

- Basics of simulation-based testing and falsification

Course Overview

1. Intro to CPS and application domains with example (e.g. Medical CPS, energy CPS, transportation CPS)

2. **Modeling formalism**: Synchronous & Asynchronous Models, Timed Models, Dynamical System Models, Hybrid Models, Basics of Control

3. **Safety:** temporal logic and automata, Monitoring Test Generation, Falsification

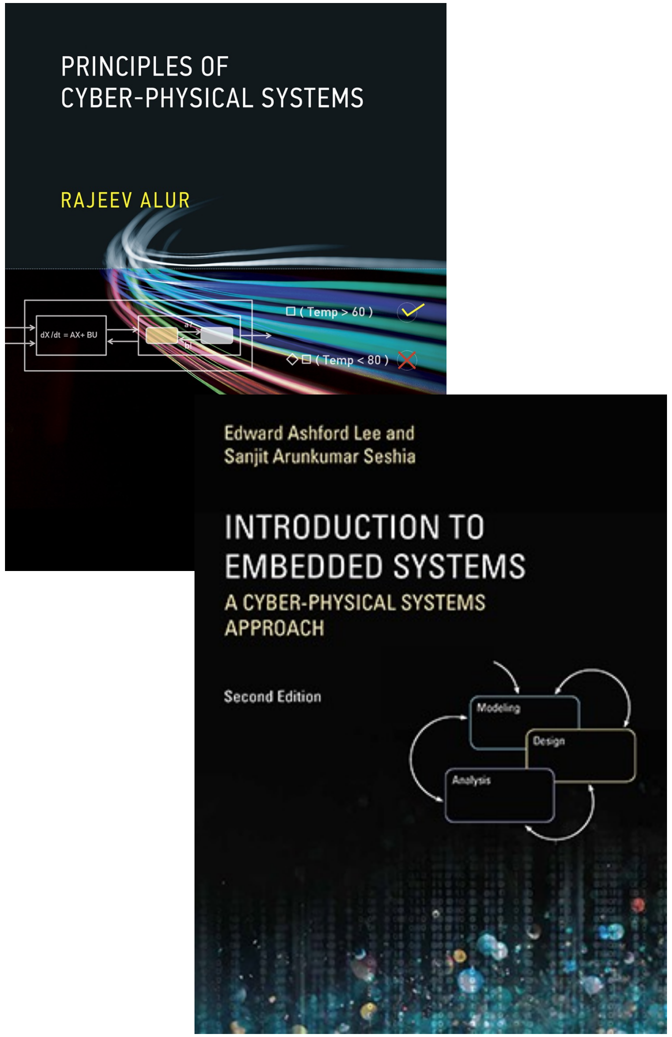
4. **Ingredients of Autonomy** for CPS: planning, decision-making, reinforcement learning

Books (they can just help you)

- Principles of Cyber-Physical Systems, Rajeev Alur, MIT Press, 2015 https://www.biblio.units.it/SebinaOpac/resource/principles-of-cyberphysical-systems/TSA3289844?tabDoc=tabloceb

- Introduction to Embedded Systems: A CPS approach Free at: https://ptolemy.berkeley.edu/books/leeseshia/ https://www.biblio.units.it/SebinaOpac/resource/introduction-to-embedded-systems-a-cyberphysical-systems-approach/TSA3289896?tabDoc=tabloceb

- Principle of Model Checking, Baier, Katoen, MIT Press, 2008



Grading

Project (teams of 1-2) with a practice development of a CPS application, verification of formal requirements and falsification or test generation experiments

You can use:

- Matlab/Simulink (simulation) or
- Python or Java if that is the preferred language (it will require additional work for handling requirements but we can help you!) or
- Open to other software solution

Report of the Project (no more than 5 pages)

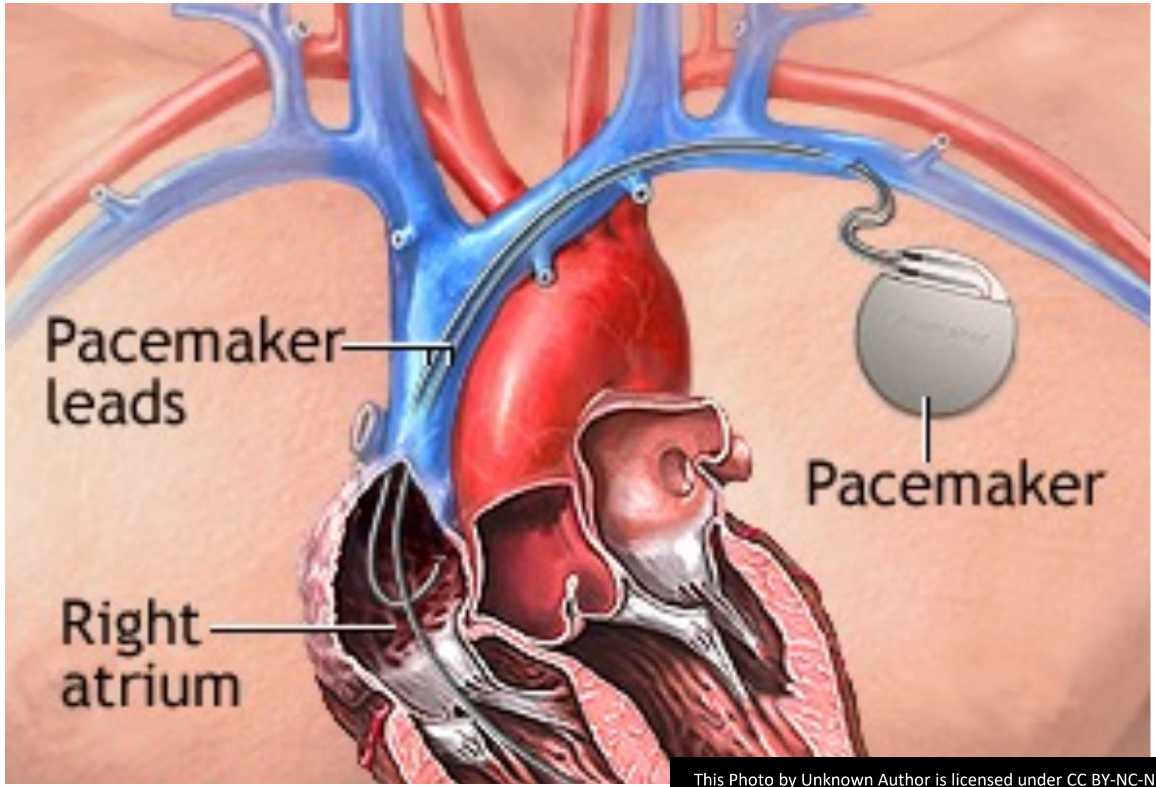
Oral exam with presentation of the Project + general questions on the topics of the course

Example Projects

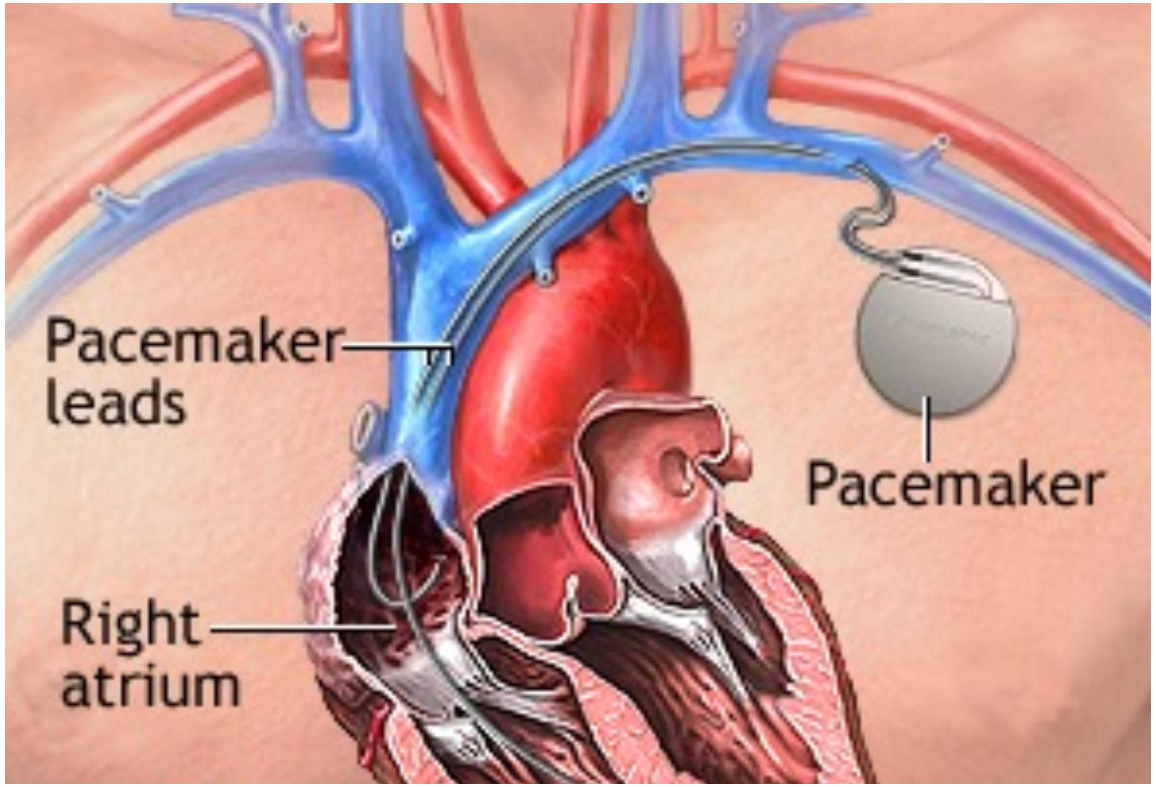
1. Create a model of the human heart and design a pacemaker
2. Create a blood-glucose dynamics model and design an automatic insulin infusion pump
3. Satellite Monitoring System
4. Temperature Control of a Continuous Stirred Tank Reactor (CSTR)
5. A traffic light, ideally equipped with a camera that detects the presence of pedestrians waiting to cross the street
6. Control policy for a simulated self driving car, that needs to be driven along a street track

Questions?

Application domains: Medical Device

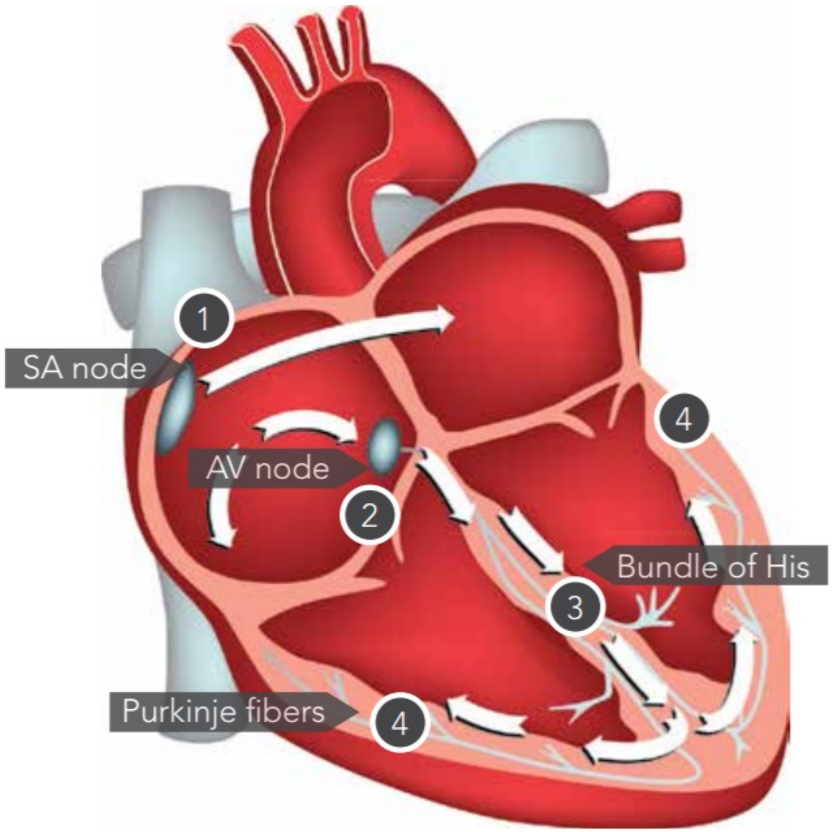


PaceMaker



Z. Jiang, M. Pajic, S. Moarref, R. Alur, R. Mangharam, *Modeling and Verification of a Dual Chamber Implantable Pacemaker*, In Proceedings of Tools and Algorithms for the Construction and Analysis of Systems (TACAS), 2012.

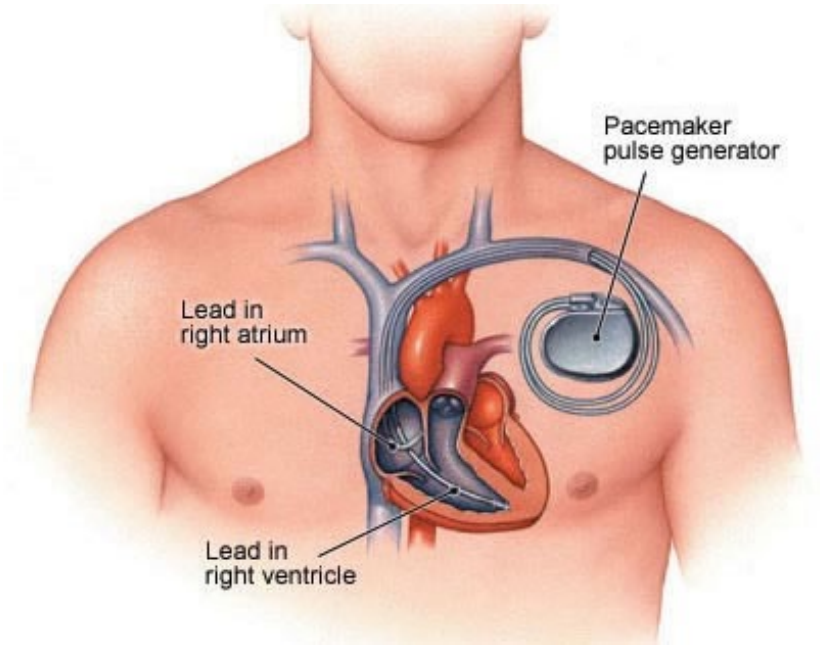
How does a healthy heart work?



- SA node (controlled by nervous system) periodically generates an electric pulse
- This pulse causes both atria to contract pushing blood into the ventricles
- Conduction is delayed at the AV node allowing ventricles to fill
- Finally the His-Pukinje system spreads electric activation through ventricles causing them both to contract, pumping blood out of the heart

Z. Jiang, M. Pajic, S. Moarref, R. Alur, R. Mangharam, *Modeling and Verification of a Dual Chamber Implantable Pacemaker*, In Proceedings of Tools and Algorithms for the Construction and Analysis of Systems (TACAS), 2012.

PaceMaker

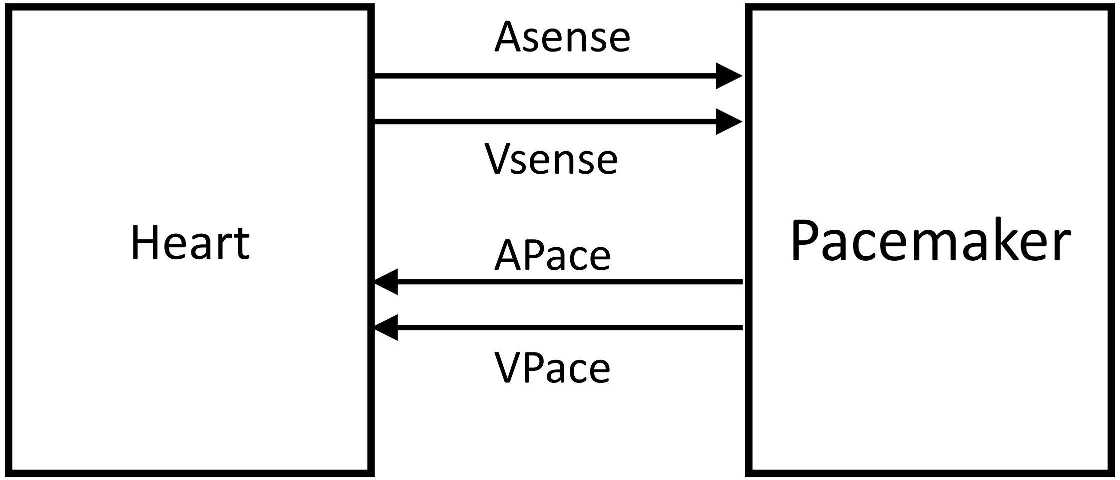


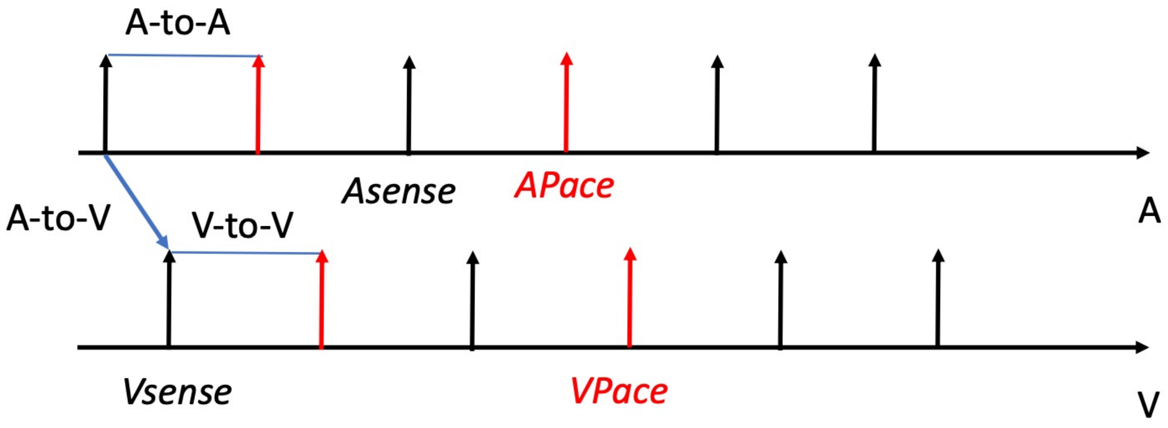
- Aging and/or diseases cause conduction properties of heart tissue to change leading to changes in heart rhythm
- Tachycardia: faster than desirable heart rate impairing hemo-dynamics (blood flow dynamics)
- Bradycardia: slower heart rate leading to insufficient blood supply
- Pacemakers can be used to treat bradycardia by providing pulses when heart rate is low

How dual-chamber pacemakers work

- Activation of local tissue sensed by the leads (giving rise to events Atrial Sense and Ventricular Sense)

- Atrial Pace or Ventricular Pace are delivered if no sensed events occur within deadlines

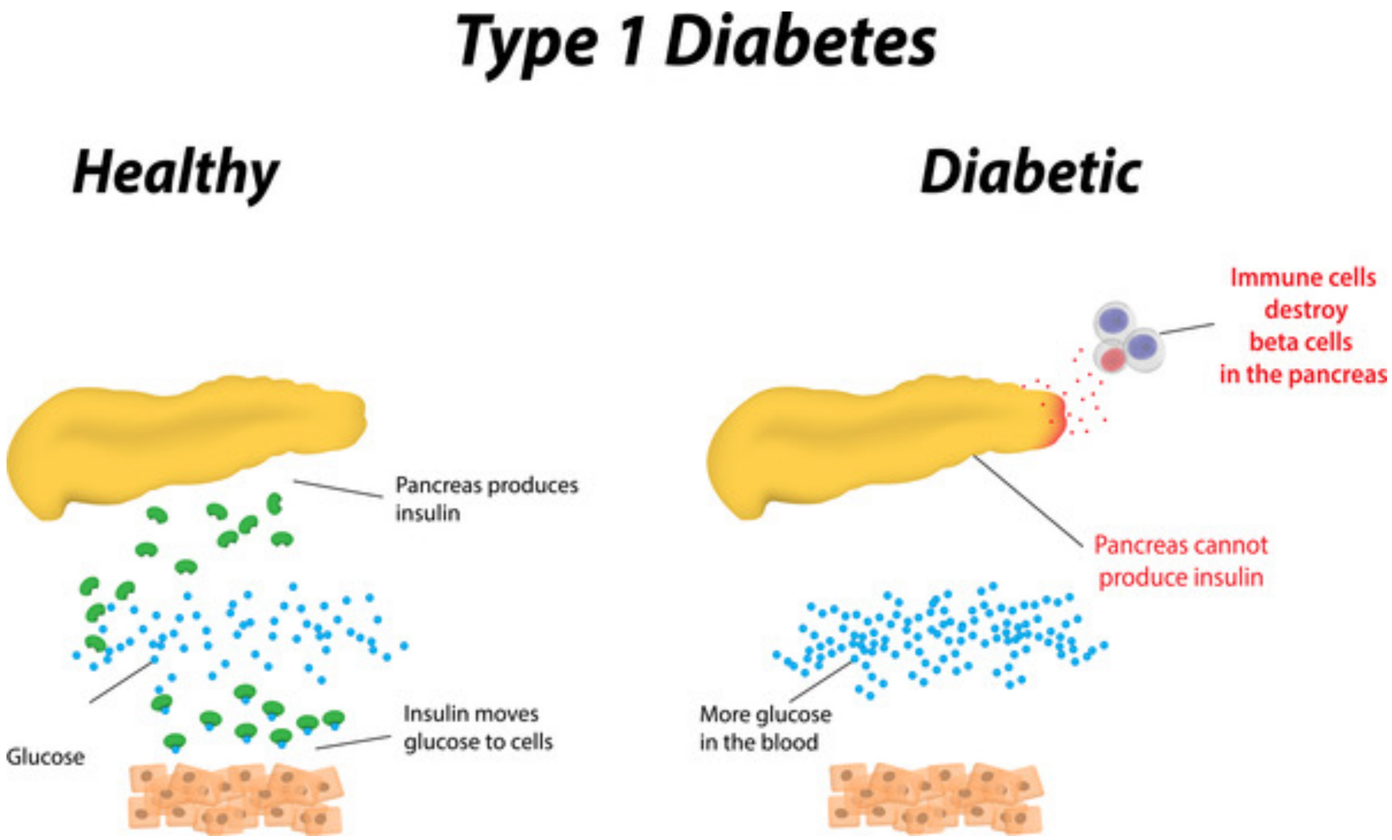




Type 1 Diabetes

Type 1 diabetes occurs when the pancreas produces little or none of the insulin needed to regulate blood glucose

They rely on external ad-ministration of insulin to manage their blood glucose levels.



Continuous Glucose Monitoring

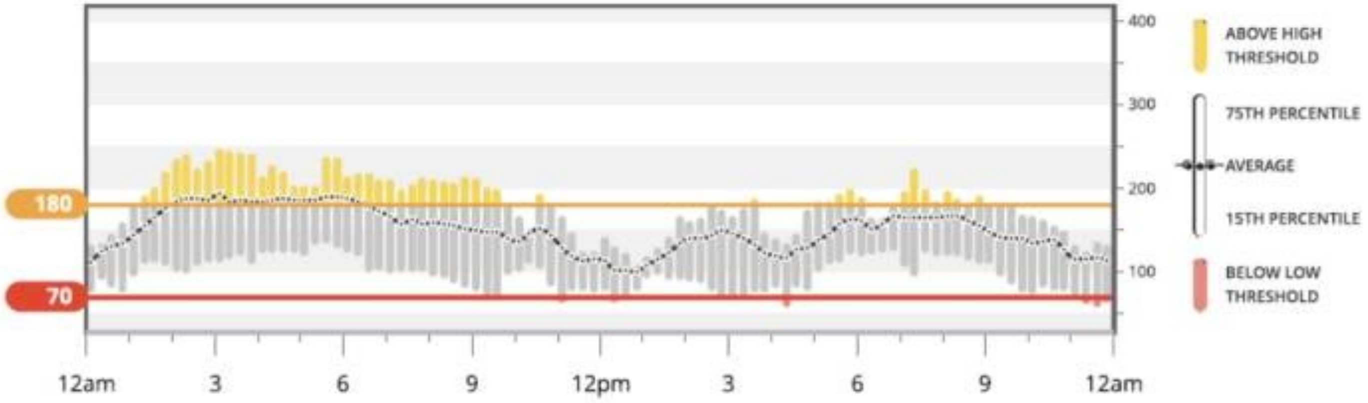
14 days | [redacted]

N/A	151 mg/dL	58 mg/dL	HIGH / MEDIUM / LOW	30.3% HIGH / 65.4% IN RANGE / 4.2% LOW	2.2
Estimated A1C	Average glucose (CGM)	Standard deviation (CGM)	Hypoglycemia risk	Time in range	Average daily CGM calibrations

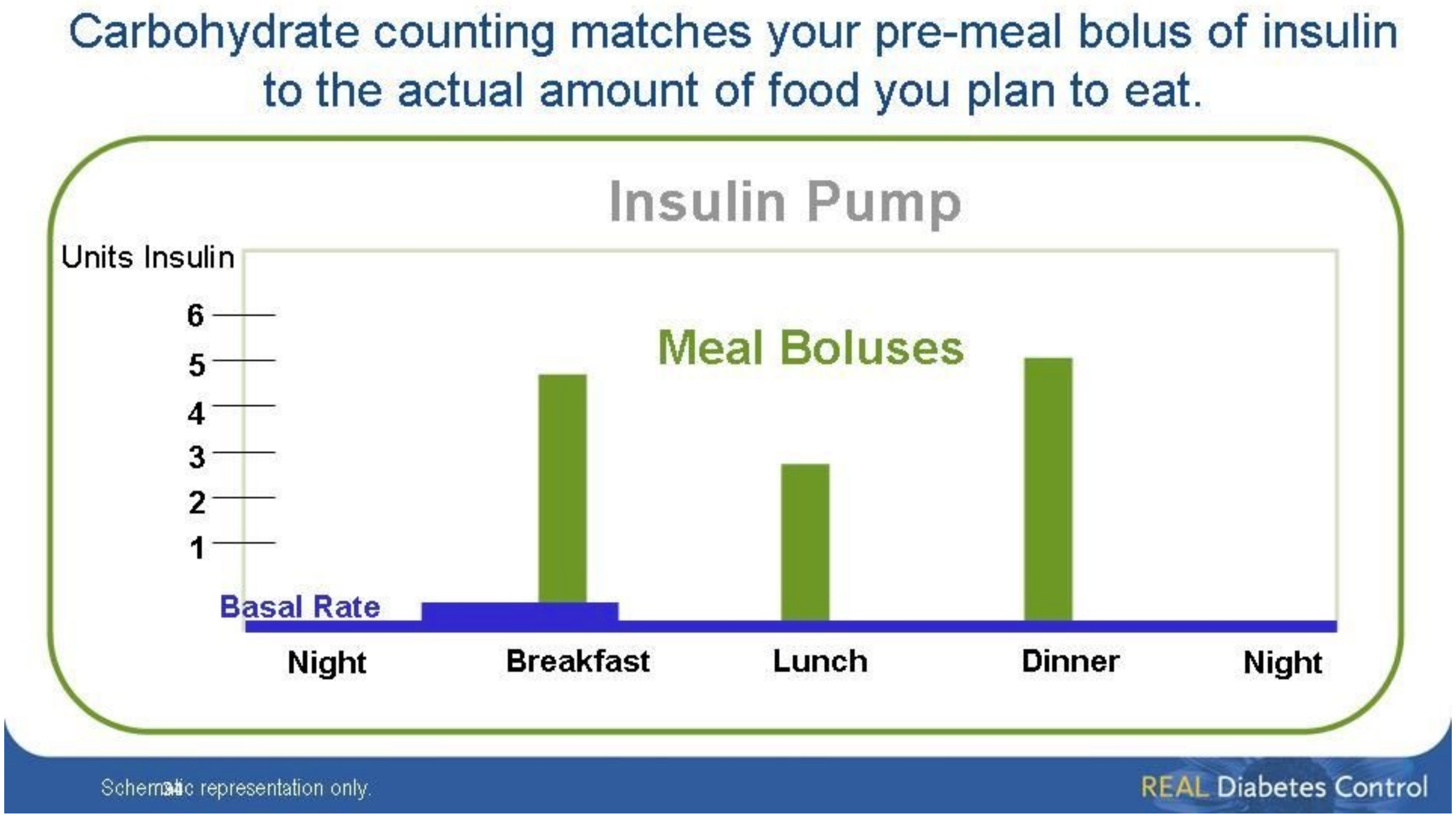
Top Patterns

[redacted]

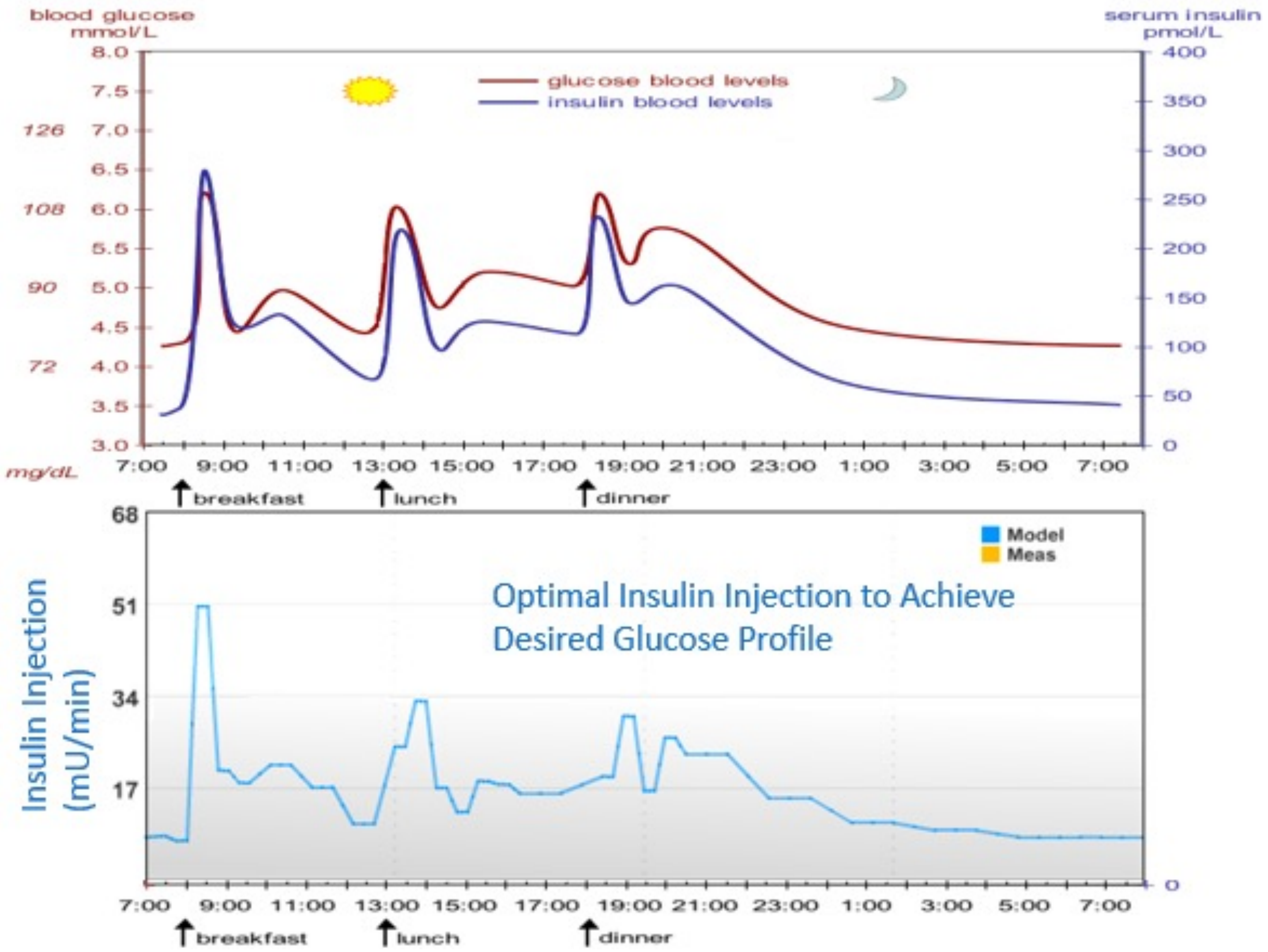
This graph shows your data averaged over 14 days



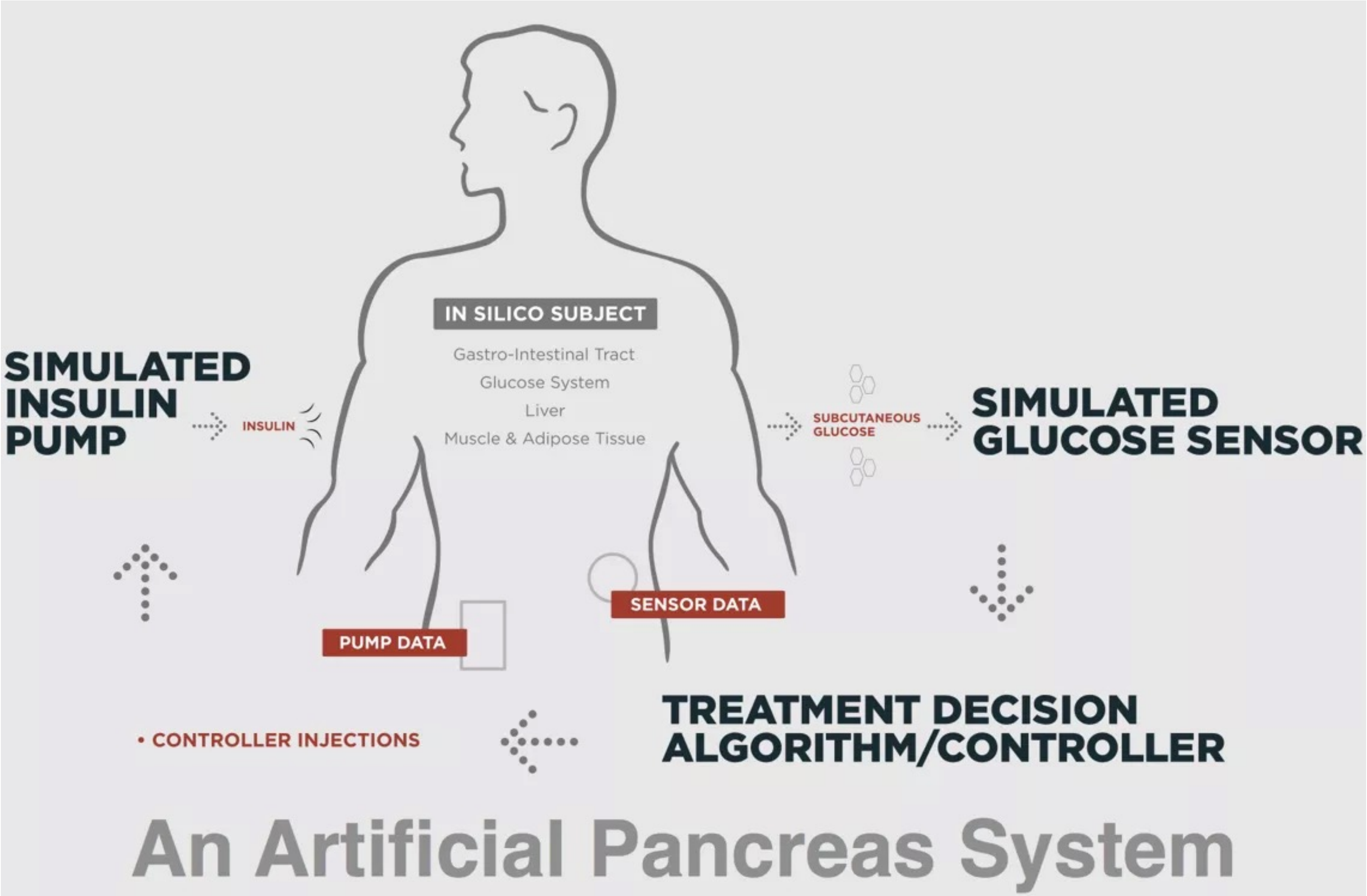
Insulin pumps



Artificial Pancreas



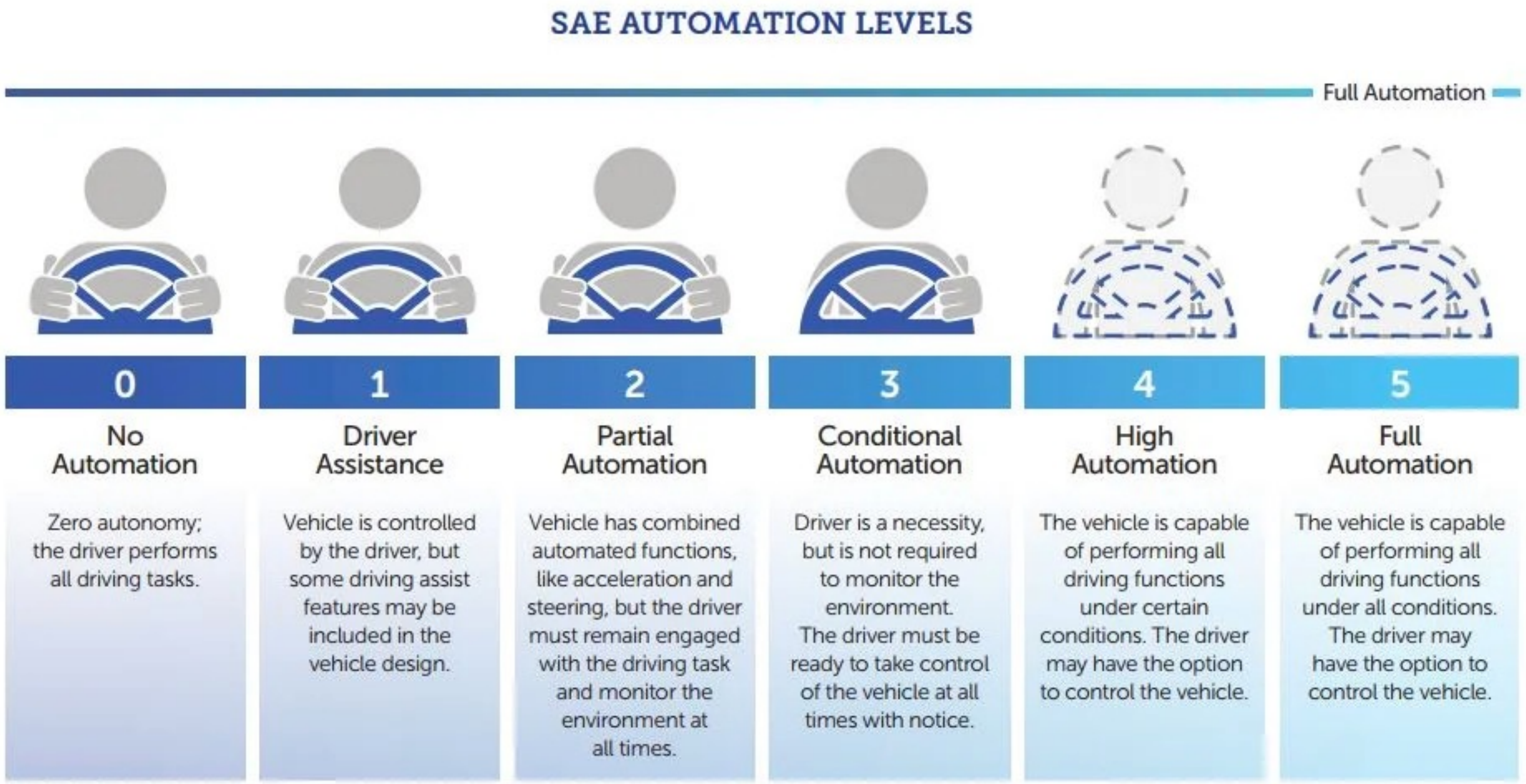
Artificial Pancreas



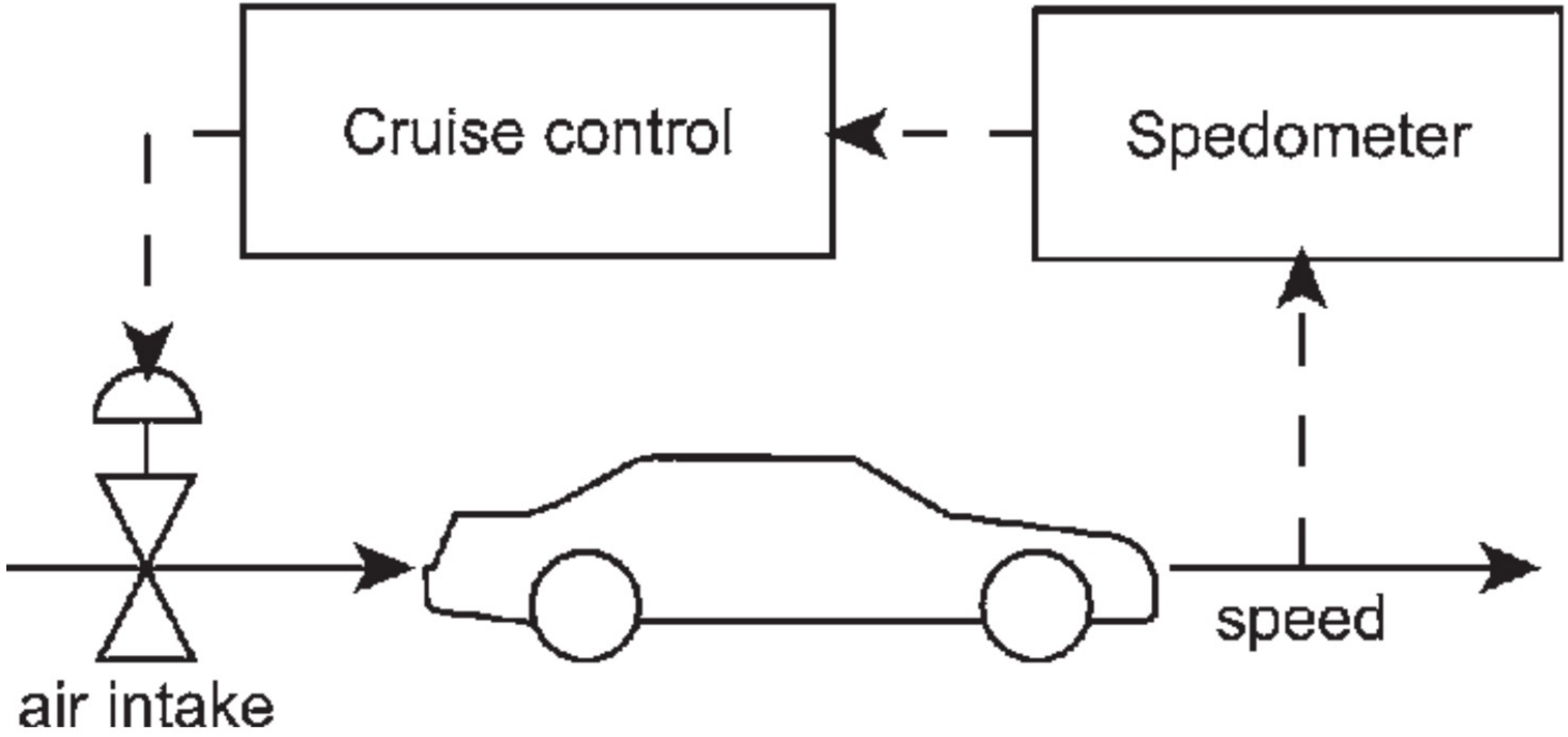
Application domains: Transportation CPS

Everything that moves will become autonomous

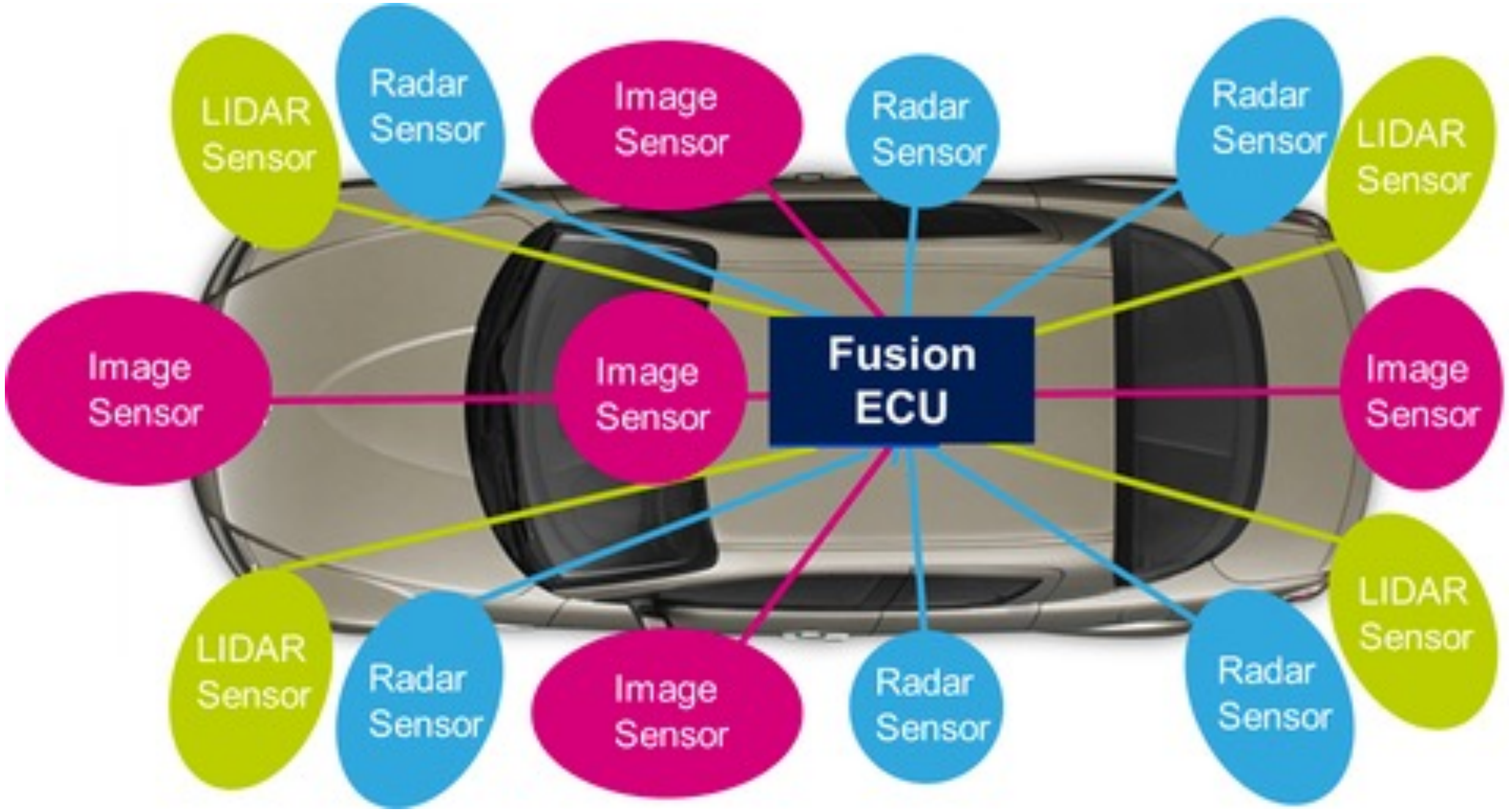
Automotive Car



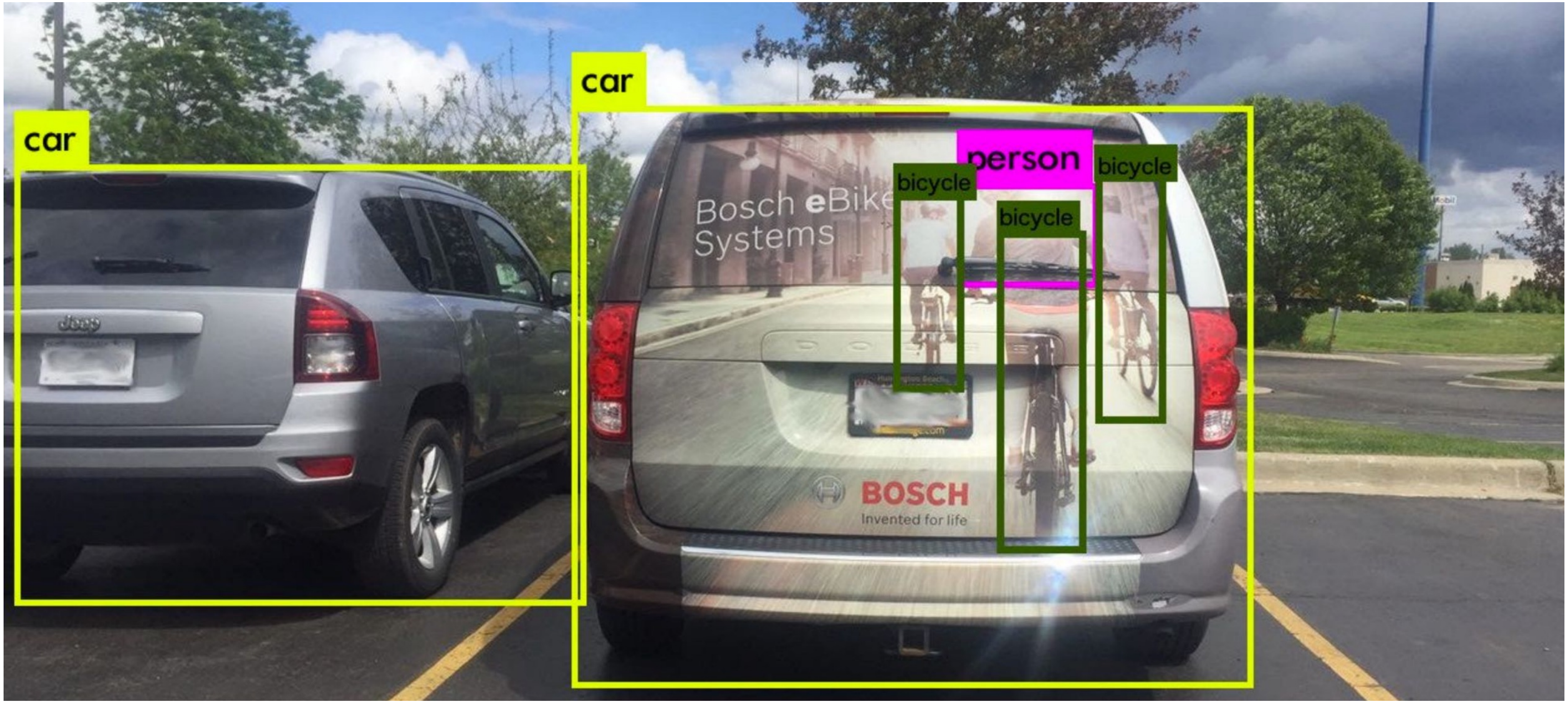
Automotive Car



Automotive Car

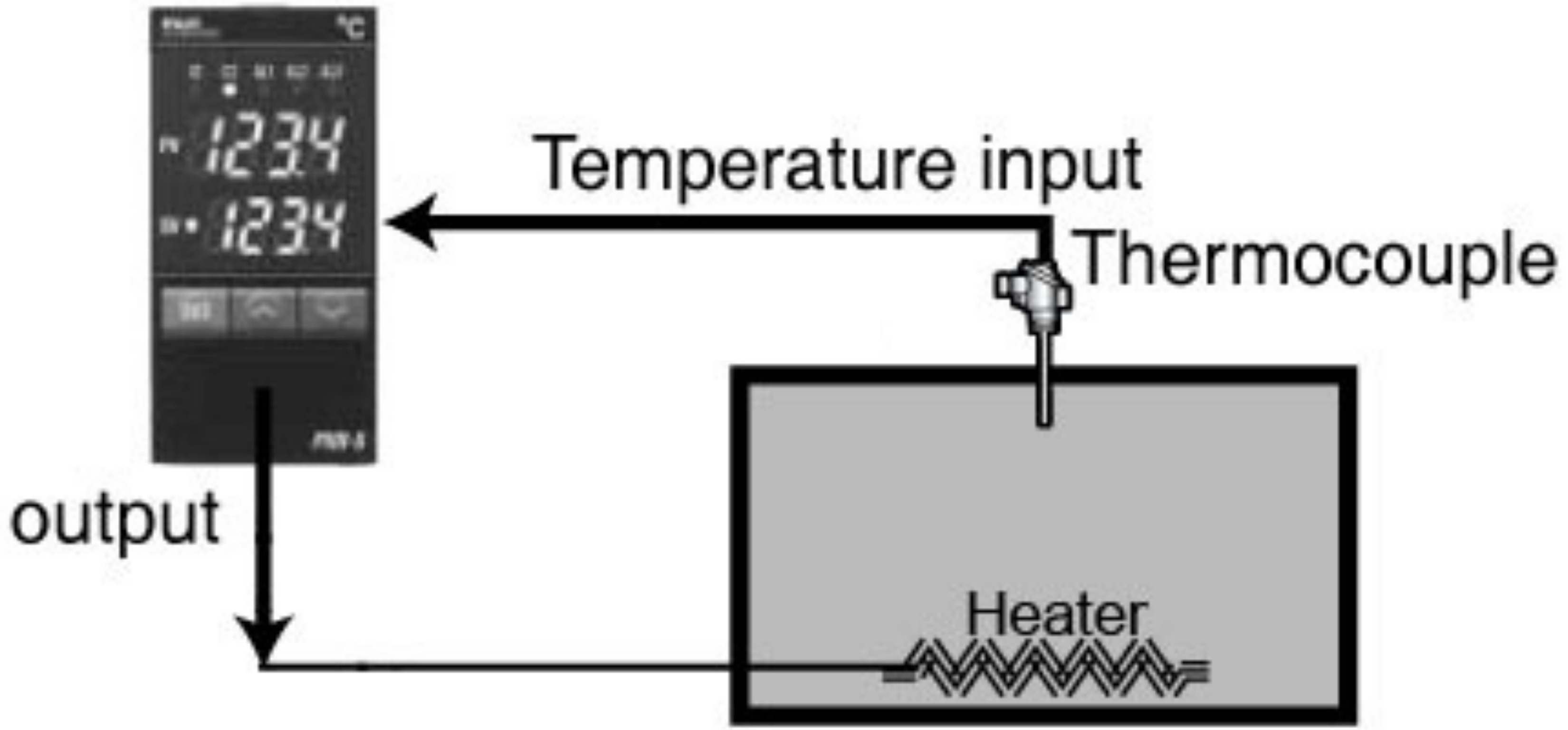


Automotive Car

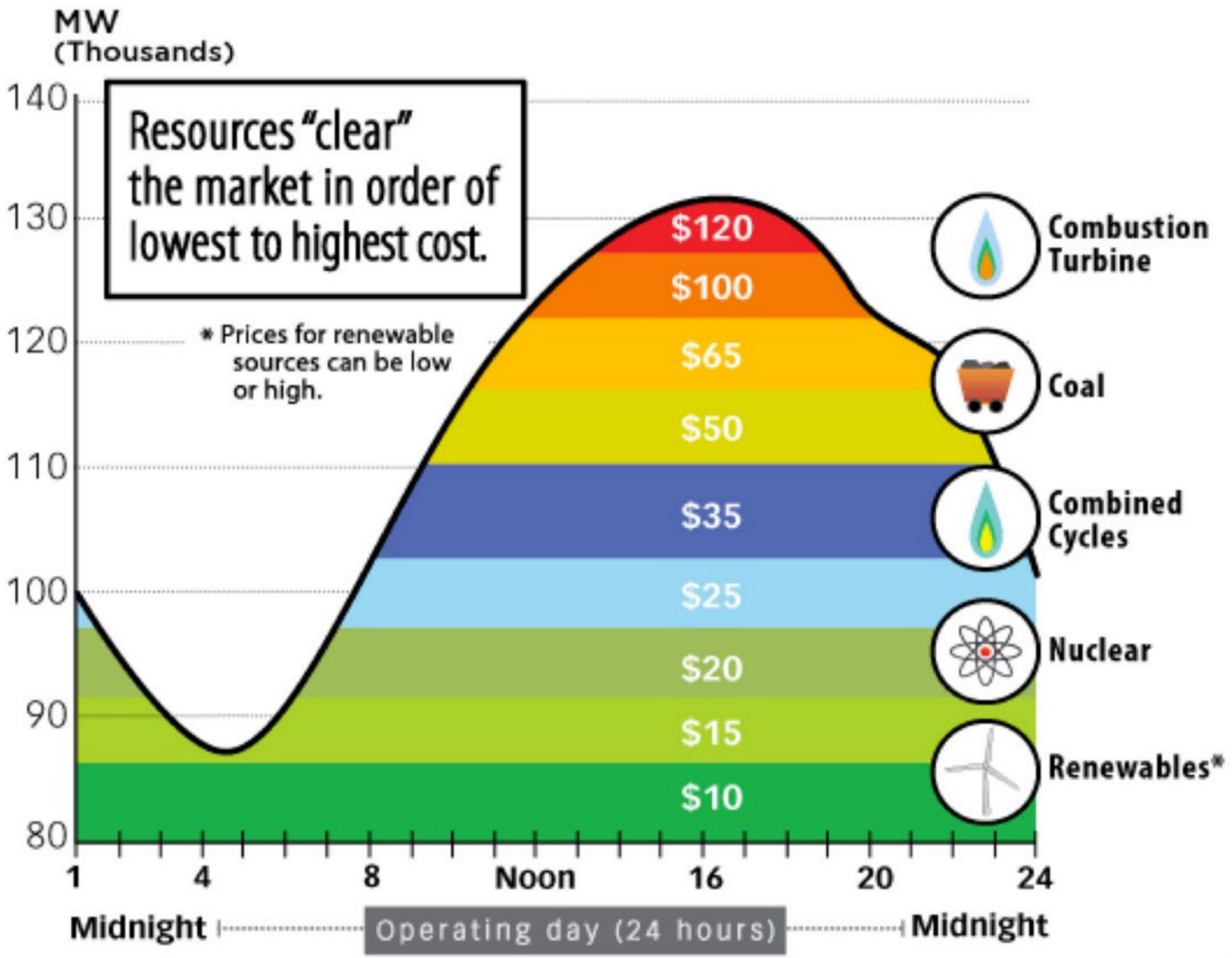


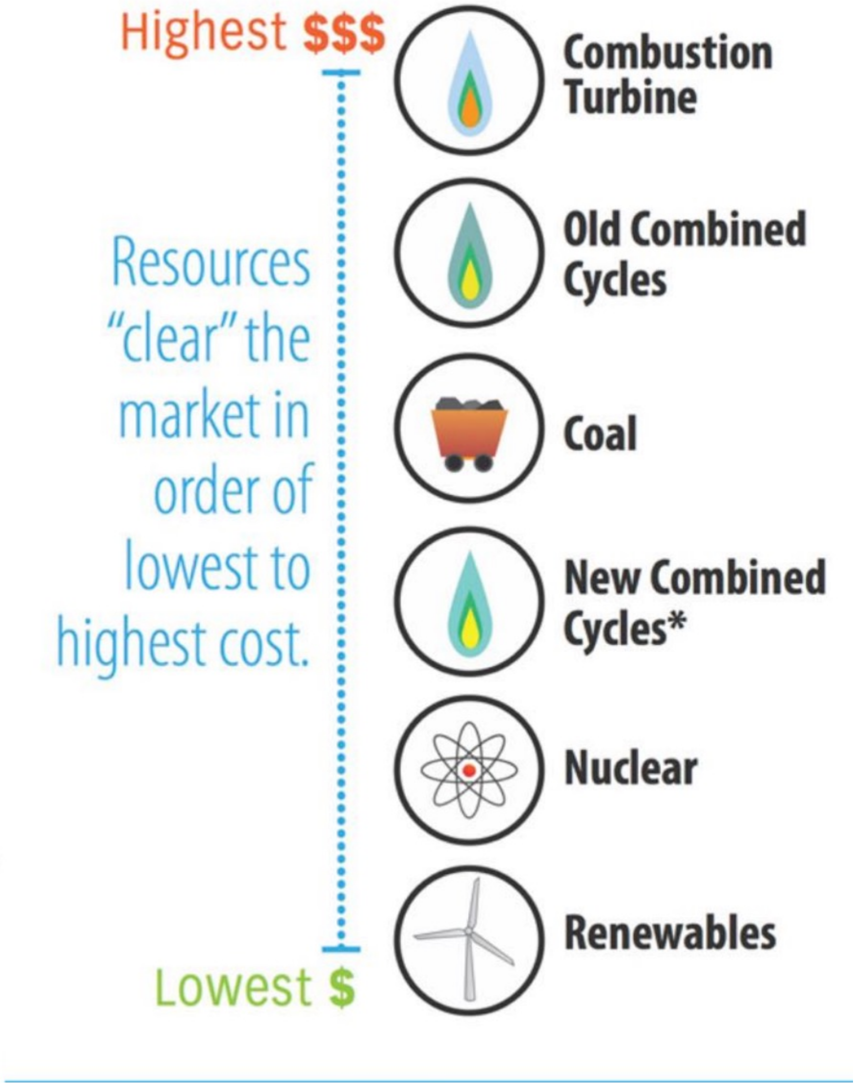
Application domains: Energy

Temperature Control



Energy Control





even-thermostats-have-a-heart