CS249r: Special Topics in Edge Computing—Autonomous Machines

Vijay Janapa Reddi
TA: Brian Plancher

Harvard John A. Paulson
School of Engineering
and Applied Sciences

Goals of the Class

Goals of the Class

1. What are **real-time autonomous machines**?

Goals of the Class

1. What are **real-time autonomous machines**?

2. What is the **enabling technology** behind them?

Goals of the Class

1. What are **real-time autonomous machines**?
2. What is the **enabling technology** behind them?
3. What are the **technical challenges** we need to solve?

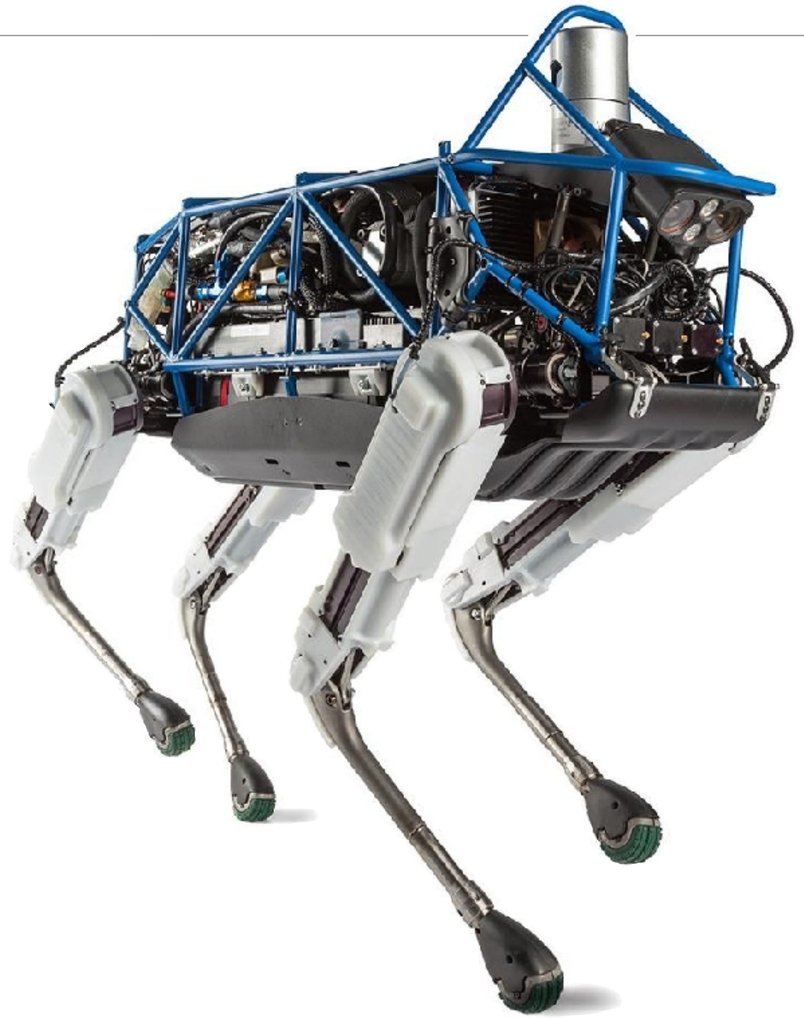
Goals of the Class

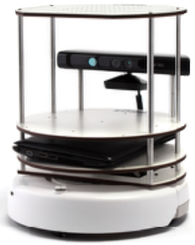
1. What are **real-time autonomous machines**?
2. What is the **enabling technology** behind them?
3. What are the **technical challenges** we need to solve?
4. What **research infrastructure** is needed to enable studies?

Goals of the Class

1. What are **real-time autonomous machines**?
2. What is the **enabling technology** behind them?
3. What are the **technical challenges** we need to solve?
4. What **research infrastructure** is needed to enable studies?
5. What **new opportunities** can we unlock for the future of AI machines?

Q1. What are real-time autonomous machines? (Types of robots)





Q2. What is the enabling technology behind them?

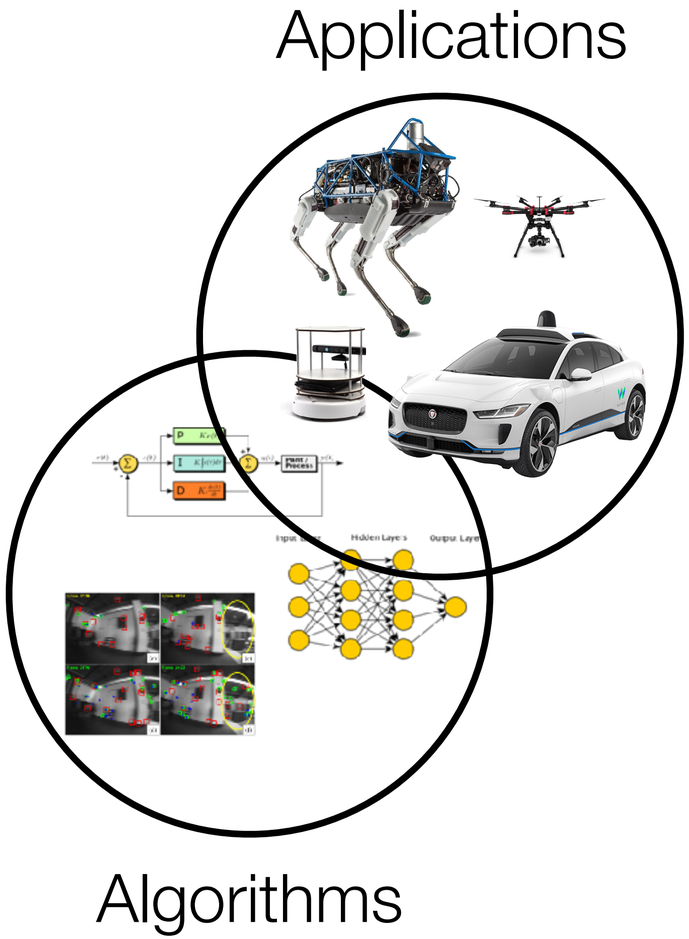
Q2. What is the enabling technology behind them?

- **Applications**
  - Requirements

Applications

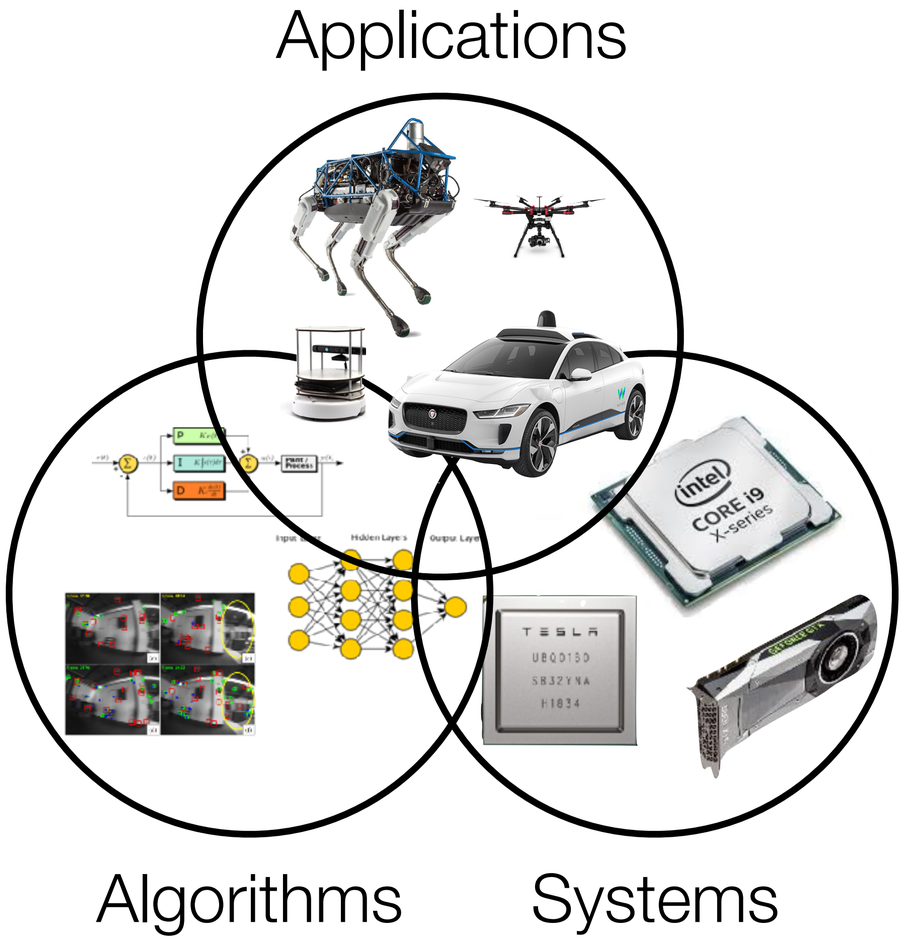
Q2. What is the enabling technology behind them?

- **Applications**
  - Requirements
- **Algorithms**
  - Methods



Q2. What is the enabling technology behind them?

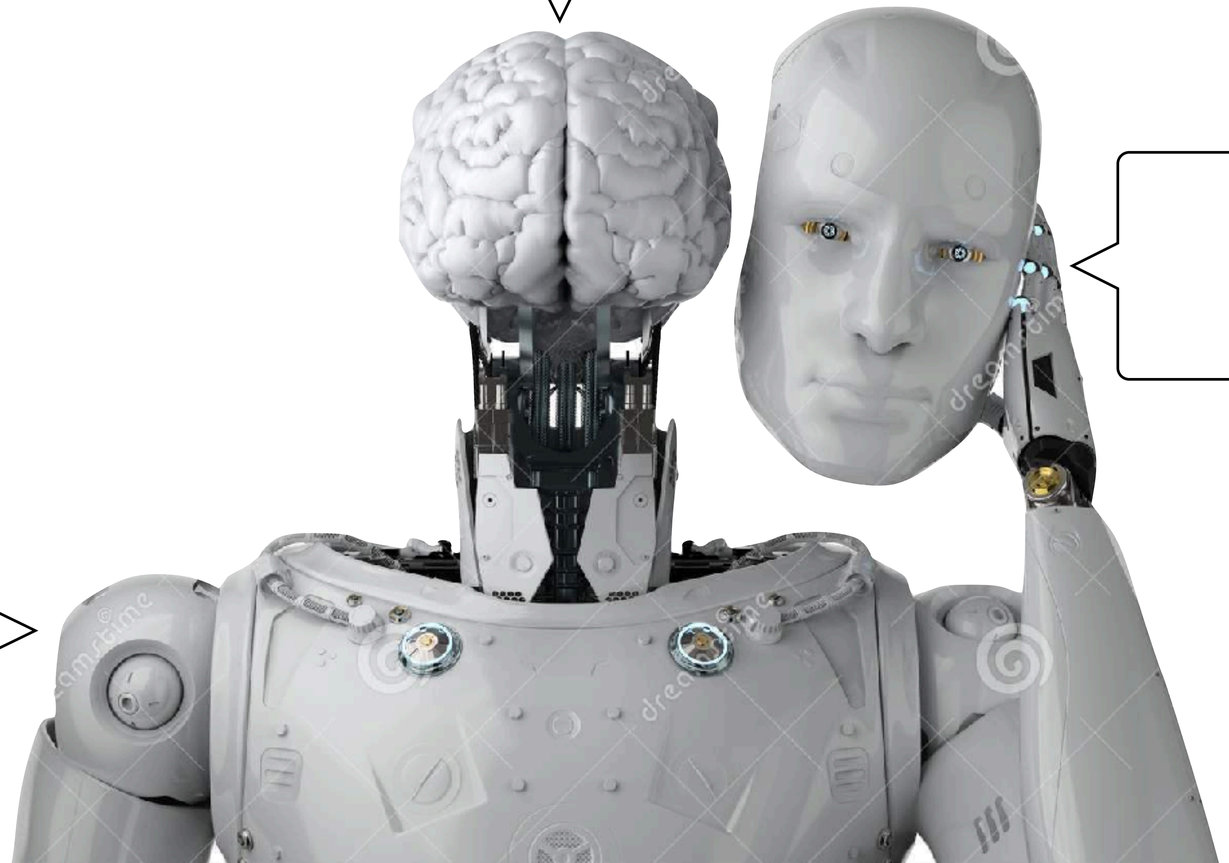
- **Applications**
  - Requirements
- **Algorithms**
  - Methods
- **Systems**
  - Capabilities



Q3. What are the technical challenges we need to solve?

- **Algorithms**
  - Unlock new capabilities through algorithm innovation
  - Traditional versus machine learning methods (pros and cons)
- **System**
  - Offboard compute versus onboard compute
  - Performance vs. power consumption trade-offs

Q3. What are the technical challenges we need to solve? (Algorithms)



Planning

Perception

Control

Q3. What are the technical challenges we need to solve? (Challenges with the Traditional “PPC” Pipeline)

Q3. What are the technical challenges we need to solve? (Challenges with the Traditional “PPC” Pipeline)

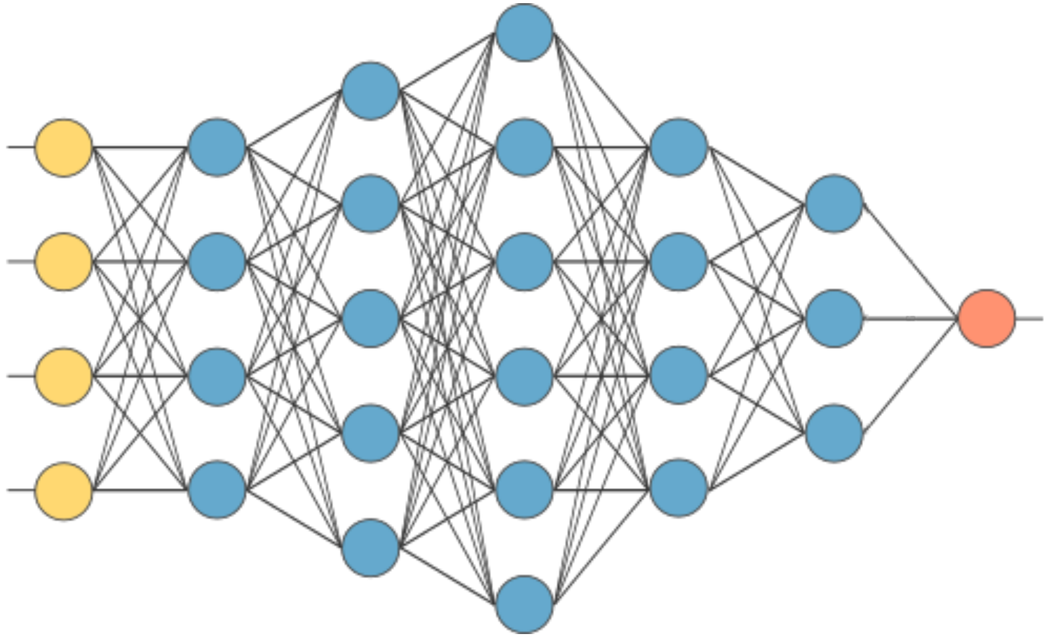
- Sensors are not perfect
- …

- Need to know the full map
- Short-term reasoning
- …

- Model based/specific
- Hard to generalize for “noise”
- …

Q3. What are the technical challenges we need to solve? (Use ML to approximate the function)

Pixels

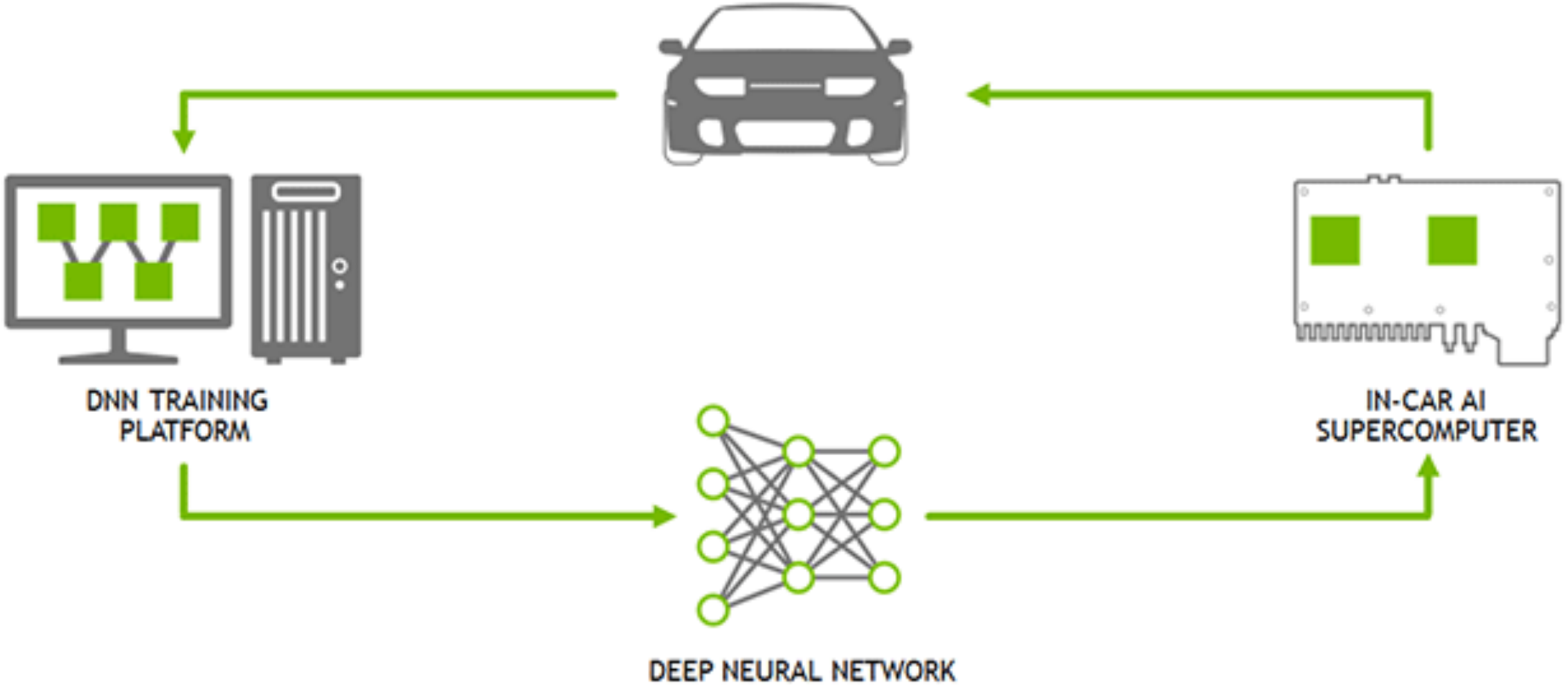


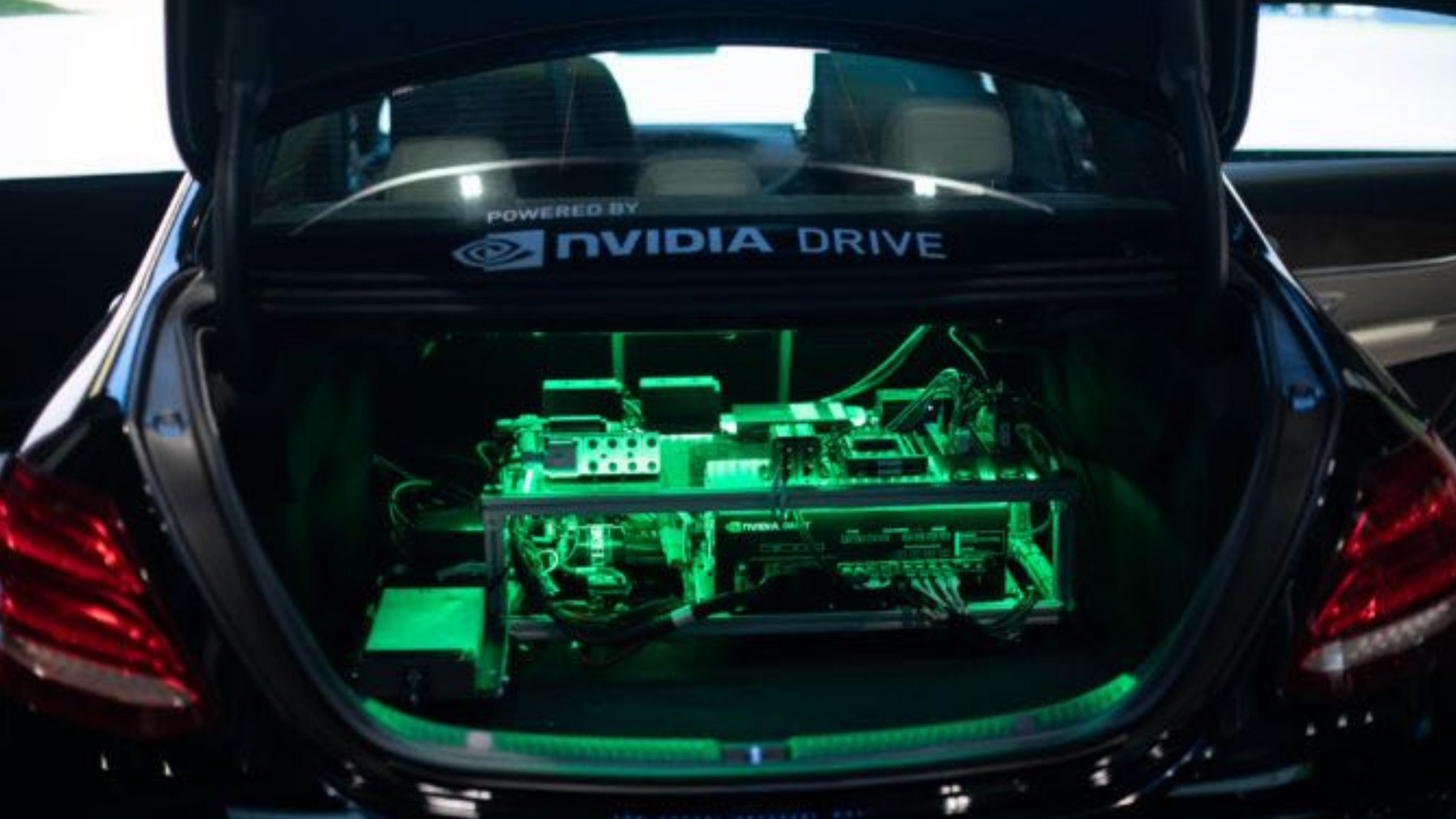
Control

Approximate the function

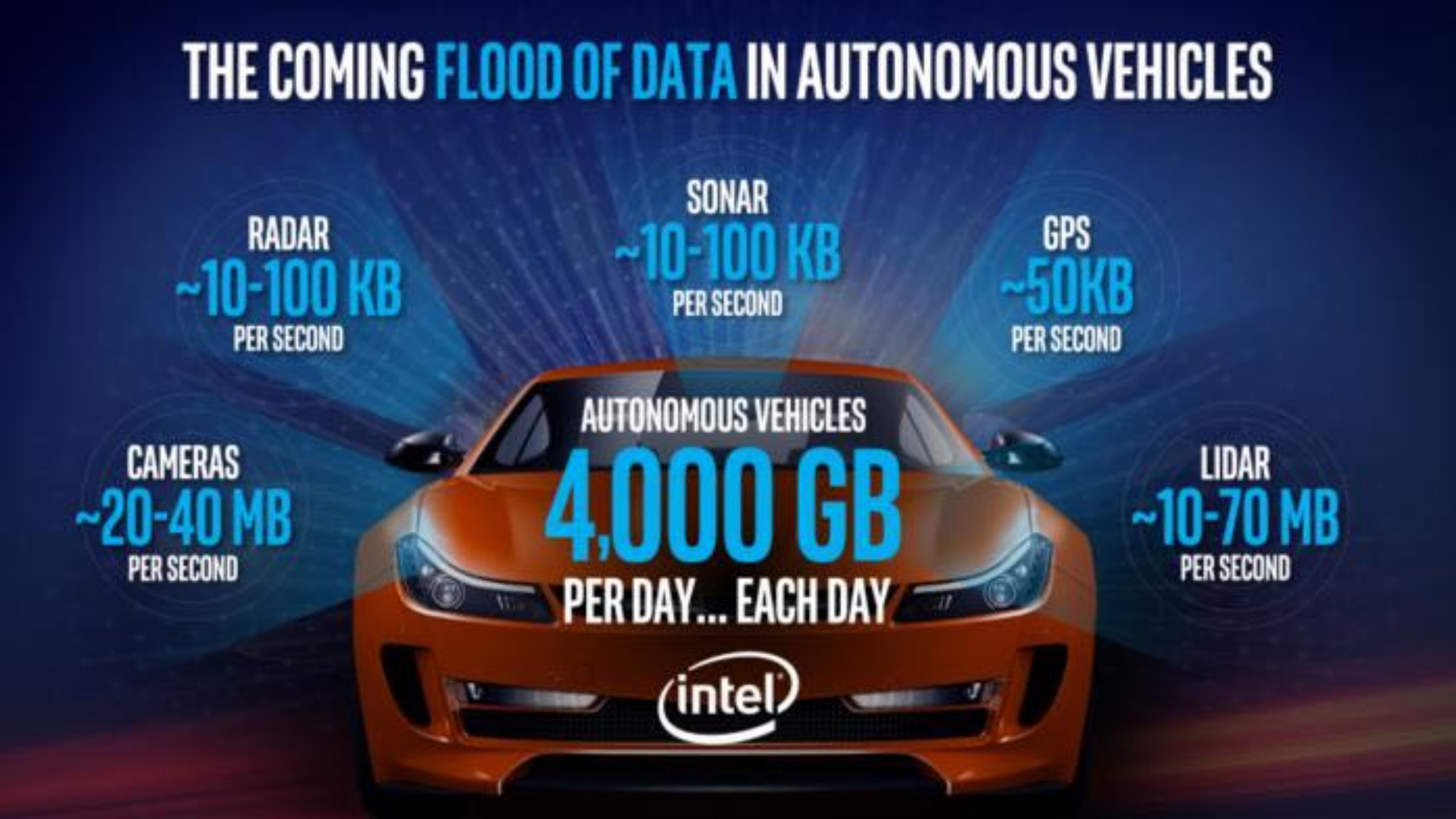
https://youtu.be/fmVWLr0X1Sk?t=30

END-TO-END DEEP LEARNING PLATFORM FOR SELF-DRIVING CARS




POWERED BY
NVIDIA DRIVE

THE COMING FLOOD OF DATA IN AUTONOMOUS VEHICLES



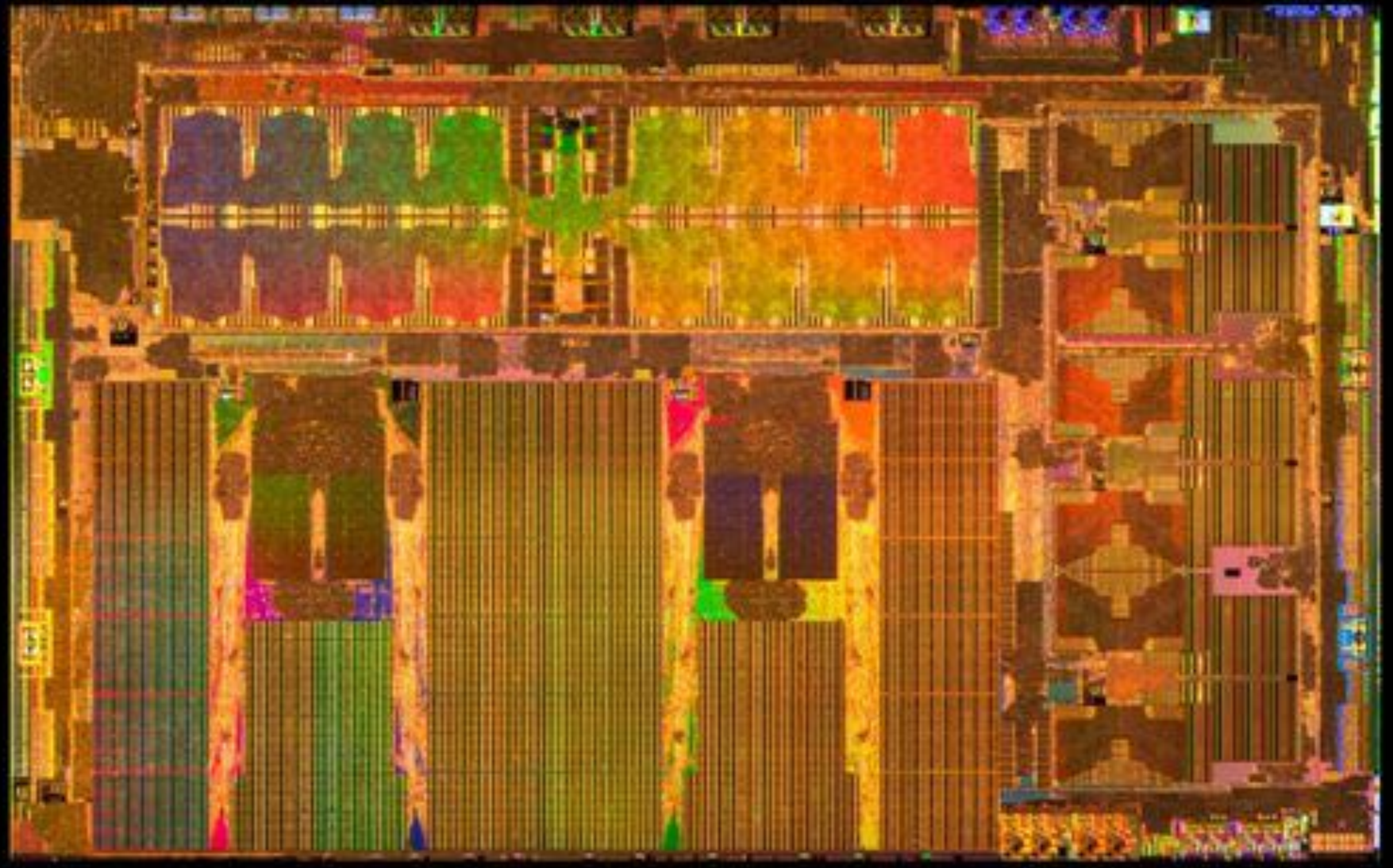
Q3. What are the technical challenges we need to solve?

- “A production car you can buy today, with just cameras and radar, generates something like **6 gigabytes of data every 30 seconds**.”

- “That takes huge computing power, which means huge electricity demands. **Prototypes use around 2,500 watts,** enough to light 40 incandescent light bulbs.”

- “To put such a system into a combustion-engined car doesn’t make any sense, because the **fuel consumption will go up tremendously**,” says Wilko Stark, Mercedes-Benz's vice president of strategy.”

FSD CHIP TOUR



14nm FinFET CMOS

260 mm²

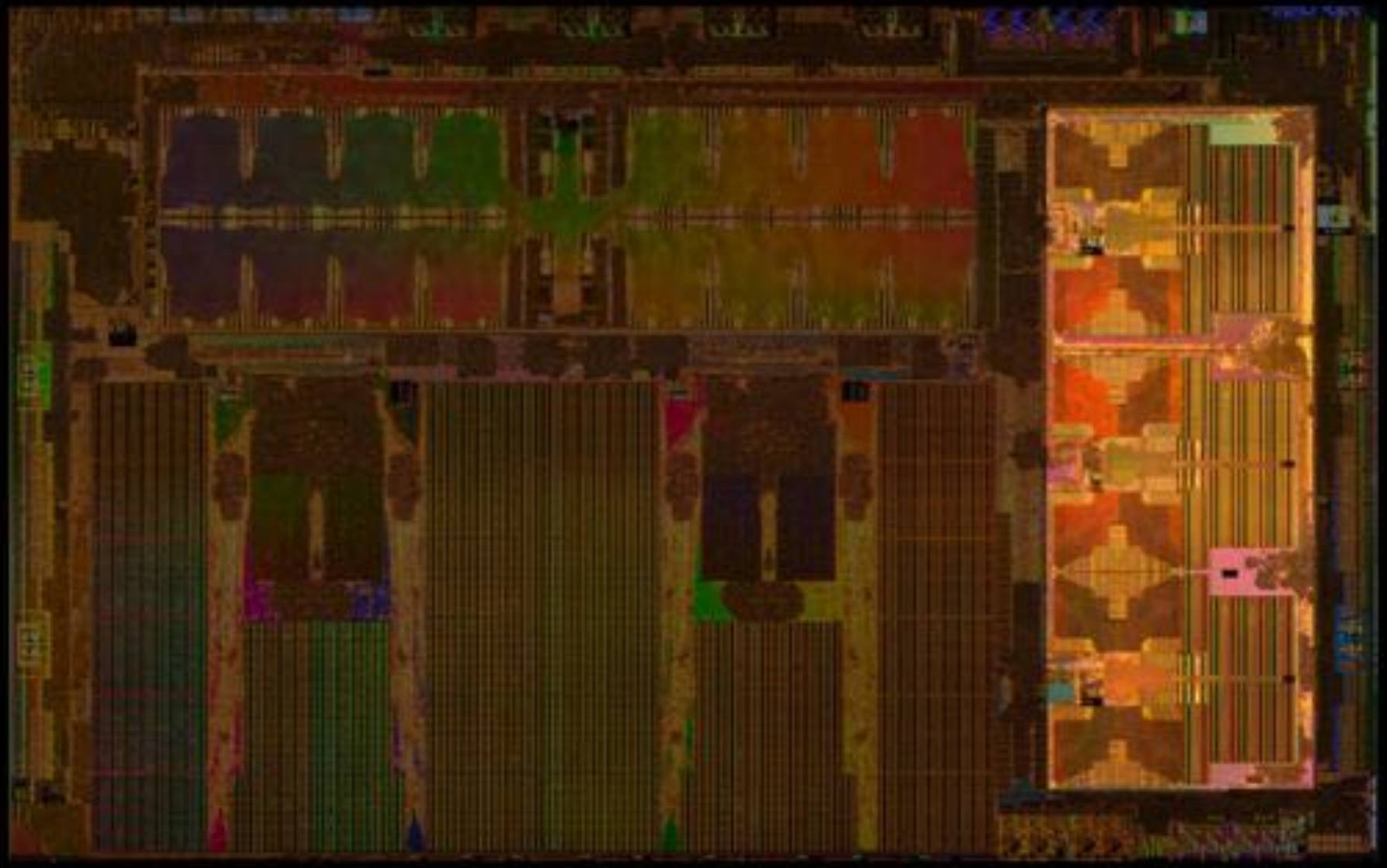
250 million gates

6 billion transistors

AEC Q100

TESLA

MAIN PROCESSOR

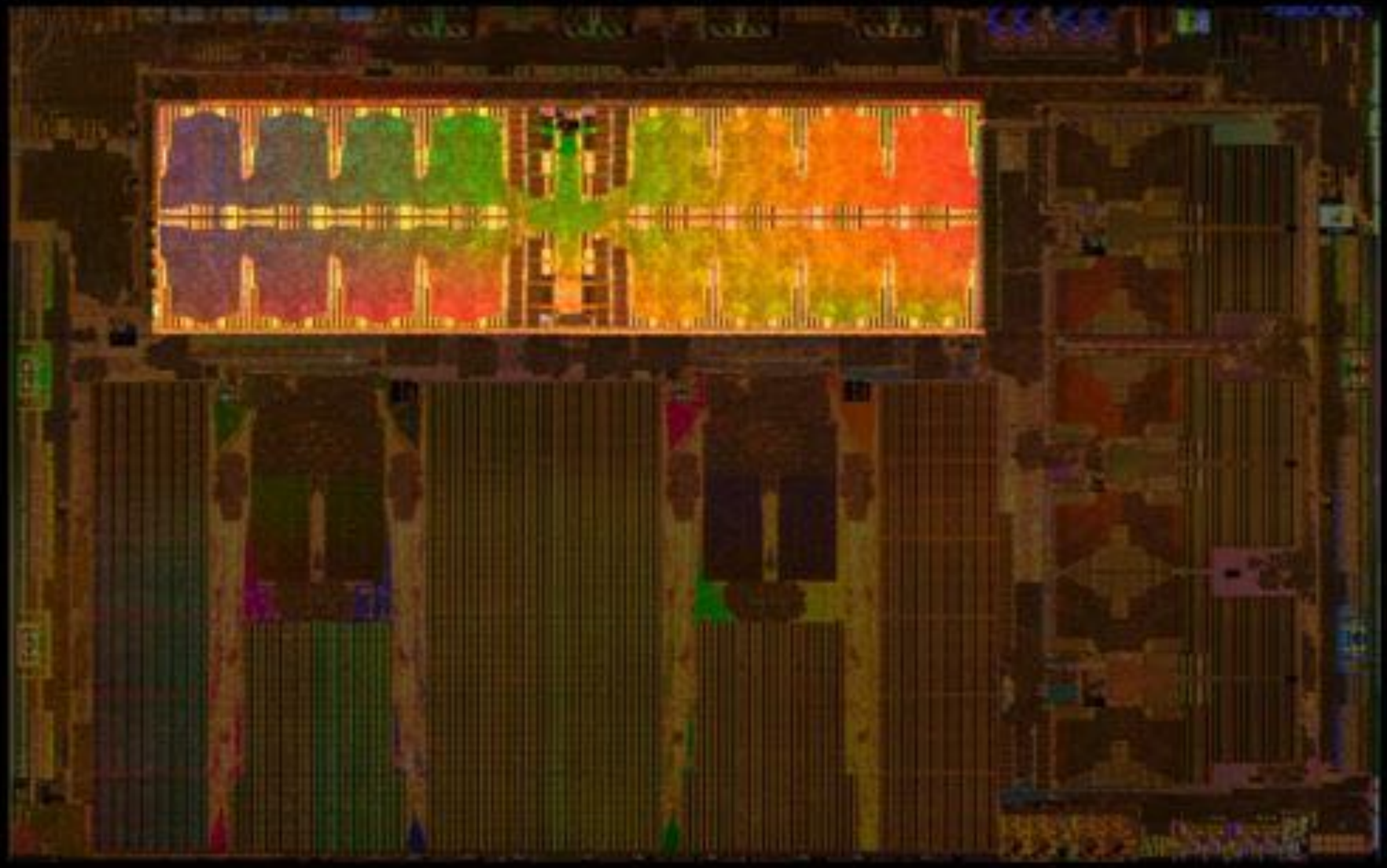


12 ARM A72 64b CPUs

2.2 GHz

TESLA

GPU



1 Ghz

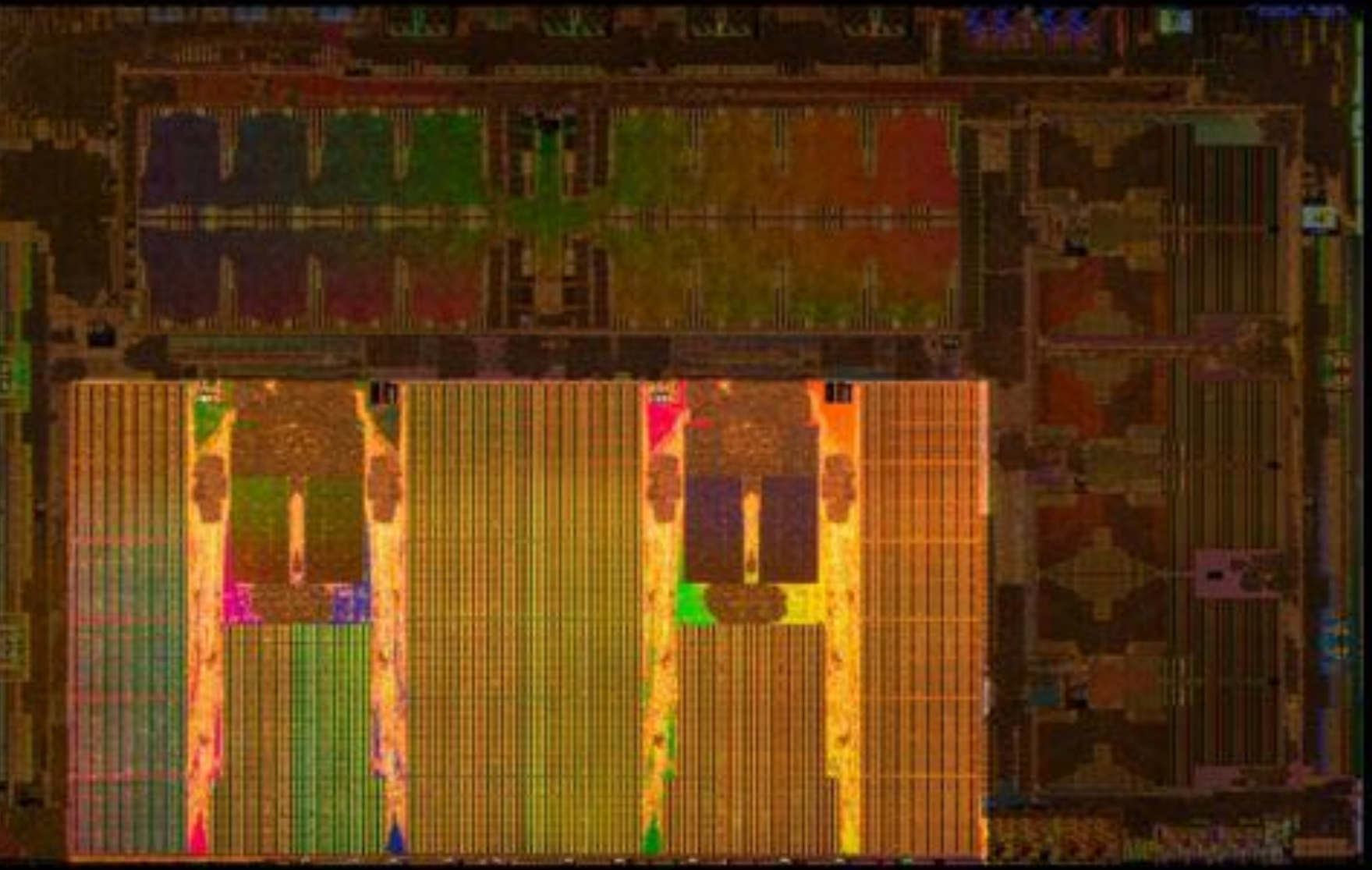
600 GFLOPS

FP32

FP16

TESLA

NEURAL NETWORK PROCESSOR



32MB SRAM

96x96 Mul/Add array

ReLU hardware

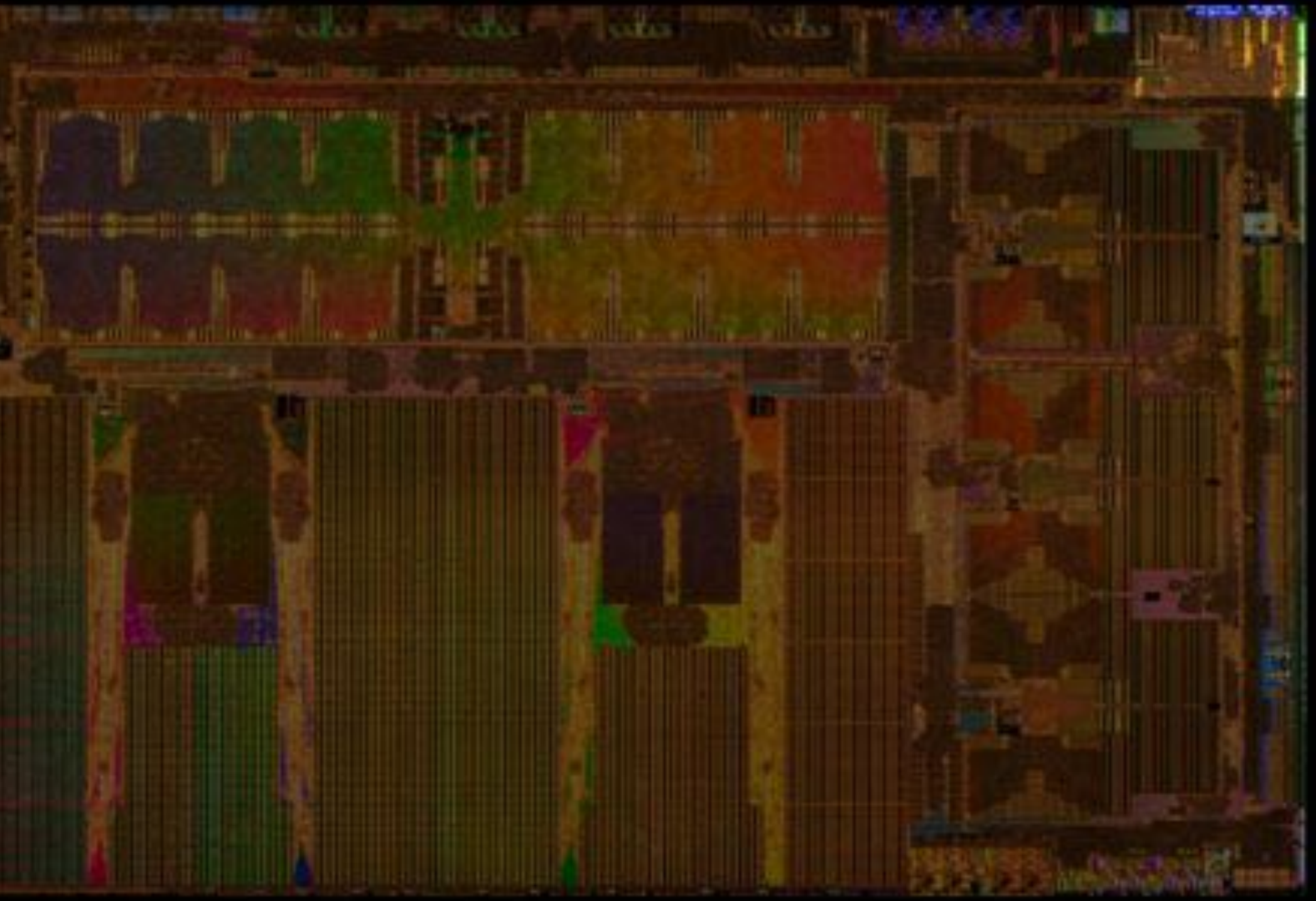
Pooling hardware

36 TOPS @ 2 GHz

2 per chip, 72 TOPS total

TESLA

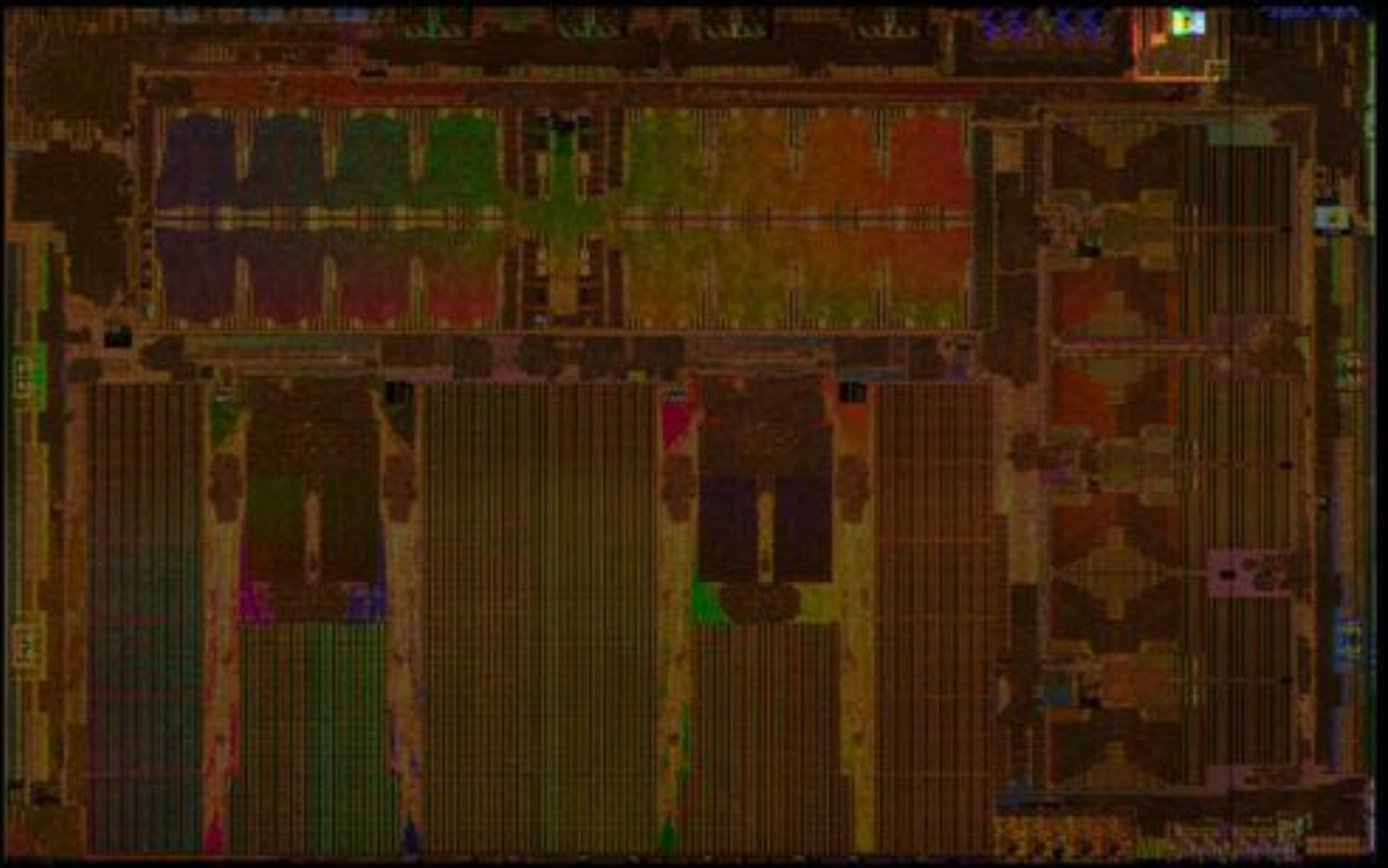
SECURITY SYSTEM



Ensure the system only runs code cryptographically signed by Tesla

TESLA LIVE

SAFETY SYSTEM



Lock step CPU

Final Drive Control

Control Validation

TESLA

PERSPECTIVE (GOPS)

Frame per second for 35 GOP network: 2100

TESLA LIVE

Q3. What are the technical challenges we need to solve? (Performance vs. Power Trade-offs in Form Factor)

87 kWh

Q3. What are the technical challenges we need to solve? (Performance vs. Power Trade-offs in Form Factor)

87 kWh

0.08 kWh

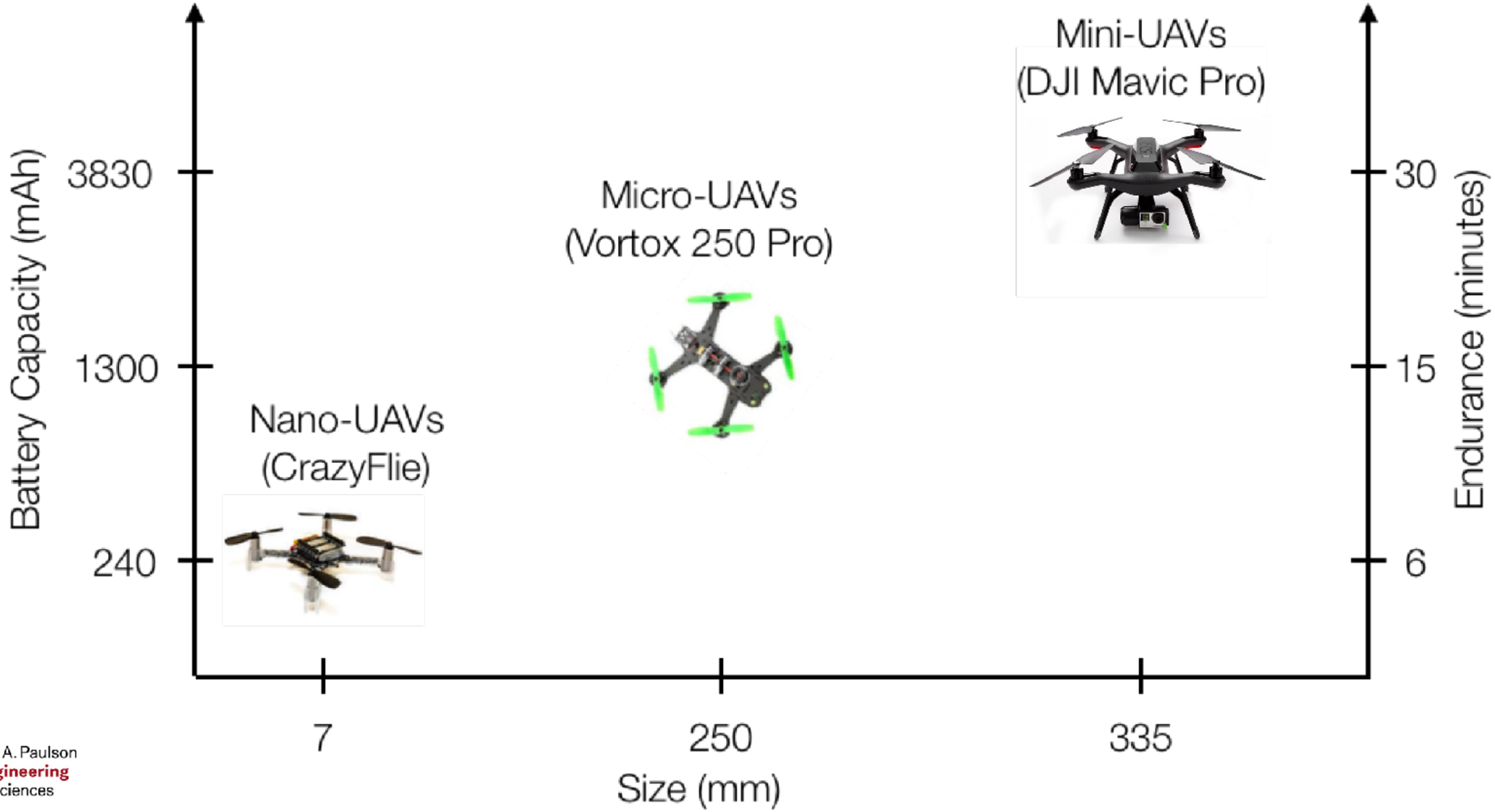
Q3. What are the technical challenges we need to solve? (Performance vs. Power Trade-offs in Form Factor)

87 kWh

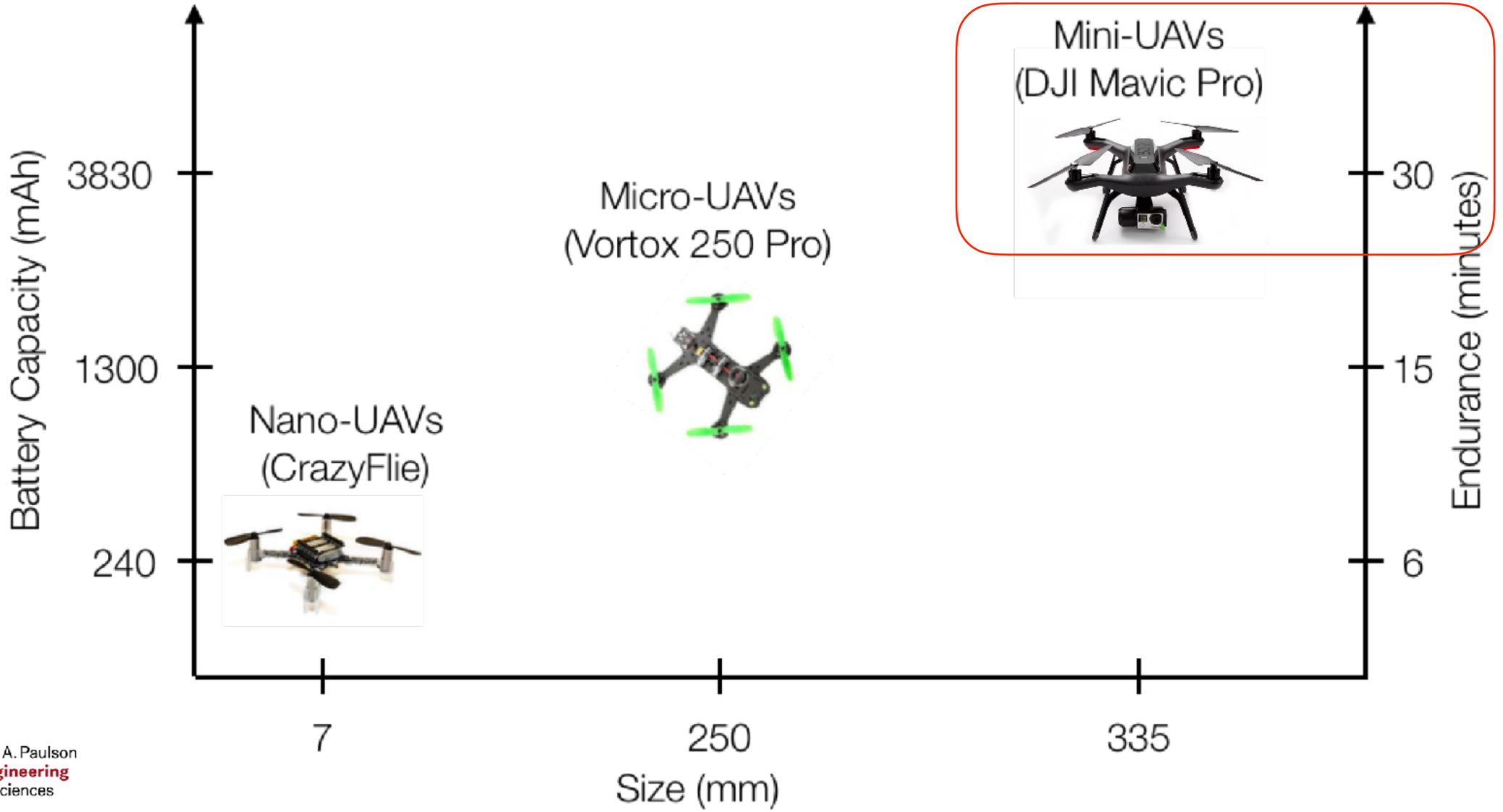
1000X

0.08 kWh

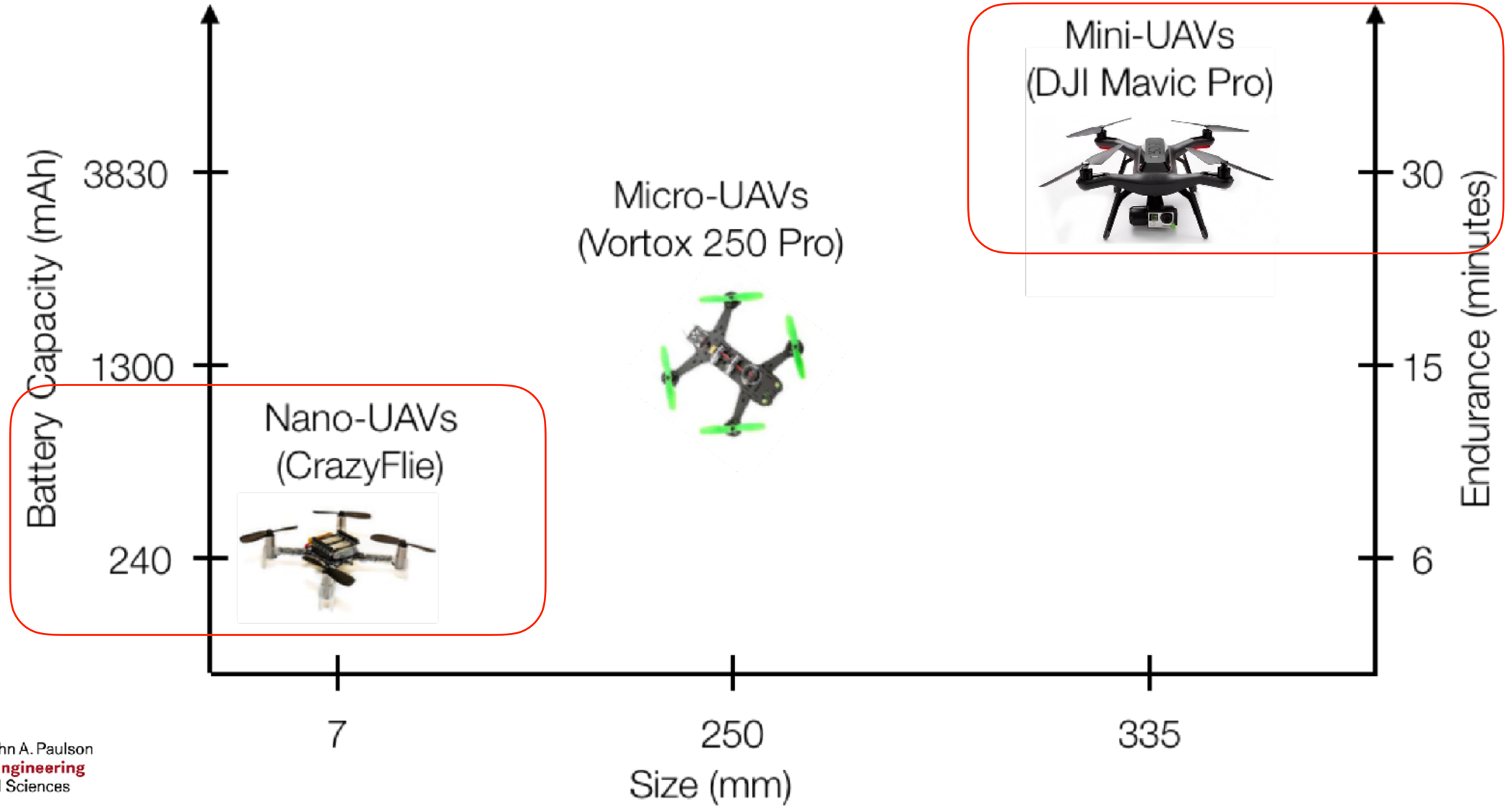
Q3. What are the technical challenges we need to solve? (Form Factor to Battery Capacity Relationship)



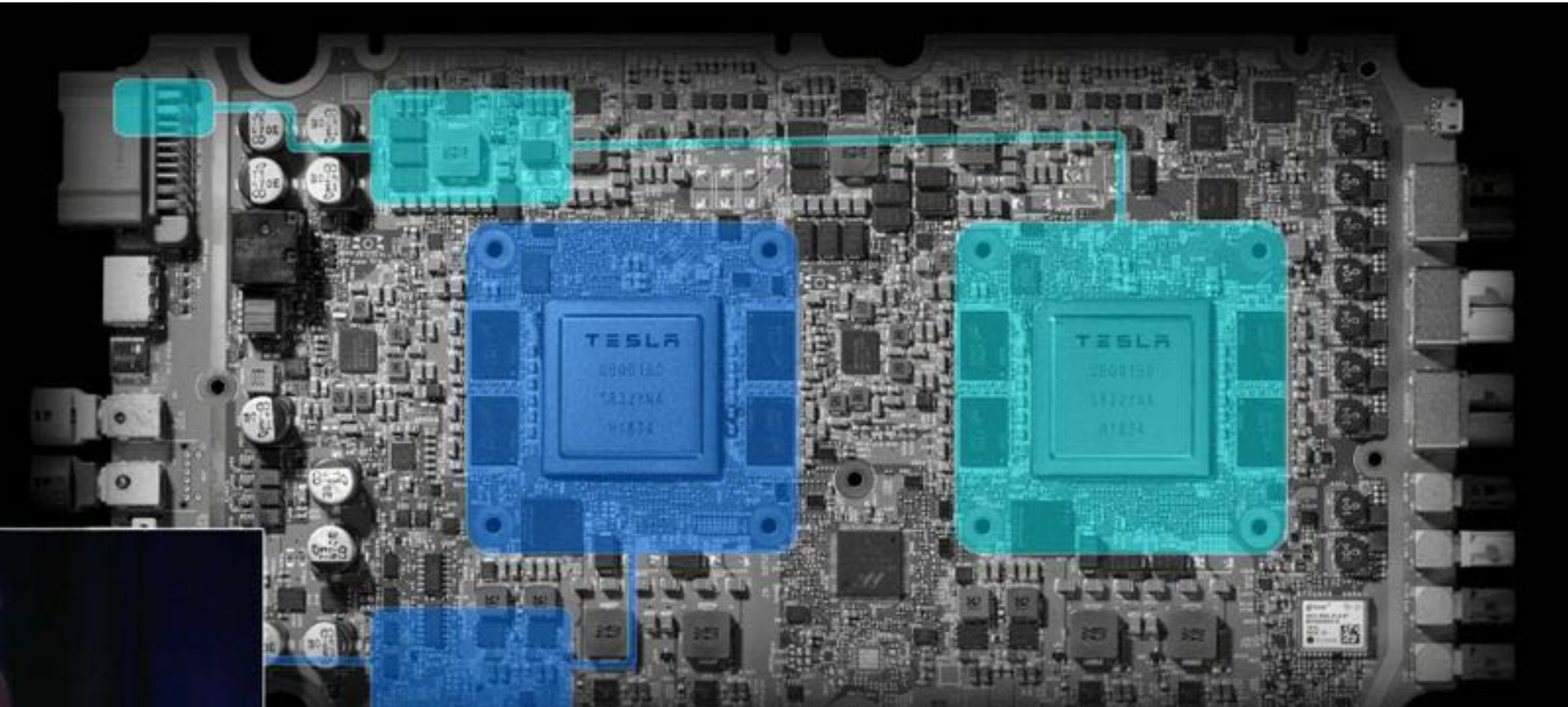
Q3. What are the technical challenges we need to solve? (Form Factor to Battery Capacity Relationship)



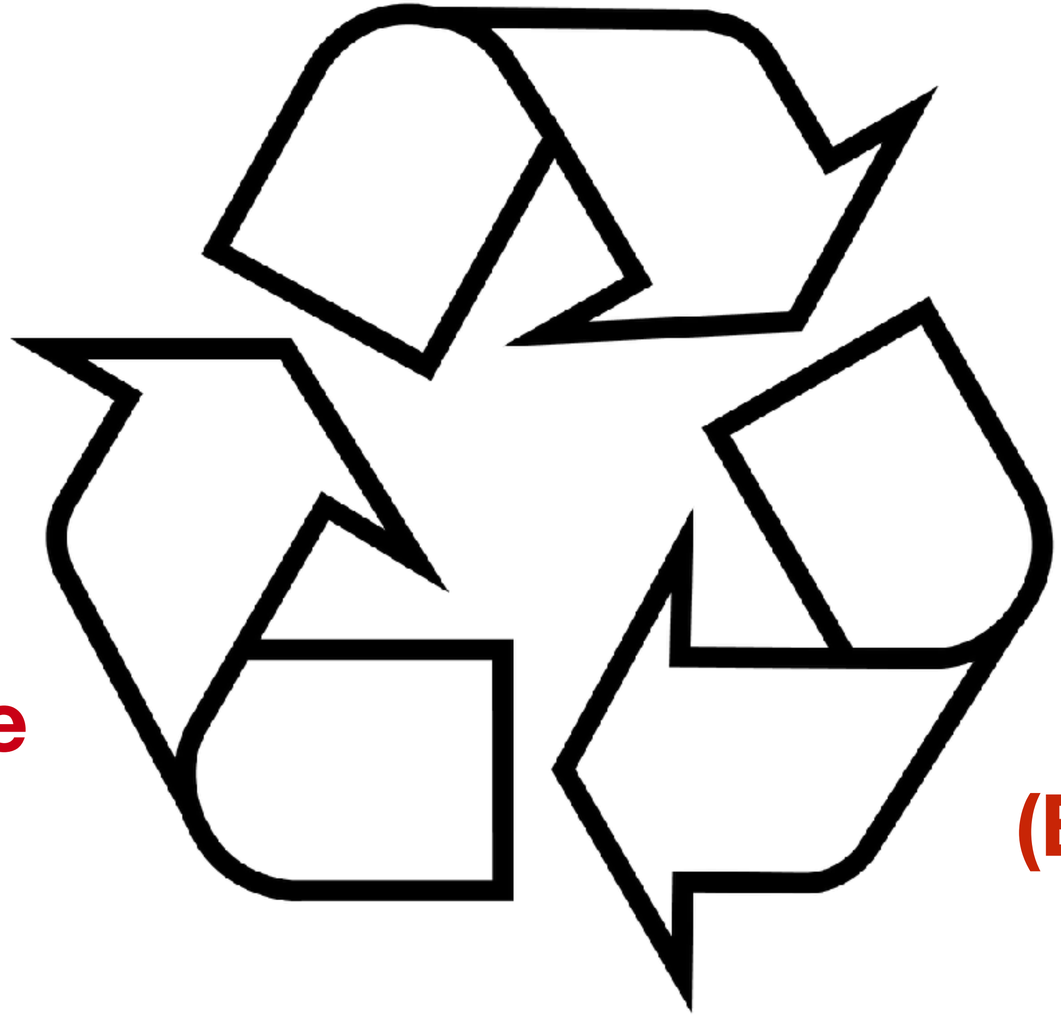
Q3. What are the technical challenges we need to solve? (Form Factor to Battery Capacity Relationship)



Q3. What are the technical challenges we need to solve?
Dual redundancy (sometimes even triple modular redundancy)



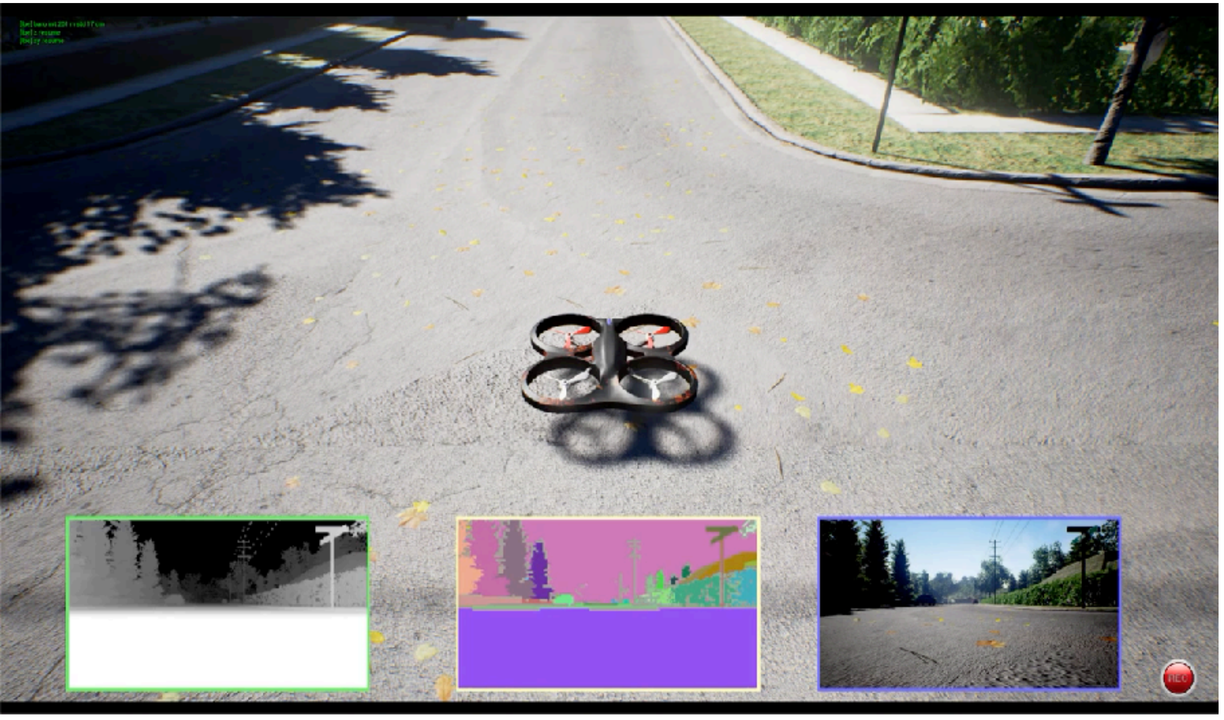
Reliability
(Safety)

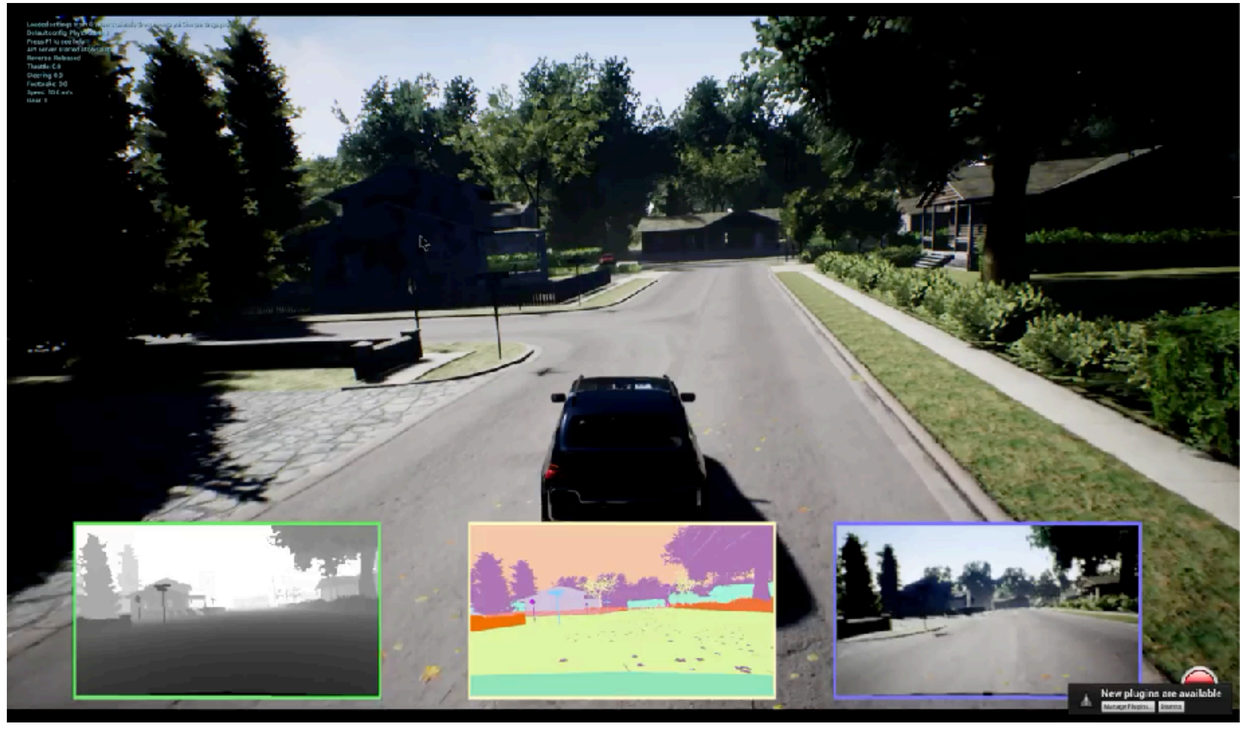


Performance
(Real-time)

Power
(Endurance)

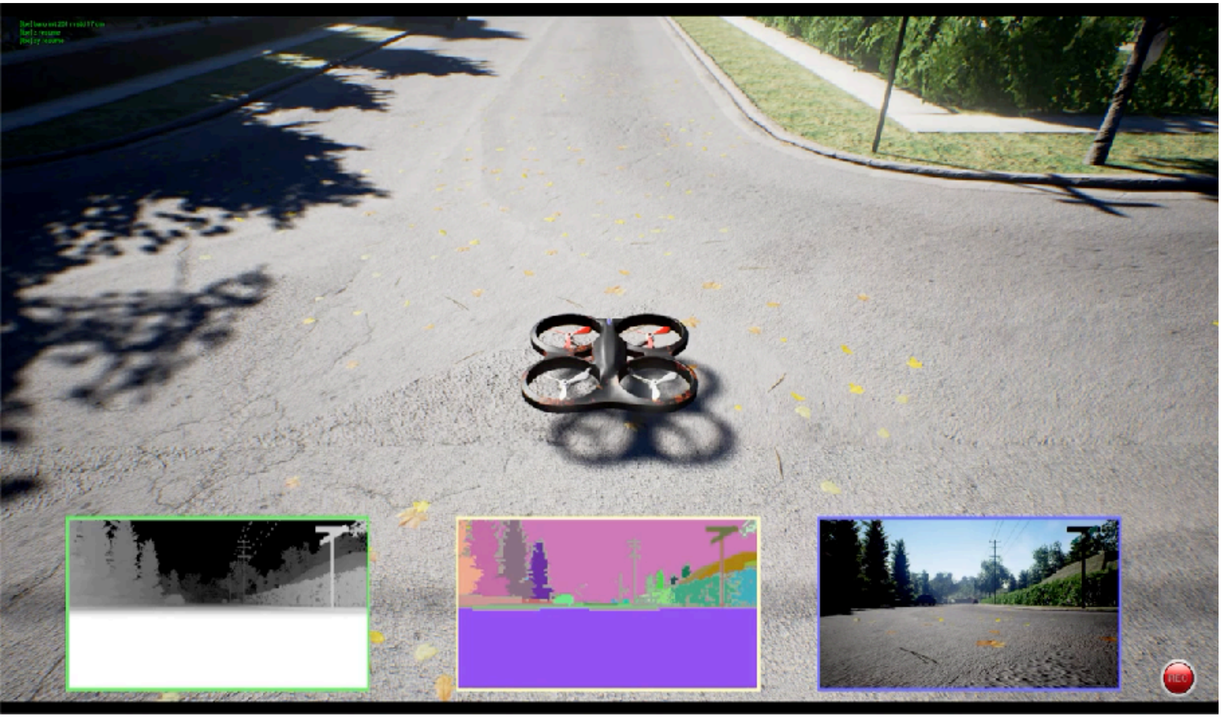
Q4. What research infrastructure is needed to enable studies? (**Simulators** for enabling cyberphysical system CPS codesign)

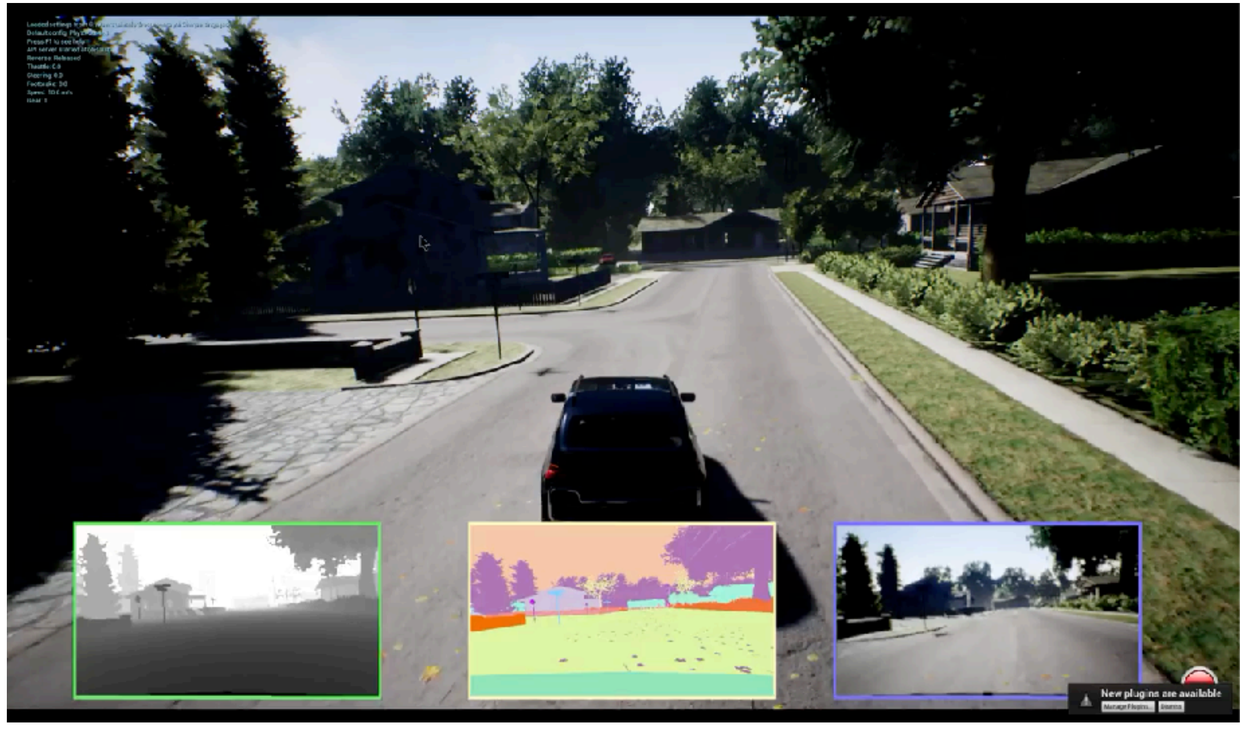




AirSim

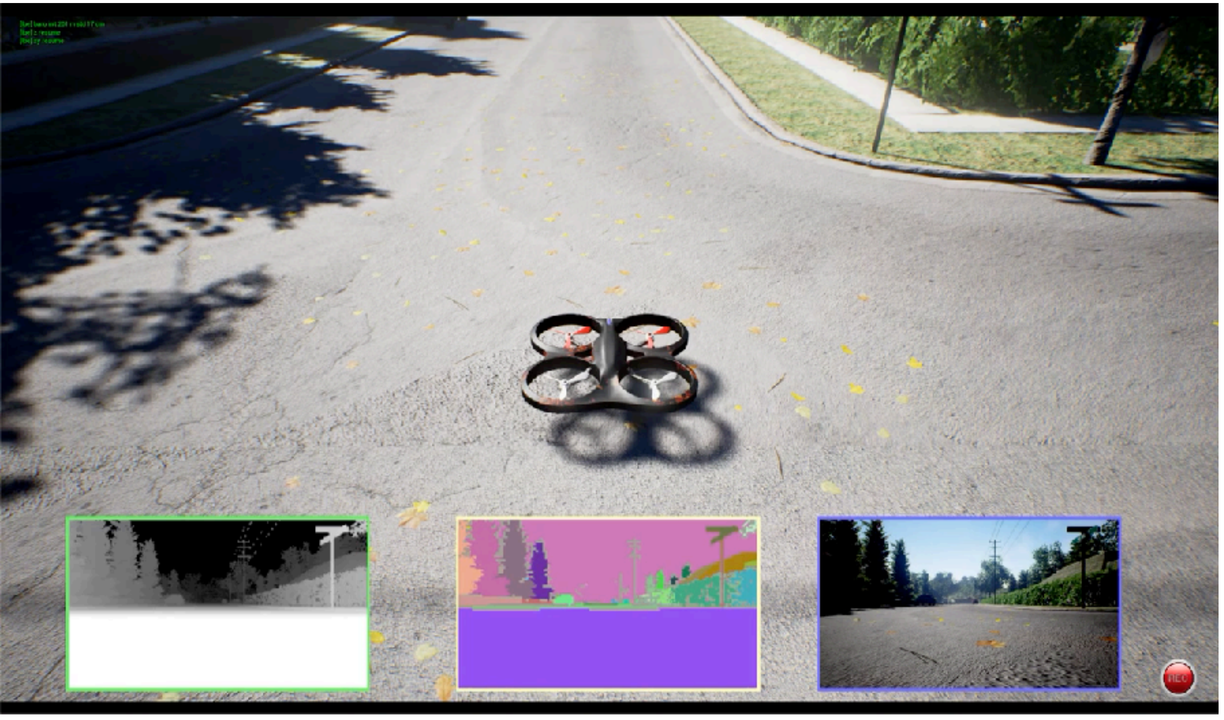
Q4. What research infrastructure is needed to enable studies? (**Simulators** for enabling cyberphysical system CPS codesign)

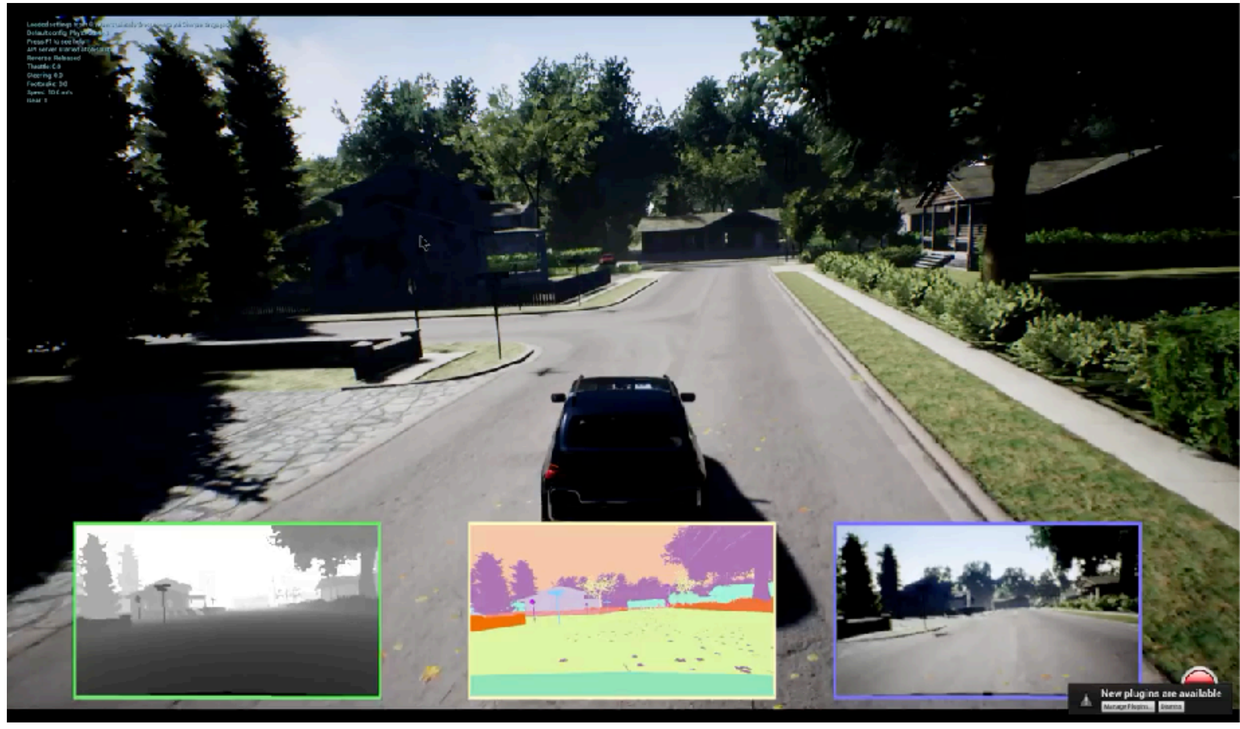




AirSim

Q4. What research infrastructure is needed to enable studies? (**Simulators** for enabling cyberphysical system CPS codesign)



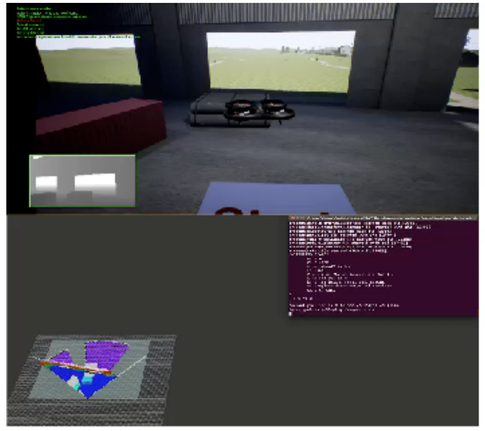


AirSim

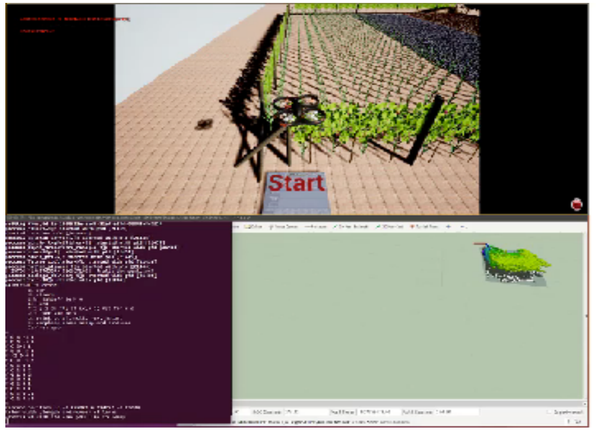
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

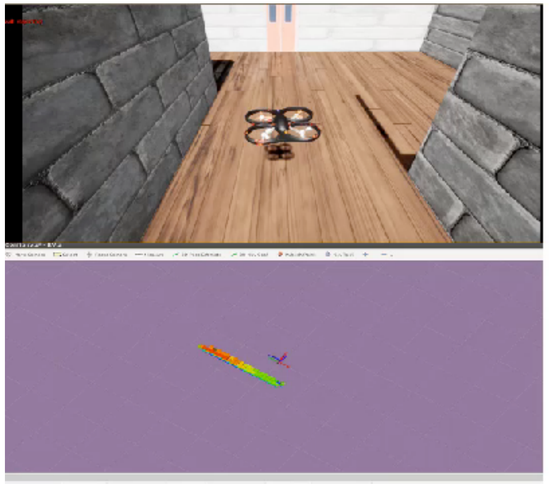
Package Delivery



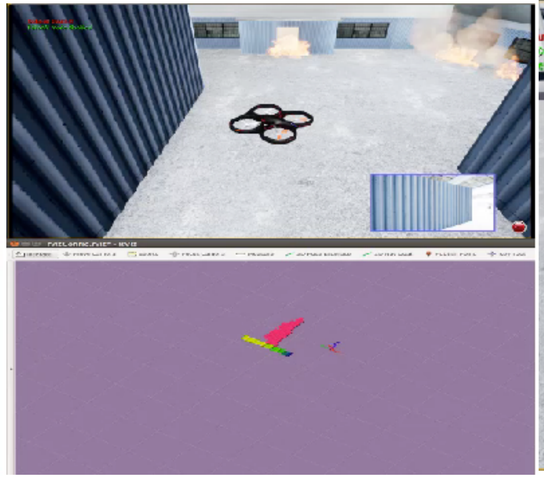
Surveying



3D Mapping



Search and Rescue

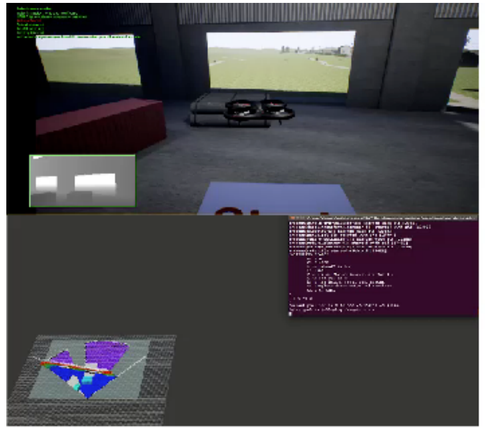


Aerial Photography

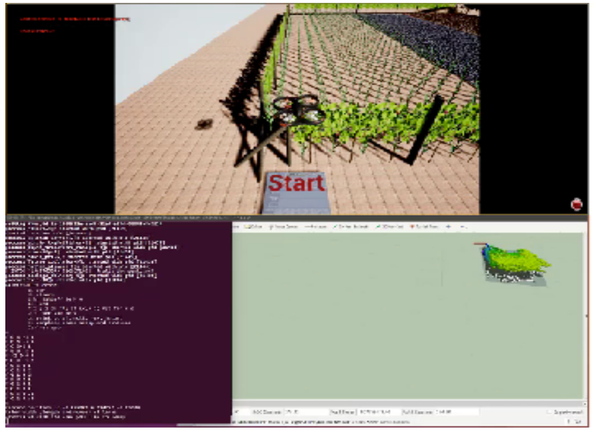
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

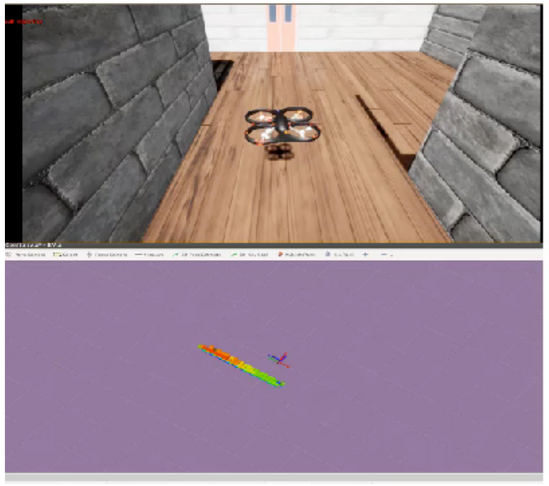
Package Delivery



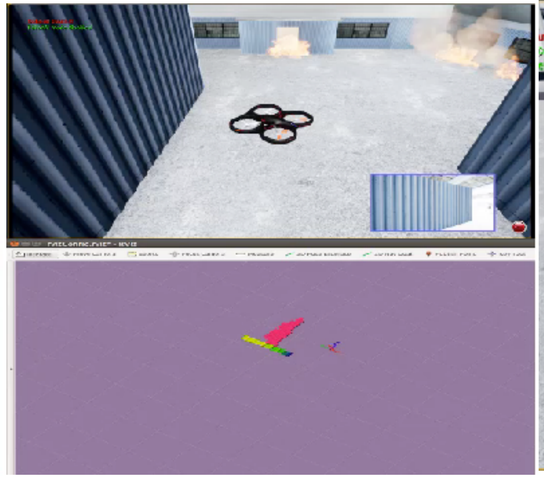
Surveying



3D Mapping



Search and Rescue

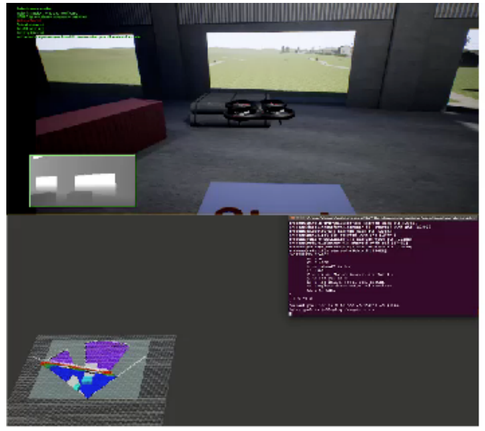


Aerial Photography

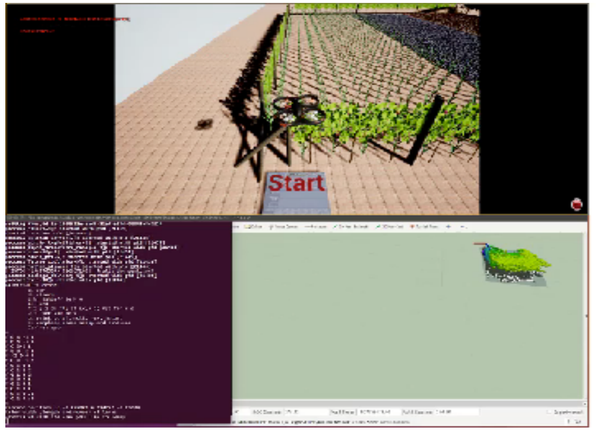
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

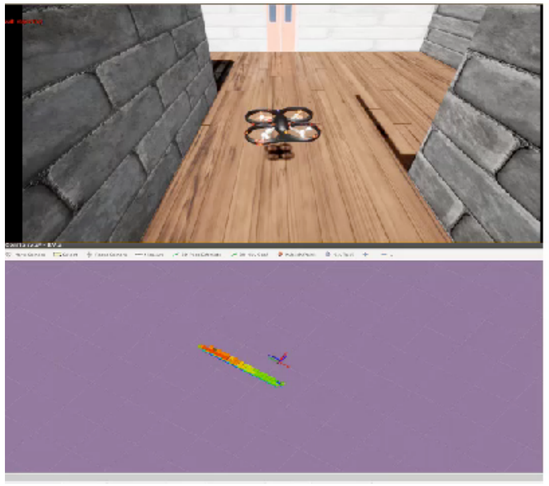
Package Delivery



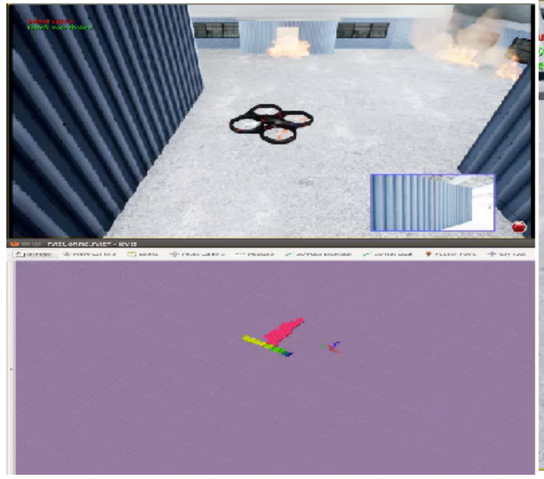
Surveying



3D Mapping



Search and Rescue

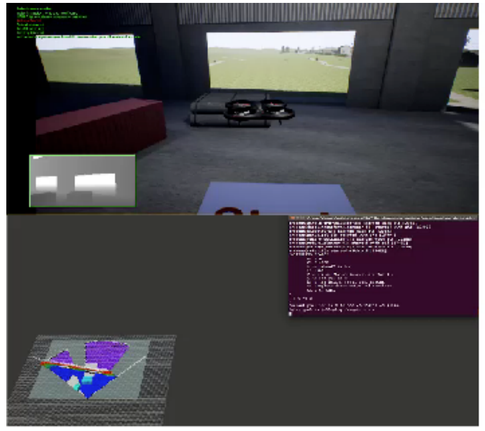


Aerial Photography

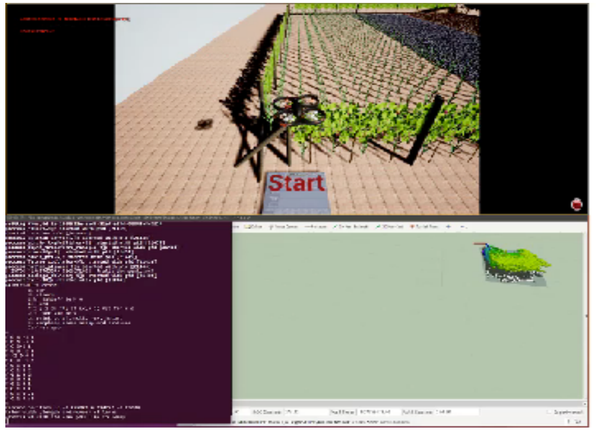
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

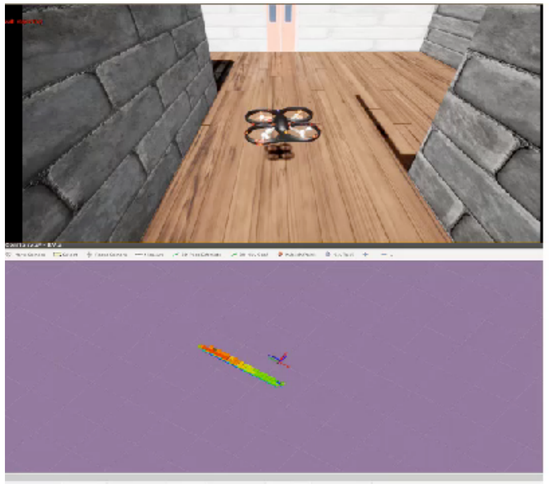
Package Delivery



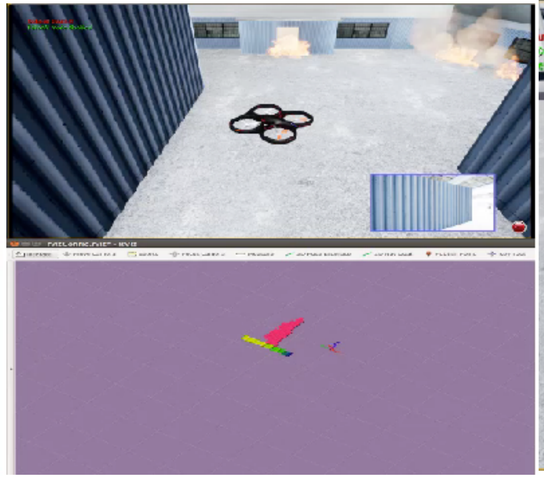
Surveying



3D Mapping



Search and Rescue

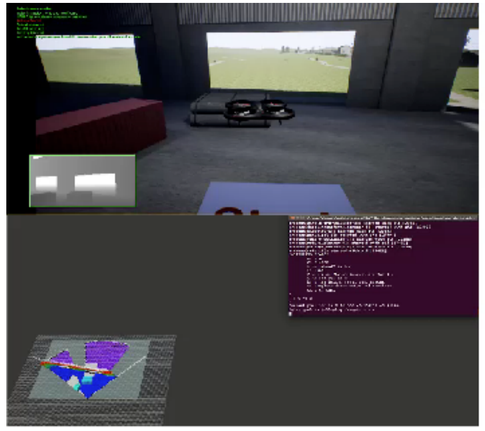


Aerial Photography

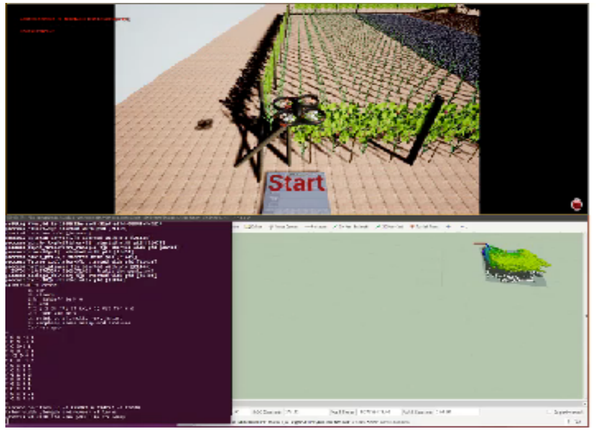
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

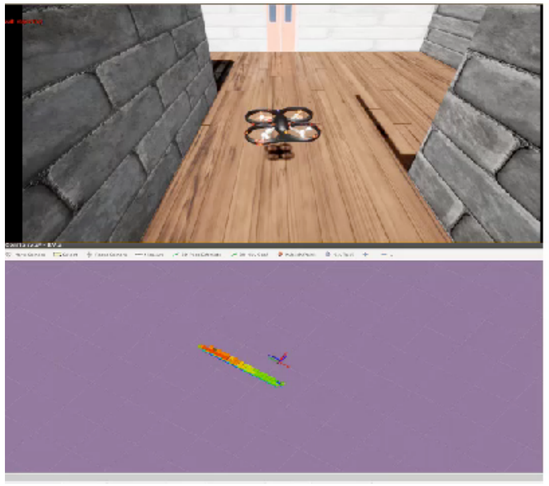
Package Delivery



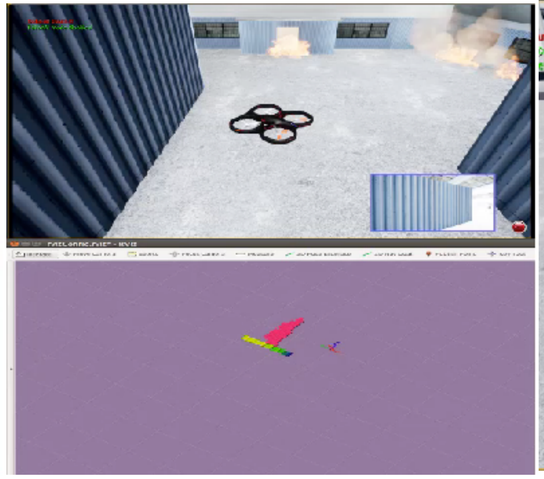
Surveying



3D Mapping



Search and Rescue

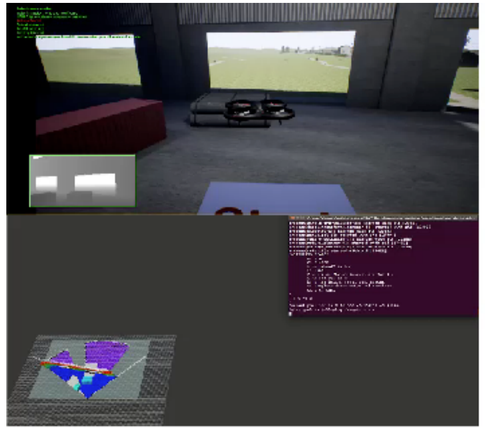


Aerial Photography

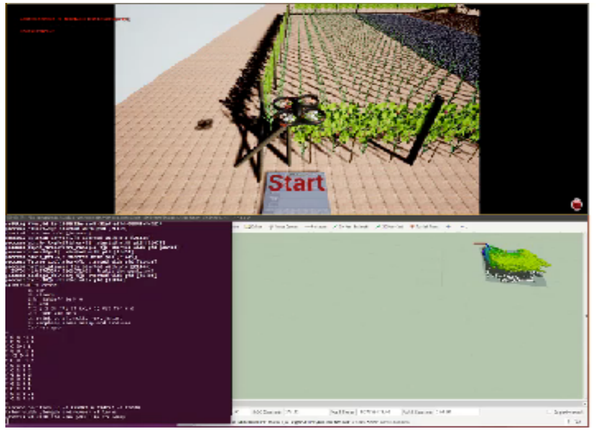
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

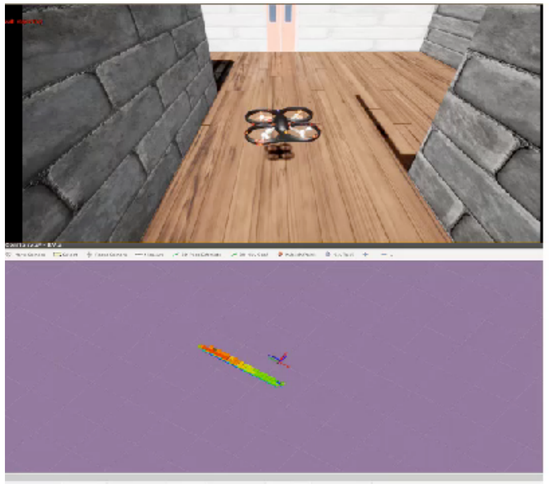
Package Delivery



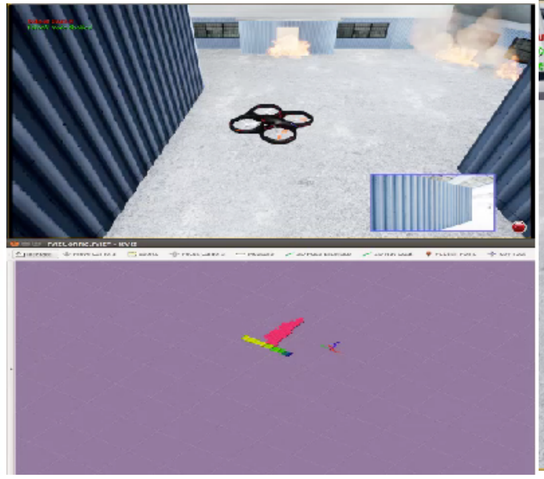
Surveying



3D Mapping



Search and Rescue

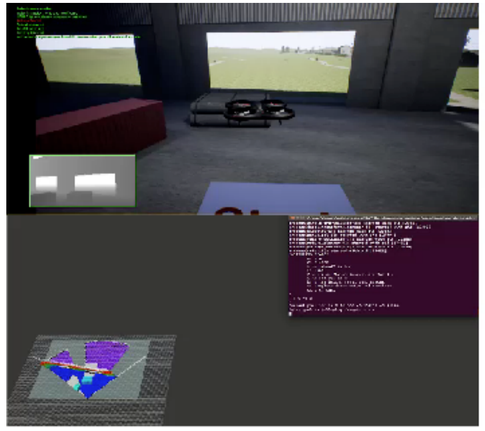


Aerial Photography

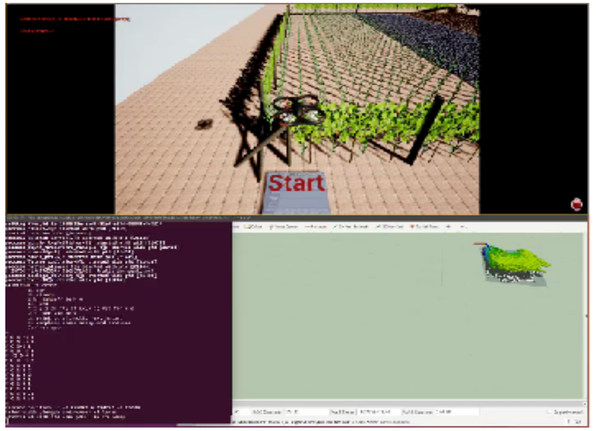
Q4. What research infrastructure is needed to enable studies? (**Benchmarks** for studying the systems systematically)

Indoor Navigation

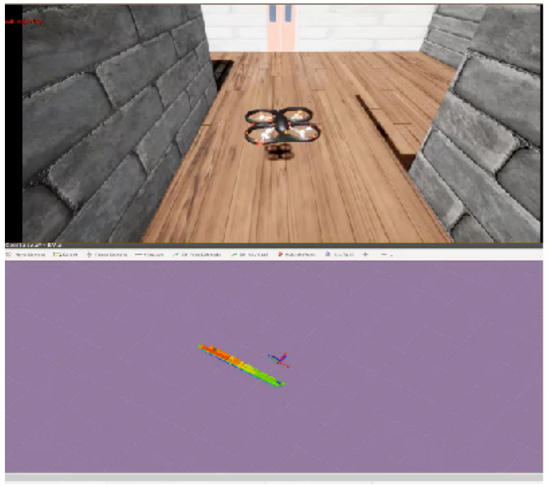
Package Delivery



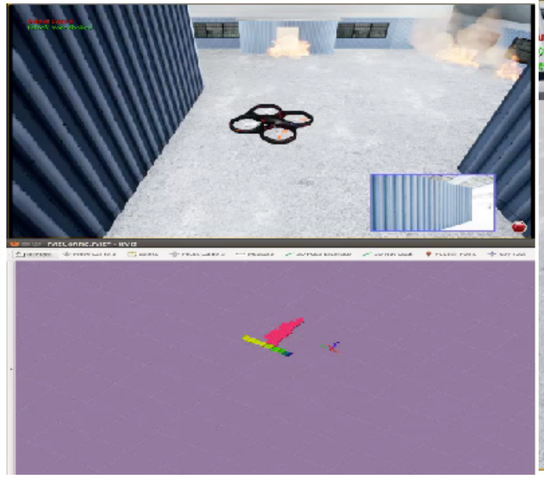
Surveying



3D Mapping

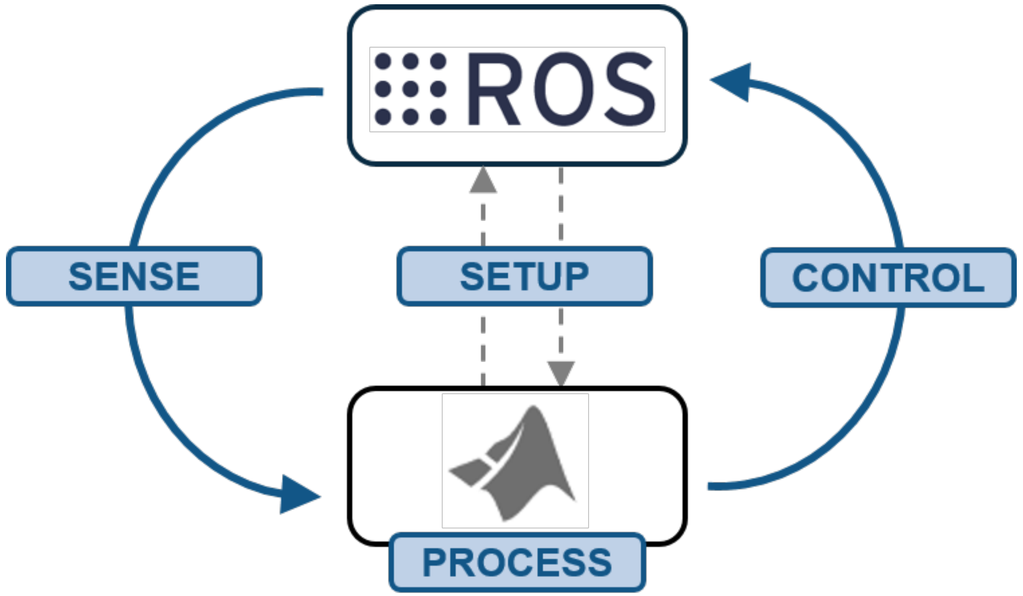


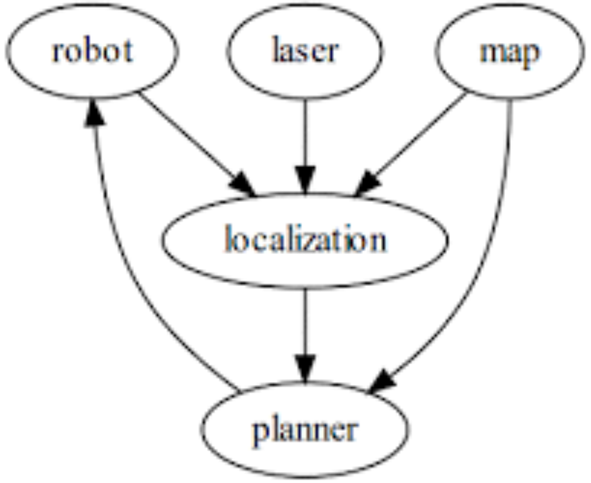
Search and Rescue



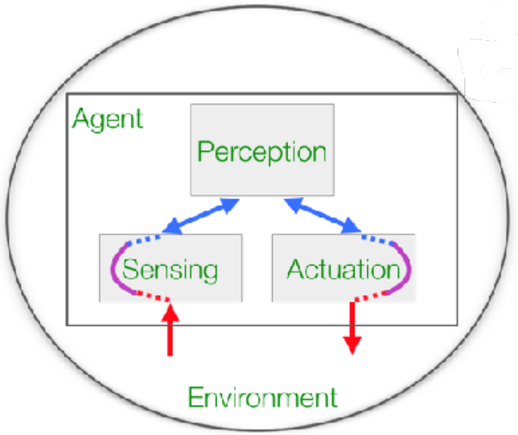
Aerial Photography

Q4. What research infrastructure is needed to enable studies? (**Middleware/Runtimes** for orchestrating the execution)





Q4. What research infrastructure is needed to enable studies? (**Hardware-in-the-loop** for orchestrating the execution)



AirSim

Unreal Engine

Sensory Data (RGBD, GPS)

Flight Control

Sensory Data (IMU)

Workload

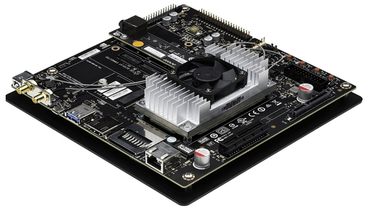
Robot Operating System

Companion Computer

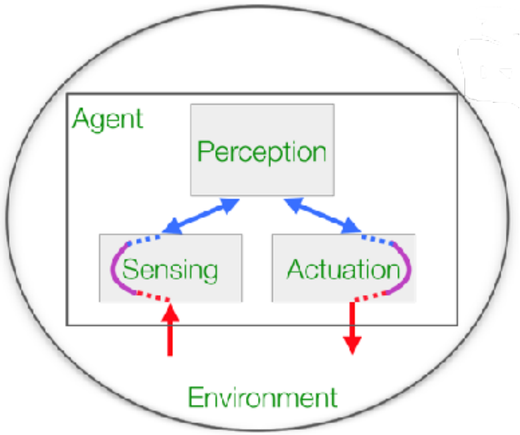
Flight Commands

Flight Stack

Autopilot Controller



Q4. What research infrastructure is needed to enable studies? (**Hardware-in-the-loop** for orchestrating the execution)



Drones, batteries, …

Environments (forest, urban, …)

Sensory Data (RGBD, GPS)

Flight Control

Sensory Data (IMU)

Applications, kernels, …

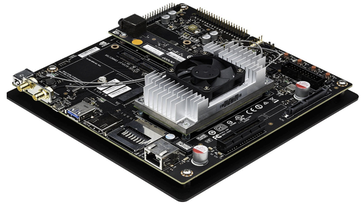
Scheduling, mapping, …

Platforms (TX2, Snapdragon, …)

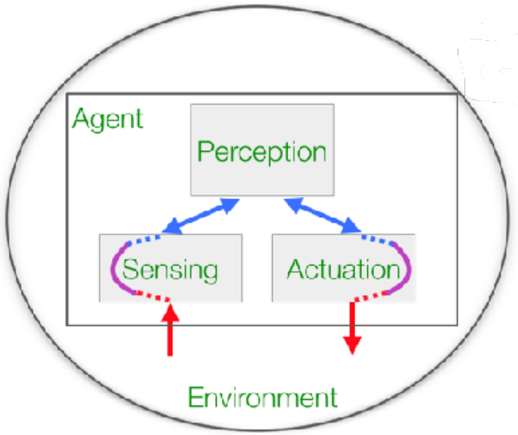
Flight Commands

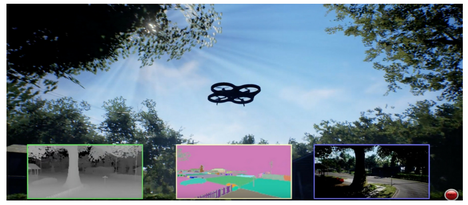
Flight Stack (PX4, ArduPilot, …)

MAVLink ISA



Q4. What research infrastructure is needed to enable studies? (**Hardware-in-the-loop** for orchestrating the execution)





Rotor energy, state of charge, …

Collisions, Proximity misses, …

Sensory Data (RGBD, GPS)

Flight Control

Sensory Data (IMU)

Profile information (hot code, …)

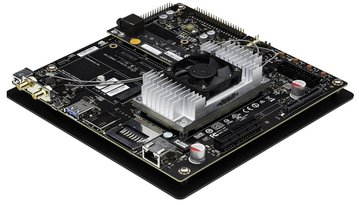
Throughput, bandwidth, …

Perf. Counters, Energy, …

Flight Commands

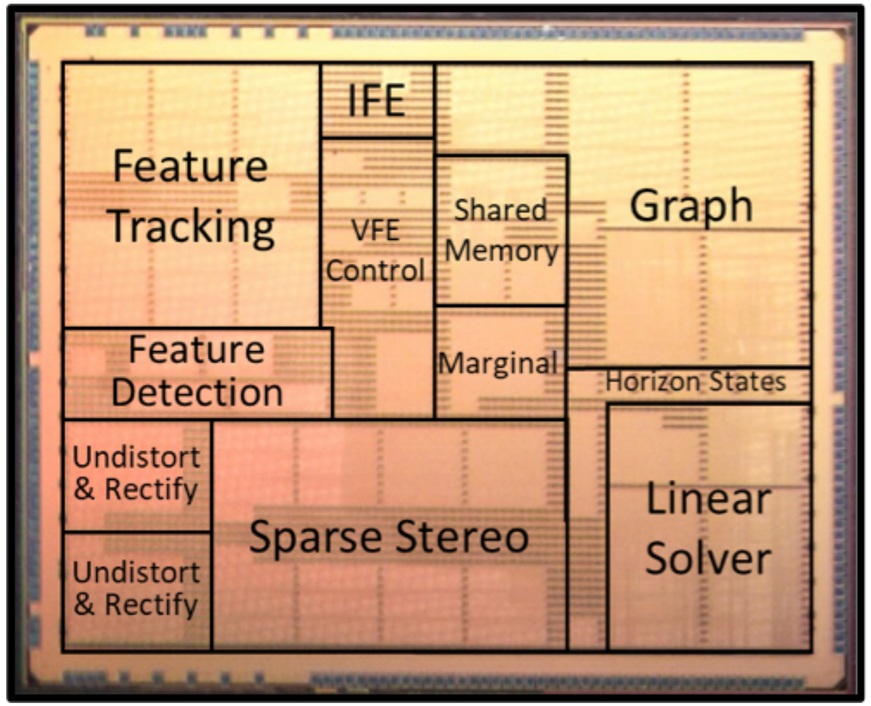
Profile information

Frequent commands, …



Q5. What opportunities can we unlock for the future of AI machines?

- Build **better algorithms**
- Build **better runtimes**
- Build **better hardware**



E.g. Navion: Enabling Autonomous Navigation of Miniaturized Robots (MIT)

Goals of the Class

Goals of the Class

1. What are **real-time autonomous machines**?
2. What is the **enabling technology** behind them?
3. What are the **technical challenges** we need to solve?
4. What **research infrastructure** is needed to enable studies?
5. What **new opportunities** can we unlock for the future of AI machines?

- Understand the **role of compute** in autonomous machines
- Learn to **bridge the gap via algorithm hardware co-design**
- Learn to **take *any* domain and specialize the compute system for it**