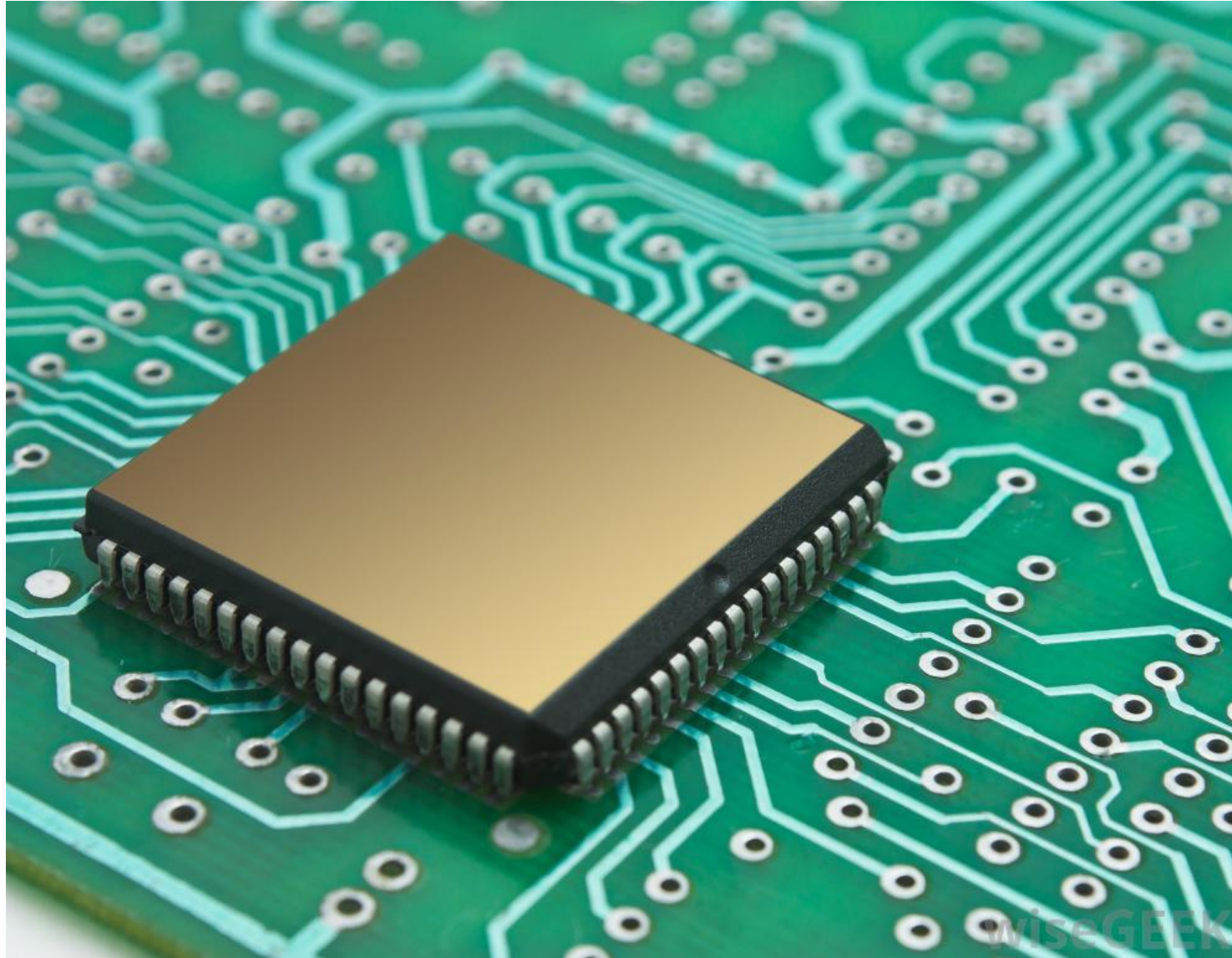

L1.3 HW/SW Platforms, Ethics

Zonghua Gu, Umeå University

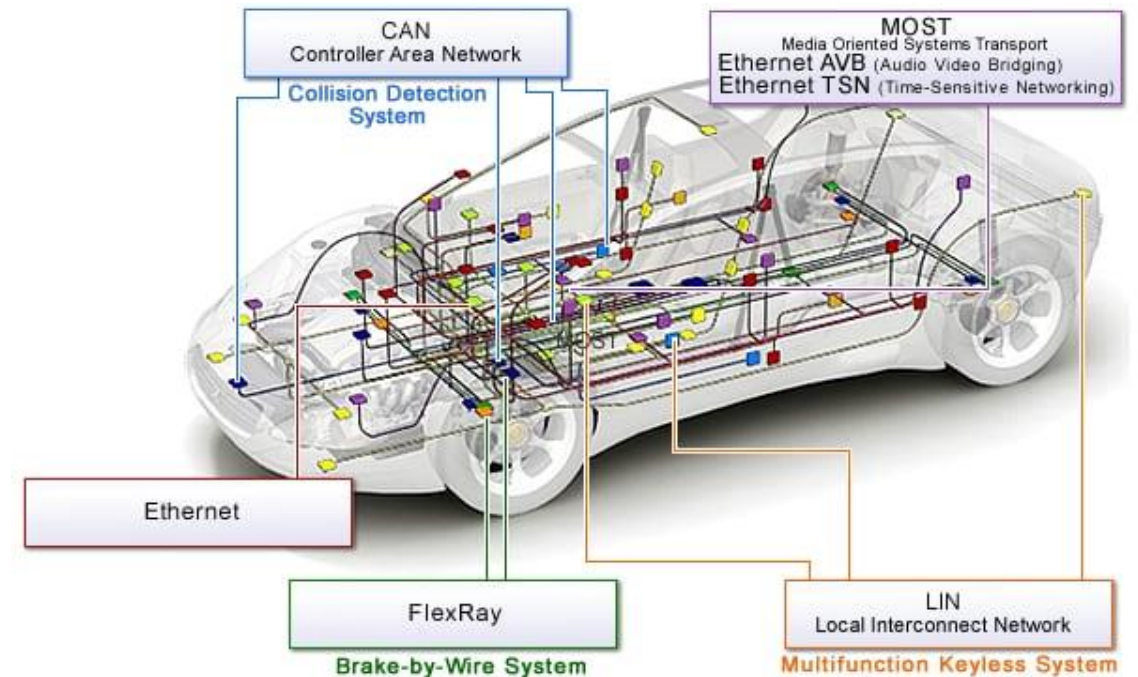
Nov. 2023

Hardware Platforms



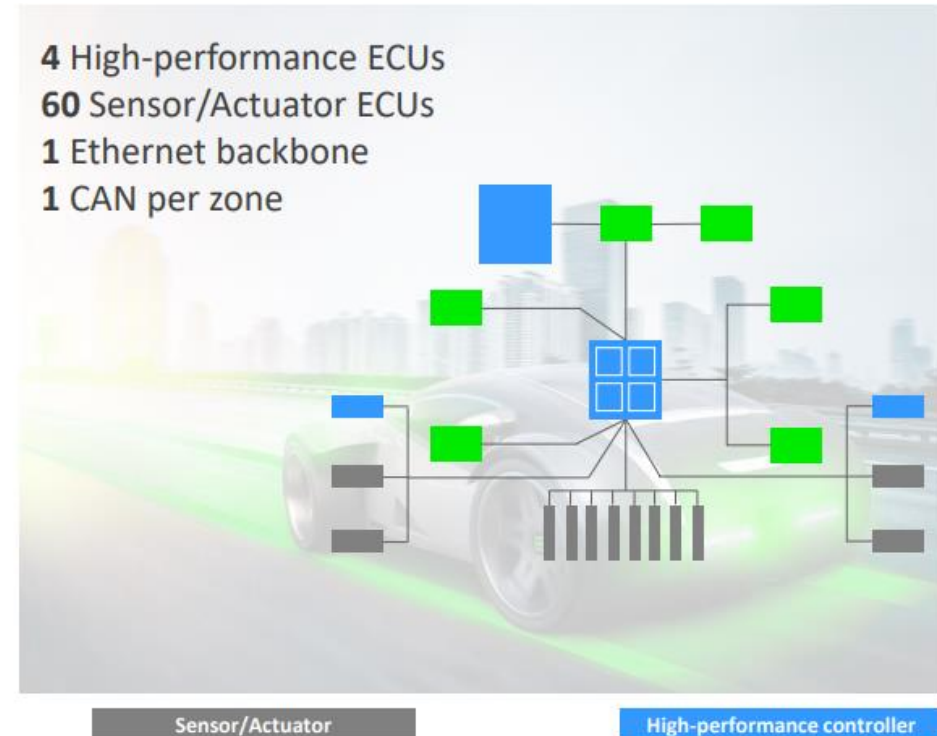
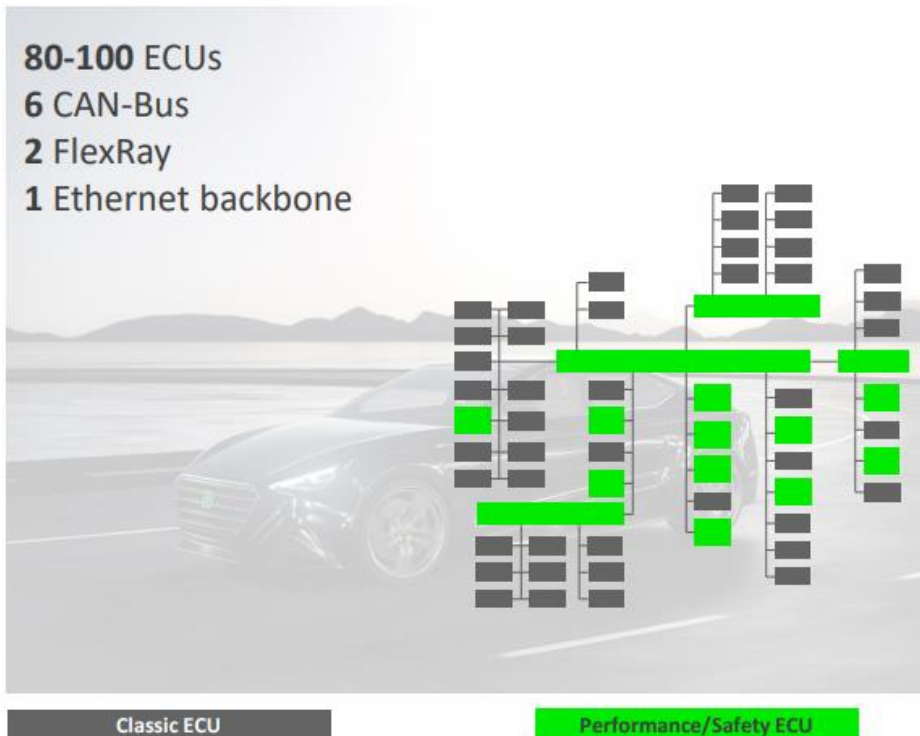
Automotive E/E Architecture

- Ethernet as high-bandwidth backbone network
 - Ethernet TSN posed to be the dominating standard protocol.
 - Regular Ethernet is also used for diagnostics
- FlexRay for safety-critical X-by-Wire, where X stands for brake, steer, drive...
 - Gradually being replaced by Ethernet
- Media Oriented Systems Transport (MOST) for multimedia transmission
 - Gradually being replaced by Ethernet
- CAN (Controller Area Network) for low-bandwidth network and interfacing with sensors/actuators
- LIN (Local Interconnect Network) for body electronics, e.g., door, light, rearview mirrors...



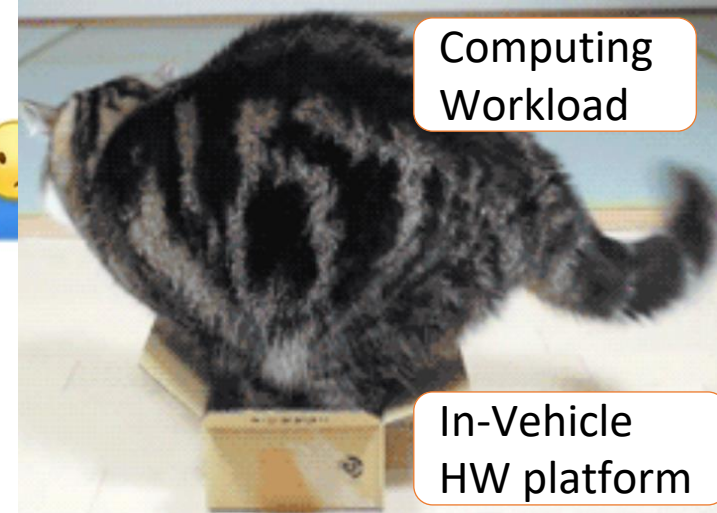
Evolution of Automotive E/E Architecture

- From many (~80-100) distributed and networked ECUs to a few (~4) high-performance ECUs with massive computing power, and large number of (~60) small ECUs for interfacing with sensors and actuators
- This helps simplify system architecture, reduce network load, and improve system reliability



Automotive E/E Architecture Trends

- Automotive E/E architecture trends:
 - HW platform consolidation
 - Centralized architecture with a few ECUs helps simplify system architecture, reduce network load, and improve system reliability
 - HW platform miniaturization
 - A trunkful of electronics hinders consumer acceptance and mass deployment
 - System-on-Chip technology helps to achieve high-performance computing with small form factor and low power consumption
- Significant HW resource constraints in terms of processor speed, memory size, and network bandwidth
 - Level of constraints depends on the application context, e.g., mobile robots and consumer drones have more severe constraints than passenger cars



Tesla's FSD Computer (2022)

Design Objectives vs. HW Constraints

- **Design Objectives**

- **Safety**

- System failures may be deadly.

- **Hard real-time**

- Deadline misses may compromise safety.

- **Security**

- Malicious attacks may compromise safety.

- **HW Constraints**

- **Size/Weight**

- Must be compact and lightweight (cannot have a trunkful of electronics)

- **Power consumption**

- Power consumption of electronics (sensors and computing hardware) for AD may be 100x that of a vehicle with regular ADAS. This drains battery and implies increased fuel consumption or reduced range for EVs

- **Cost**

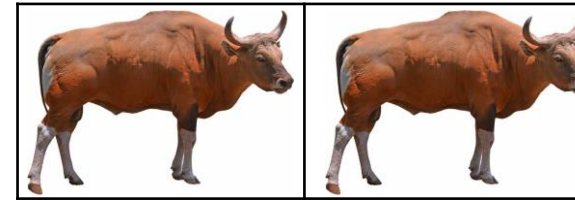
- Important for mass deployment.
- Cost of electronics in an experimental AV often exceeds cost of the original vehicle.

SoC Hardware for AVs

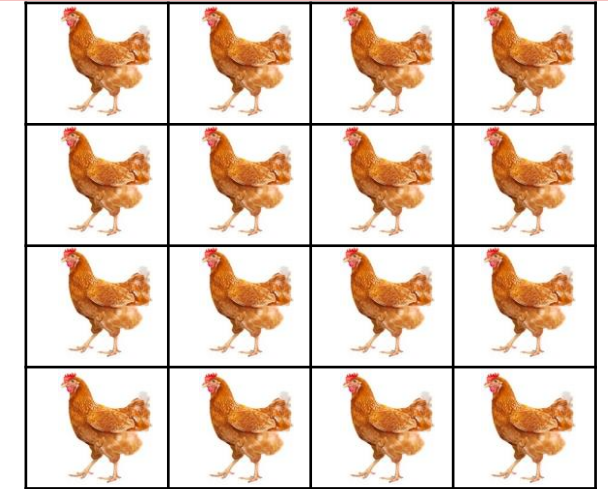
- The most compute-intensive workload is Deep Learning inference
- Many vendors provide SoC (System-on-Chip) products that integrate CPU cores with specialized computational engines for Deep Learning:
 - GPU (Graphics Processing Unit)
 - NVIDIA is the only serious player.
 - Other GPU vendors, e.g., AMD, ARM, Intel, focus on computer graphics instead of general-purpose computing (GPGPU).
 - FPGA (Field-Programmable Gate Arrays)
 - Xilinx, Intel Altera
 - ASIC (Application-Specific Integrated Circuit)
 - An explosion of specialized ASICs for Deep Learning in recent years, with hundreds of companies and products ranging from high-performance to embedded.
 - DSP (Digital Signal Processor)
 - Mainly for image preprocessing, e.g., products from Texas Instruments

CPU vs. GPU/FPGA/ASIC

- CPU is designed for general-purpose workloads
- GPU has much simpler control logic than CPU, hence has more computational elements (Arithmetic Logic Units)
 - Specialized for highly-parallel workloads, e.g., matrix-multiply, which is a core operation in Deep Learning training and inference
 - Similarly for FPGA and ASIC



CPU (2 oxen)

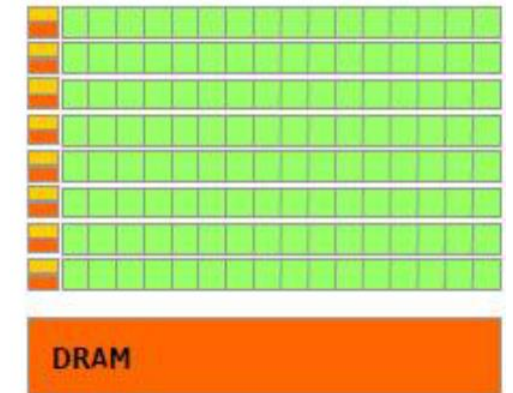


GPU (1024 chickens)

“If you were plowing a field, which would you rather use? 2 oxen, or 1024 chickens?” S. Cray.



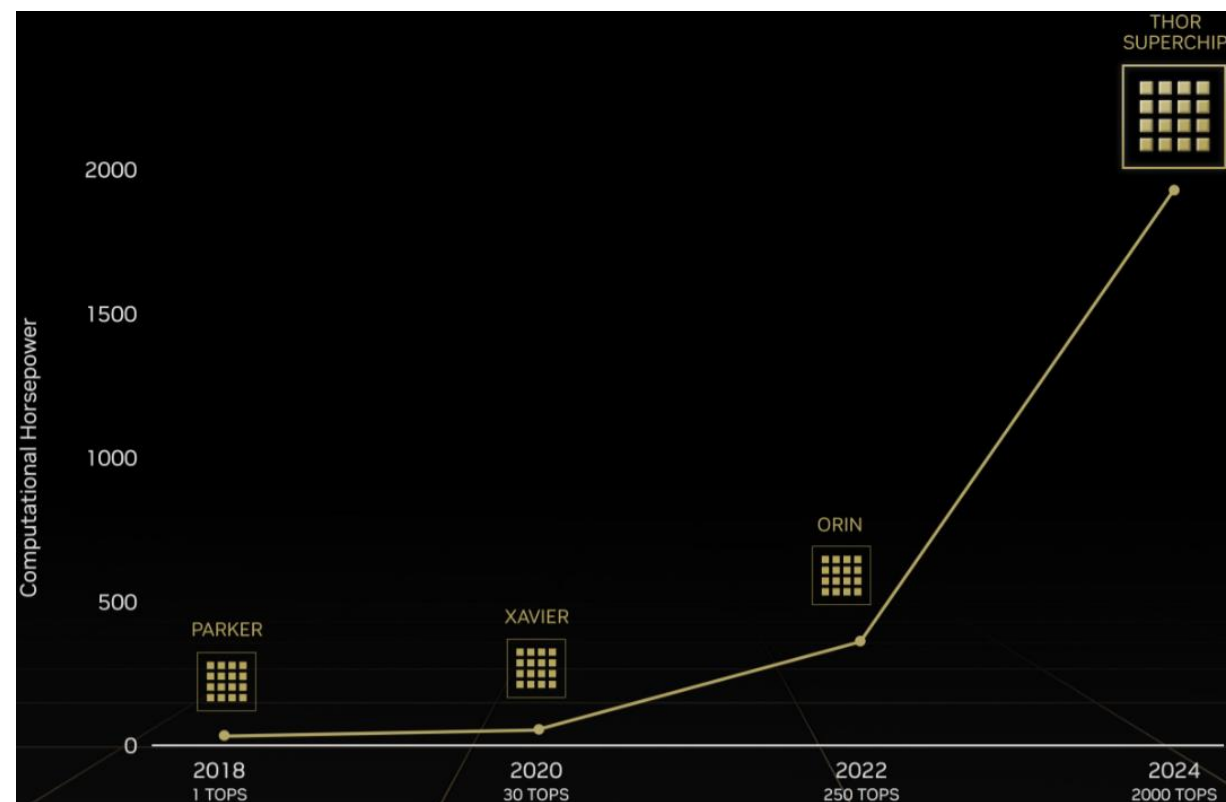
CPU



GPU

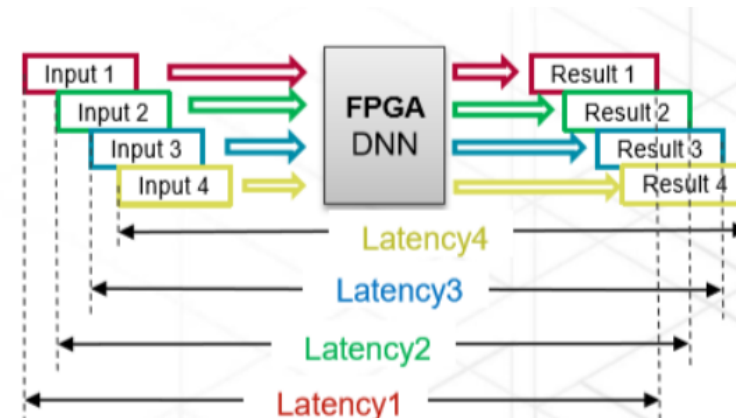
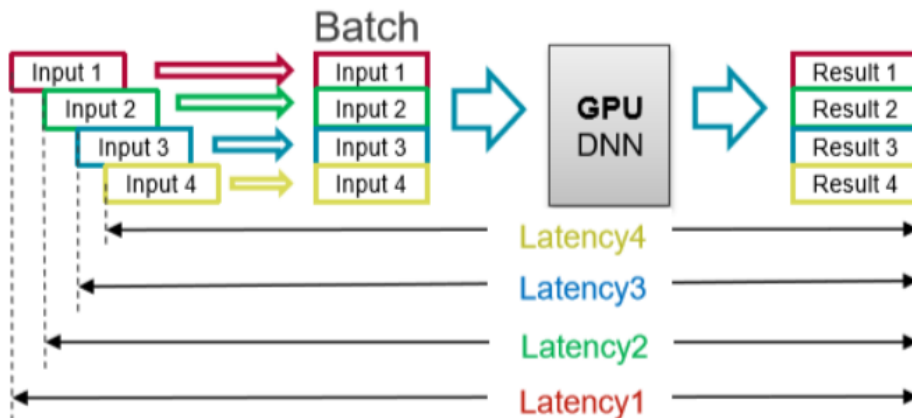
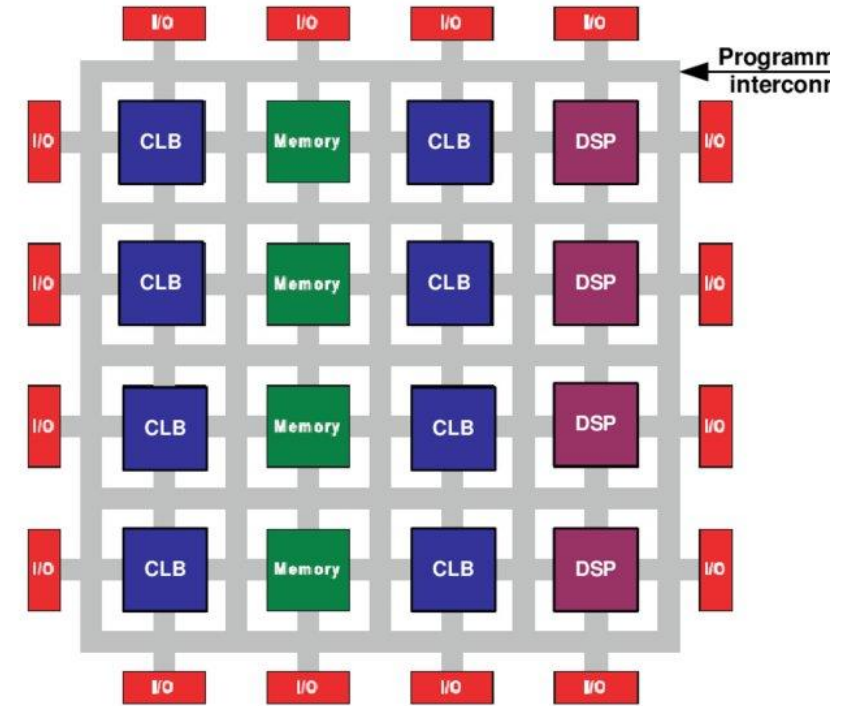
NVIDIA DRIVE GPU-based HW Platform

- A family of products, ranging from the low-end Parker to the latest high-end THOR with 2000 TOPS (Tera Operations Per Second)
- Besides CPU and GPU cores, also includes NVDLA (NVIDIA Deep Learning Accelerator), an ASIC for Deep Learning inference



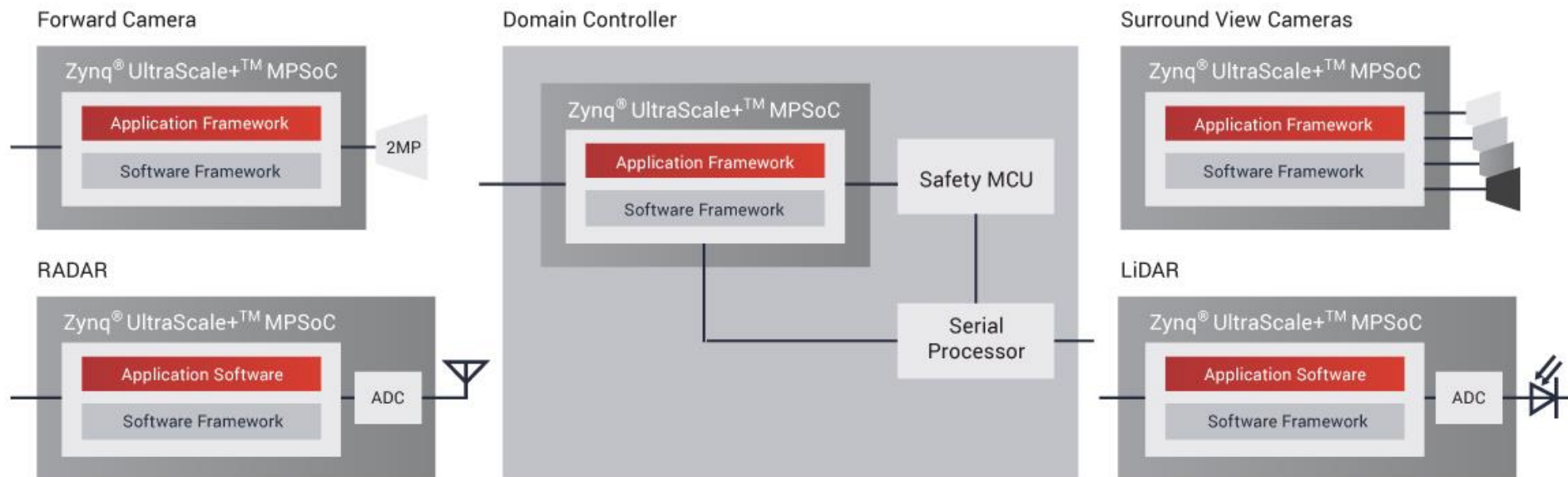
FPGA

- FPGA is reprogrammable hardware, consisting of an array of Configurable Logic Blocks (CLBs) and interconnections which can be configured at design time or runtime. FPGA has advantages over GPU for Deep Learning inference tasks
- GPU performs computation in batches for efficient exploitation of SIMD (Single Instruction, Multiple Data) computation model
 - This is ideally suited for training tasks, with well-known algorithms such as Stochastic Gradient Descent with mini-batches
 - But not ideal for inference tasks
 - Larger batch size leads to high throughput, but also high and nondeterministic latency for each data item
 - Smaller batch size leads to low computation efficiency
- FPGA can perform “batch-less” inference
 - Low and deterministic latency for any batch size



FPGA for Automotive

- FPGAs can be integrated into smart sensors (camera, Lidar, radar), or serve as central compute engine in domain controller or AD computer
 - Xilinx FPGAs have 90% market share in Lidar signal processing
 - Intel went into the automotive market, with acquisition of Altera in 2015



ASIC

- ASICs for Deep Learning are often called Neural Processing Units or AI accelerators
- ASICs, thanks to dedicated circuit design, may achieve up to 10x in computation efficiency and power consumption compared to CPU/GPU, and less dramatic, but still significant improvement compared to FPGA. The drawback is loss of programmability and flexibility.
- Almost every chip vendor provides some kind of AI accelerator, e.g. Google's Tensor Processing Unit (TPU)

MobileEye EyeQ Series

- ASICs for computer vision

2018



- > MID - ADAS front camera
- > High – ADAS/AD Trifocal

M: 1.1 TOPs @ 4.5W
H: 2.2 TOPs @ 6.5W

100°/1.7MP
52°/1.3MP



9

2021



- > MID - ADAS/AV front camera
- > High – AD surround

M_L: 7 TOPs @ 7.5W
M: 12 TOPs @ 17W
H: 24 TOPs @ 34W

120°/8MP
Front + Rear
Super Vision



2024/5



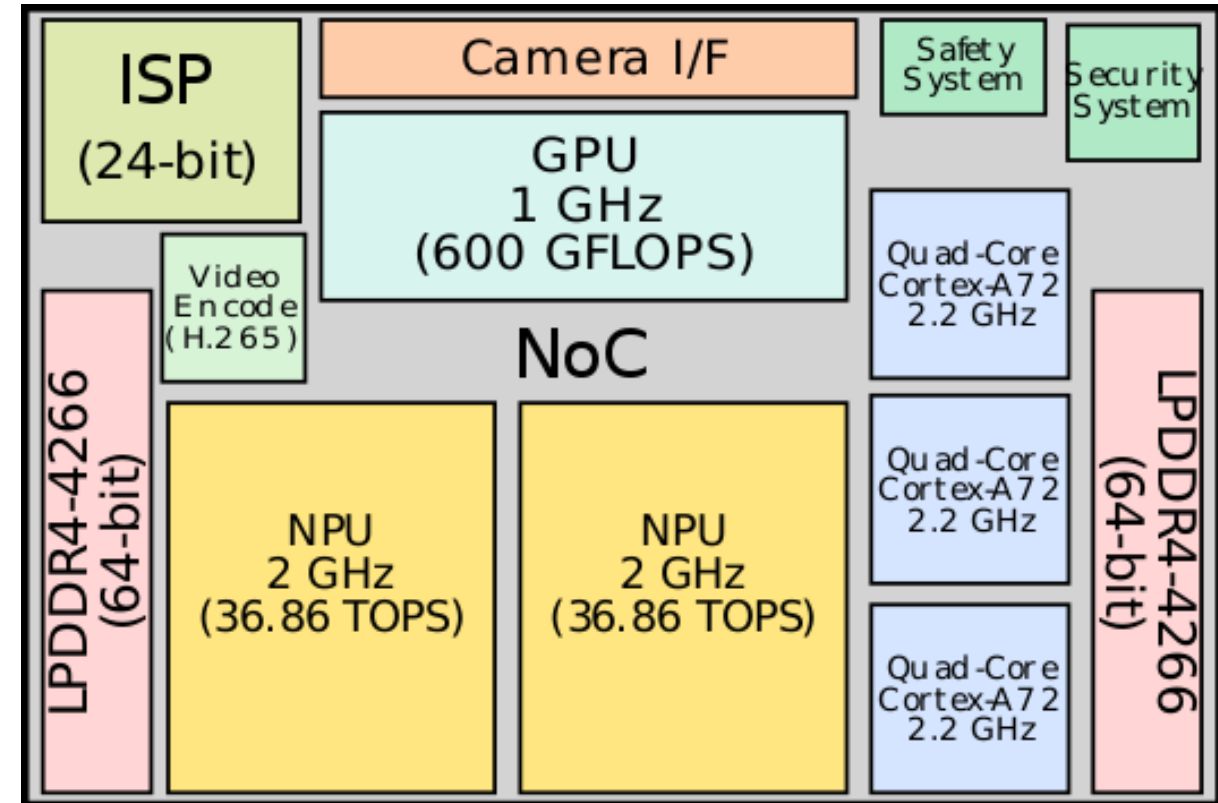
- > Light – ADAS/AD front camera
- > High – AD Surround & Visualization

L : 7.7TOPs@4.5W
H_L: 42TOPs @ 17W
H : 67TOPs @ 35W

120°/8MP
Front + Rear
Super Vision
+ Redundancy for L4

Tesla FSD Chip

- It incorporates 3 quad-core Cortex-A72 clusters for a total of 12 CPUs operating at 2.2 GHz, a GPU operating 1 GHz, 2 Neural Processing Units (NPUs) operating at 2 GHz, and various other hardware accelerators.
 - NPU: ASIC for Deep Learning inference

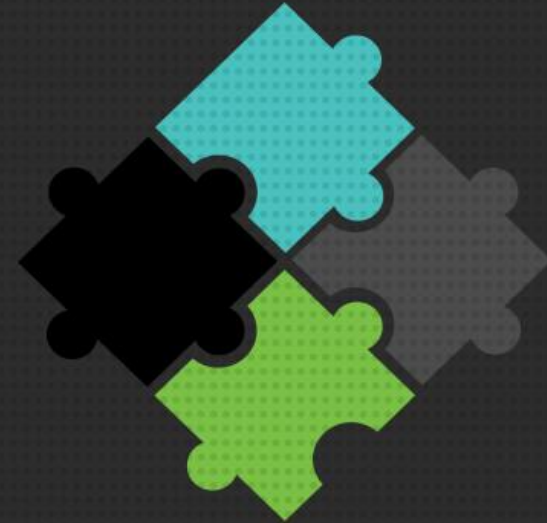


Products from Automotive OEMs and Suppliers

- Some companies offer integration solutions based on products from chip vendors like NVIDIA.
 - Delphi/Audi zFAS (zentrales Fahrerassistenz-Steuergeraet)
 - based on NVIDIA Tegra K1 and Mobileye EyeQ3
 - ZF ProAI
 - based on NVIDIA DRIVE PX2
 - Bosch AI Car Computer
 - based on NVIDIA DRIVE AGX Xavier
 - Many others
 - Continental ADCU; Visteon DriveCore; NXP BlueBox; Renesas R-Car...

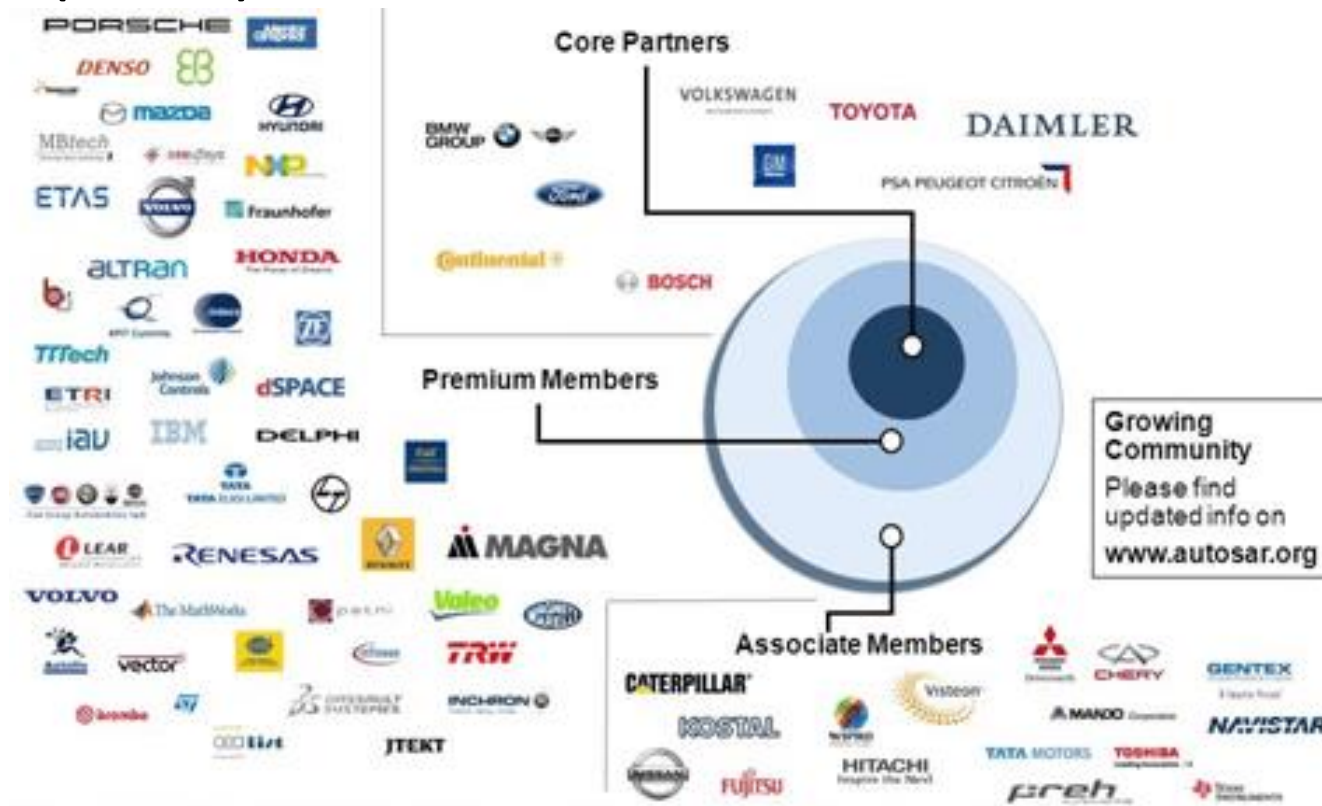
Software Platforms

SOFTWARE: **PRODUCT VS PLATFORM**



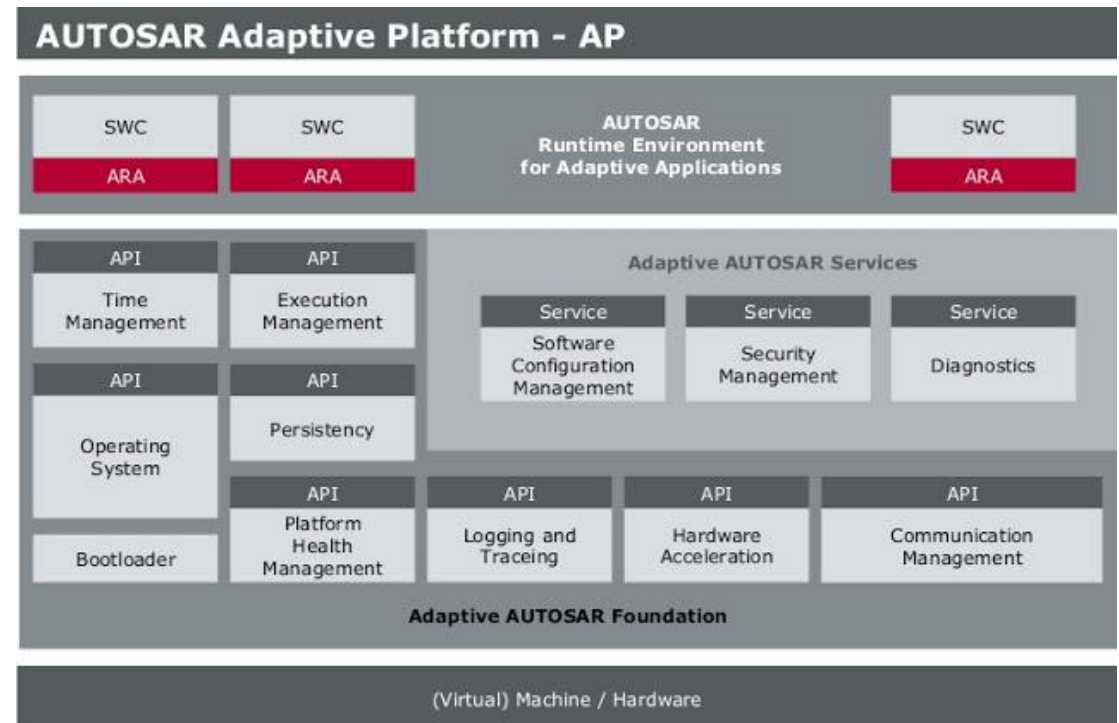
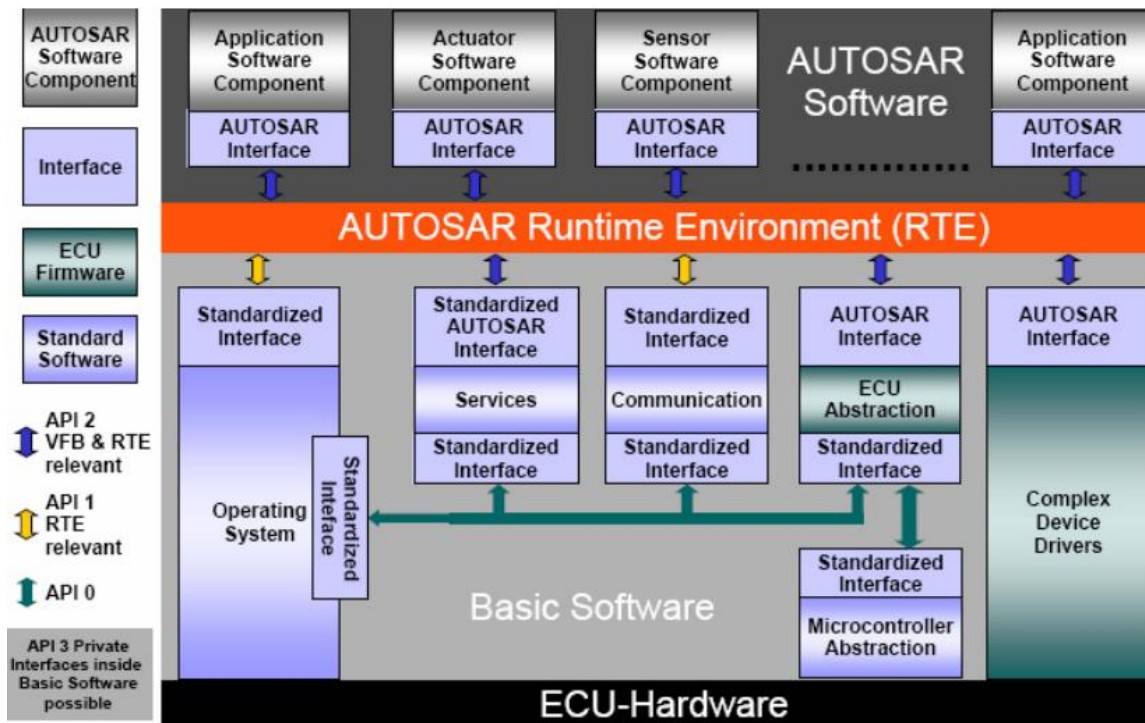
AUTOSAR Consortium

- AUTomotive Open System ARchitecture (AUTOSAR) is a global development partnership of automotive interested parties founded in 2003. It pursues the objective to create and establish an open and standardized software architecture for automotive Electronic Control Units (ECUs)



AUTOSAR Classic and Adaptive Platforms

- AUTOSAR-CP (Classic Platform) is an industry standard for resource-constrained safety-critical ECUs
- AUTOSAR-AP (Adaptive Platform) is an industry standard for high-performance multicore automotive ECUs
 - AUTOSAR-AP allows dynamic linking of services and clients during ECU runtime, which facilitates Over-the-Air (OTA) Update



Integration of Multiple SW Platforms

- AUTOSAR CP (labeled C) is used for safety-critical ECUs for low-level control and interfacing with actuators
- AUTOSAR AP (labeled A) is used for high-performance AD computer.
- Non-AUTOSAR (labeled N) may be Linux or Android, for non-safety-critical IVI (In-Vehicle Infotainment) and COTS (Commercial Off-the-Shelf) applications.

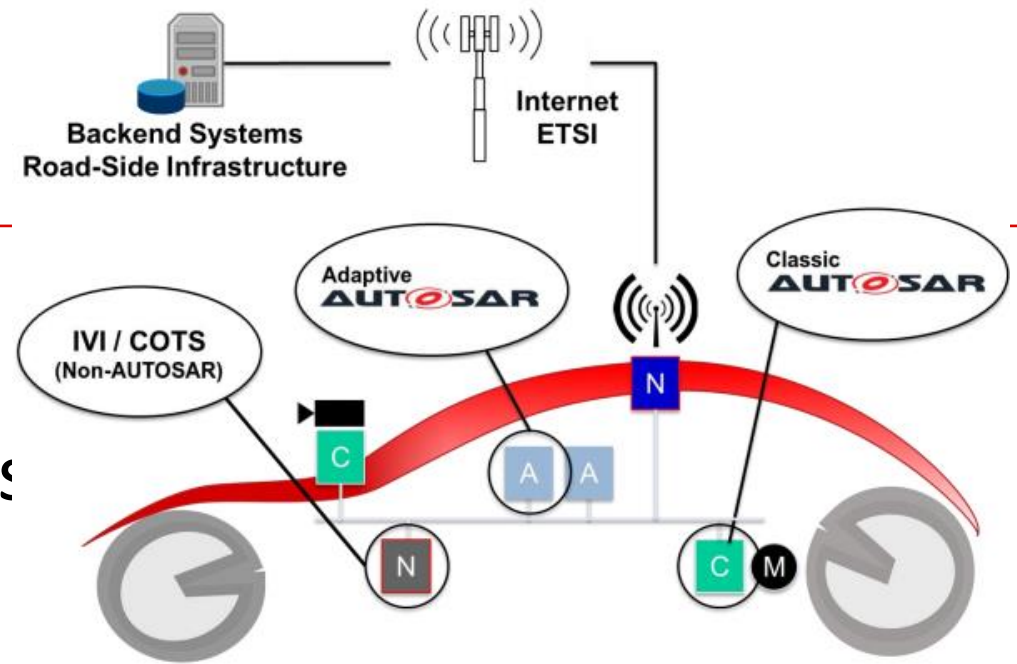


Figure 2-1 Exemplary deployment of different platforms

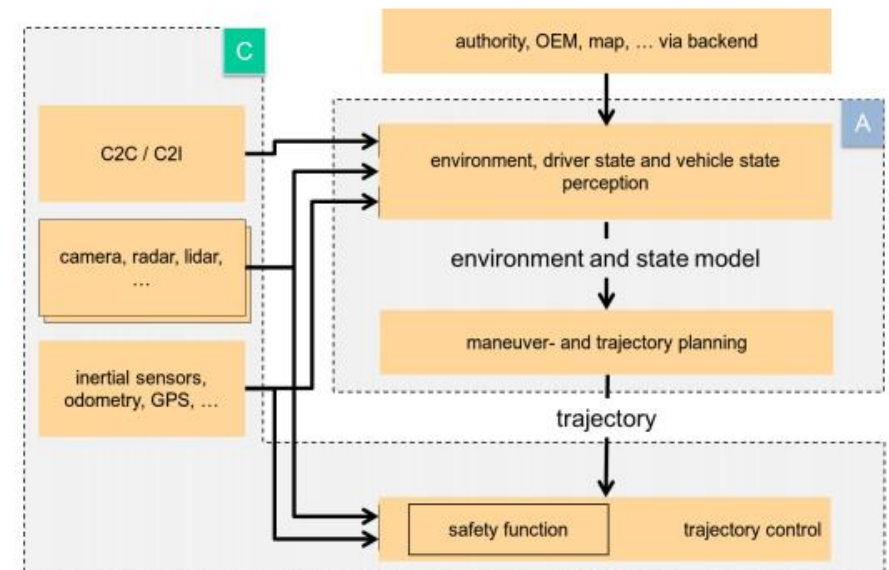
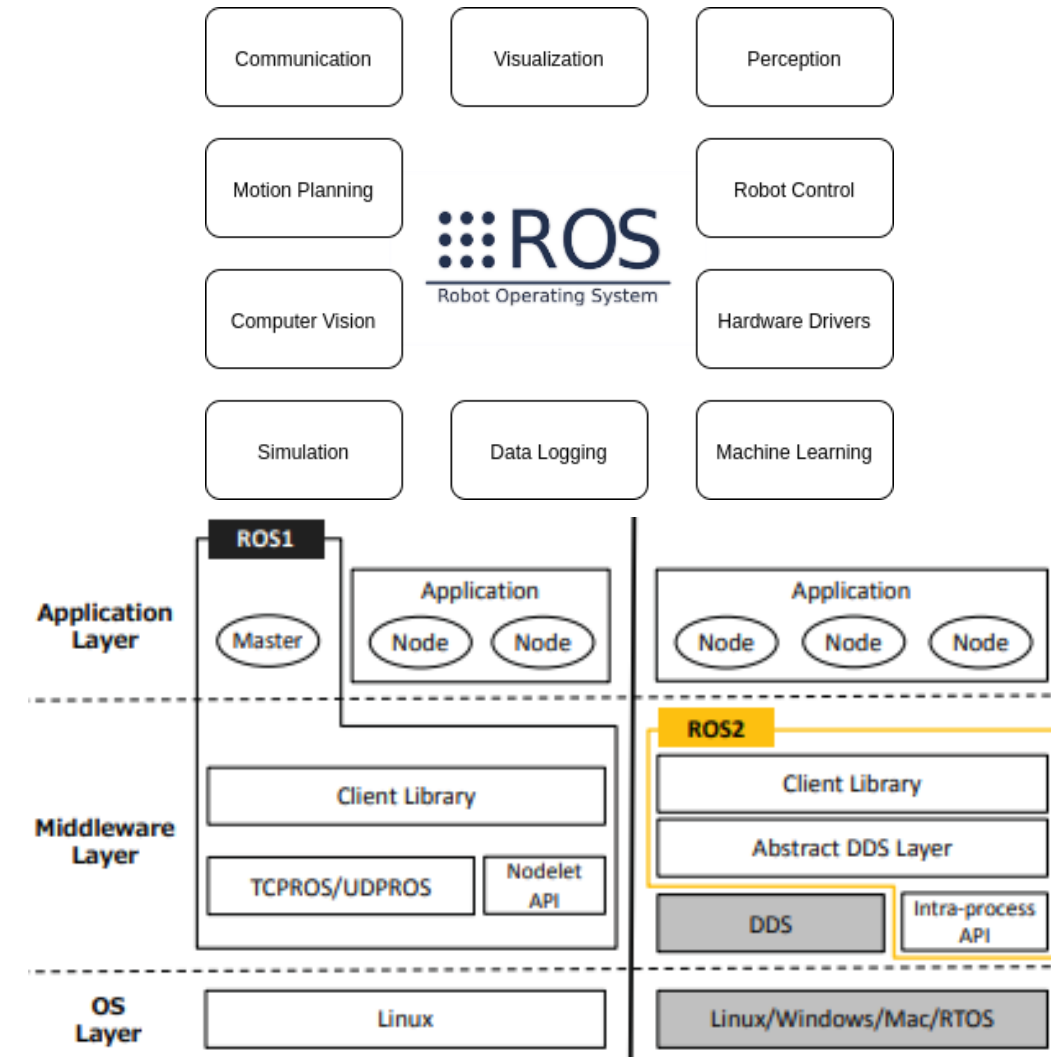


Figure 2-2 Exemplary interactions of AP and CP

Robot Operating System (ROS)

- ROS is a set of software libraries and tools for building robotic applications. Many companies use ROS to develop AVs. It uses the publish-subscribe paradigm for inter-node communication.
 - ROS has a Master node that provides naming and registration services to the rest of the nodes.
 - ROS 2 removed the Master node, and uses publish-subscribe middleware DDS (Data Distribution Service)
- Since ROS uses Linux as the underlying operating system, it is difficult to pass high-level of safety certification



Apex.OS

- “Safe and certified software framework for autonomous mobility systems.”
- Aims to be certified as a Safety Element out of Context (SEooC) up to Automotive Safety Integrity Level (ASIL)-D
 - Hard real-time, static memory allocation (no new() or malloc()), callbacks vs. waitset, security, testing, real-time I/O logging...
 - Real-Time Linux or QNX as the RTOS



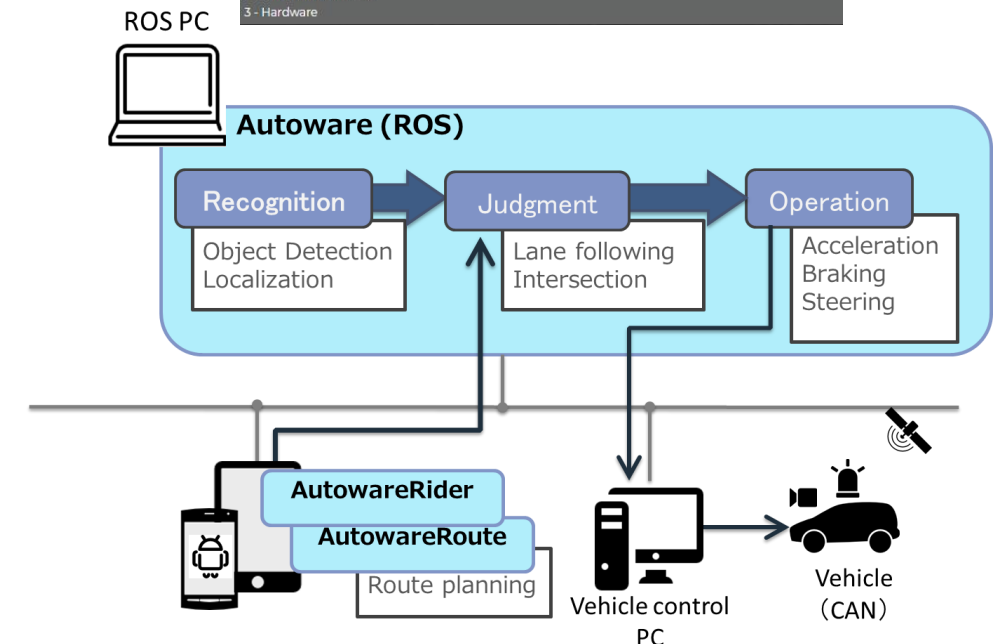
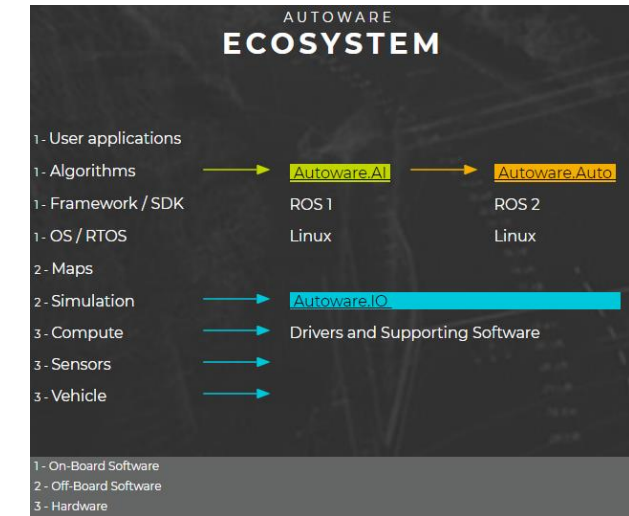
NVIDIA DRIVE Software Framework

- An open-source framework for AD (only for NVIDIA hardware).
 - **DRIVE OS** is a foundational software stack consisting of an embedded Real Time OS (RTOS), hypervisor, CUDA libraries, Tensor RT, and other modules that give you access to the hardware engines.
 - **DriveWorks SDK** enables developers to implement AV solutions by providing a comprehensive library of modules, developer tools, and reference applications.
 - **DRIVE AV** provides perception, mapping, and planning modules that utilize the DriveWorks SDK.
 - **DRIVE IX** provides full cabin interior sensing capabilities needed to enable AI cockpit solution.



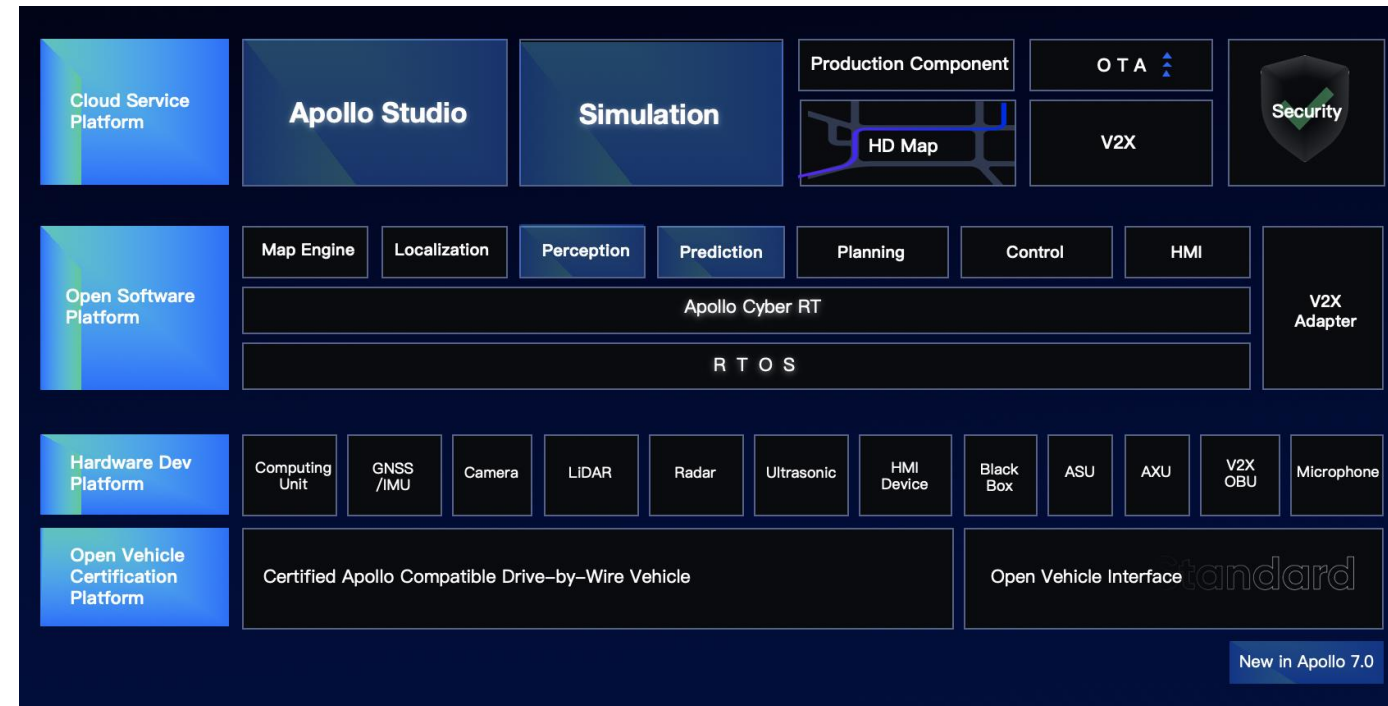
Autoware

- Open-source AD platform from Japan.
 - Autoware.AI (<https://www.autoware.ai>) is based on ROS-1
 - Autoware.auto (<https://www.autoware.auto>) is the new version based on ROS2

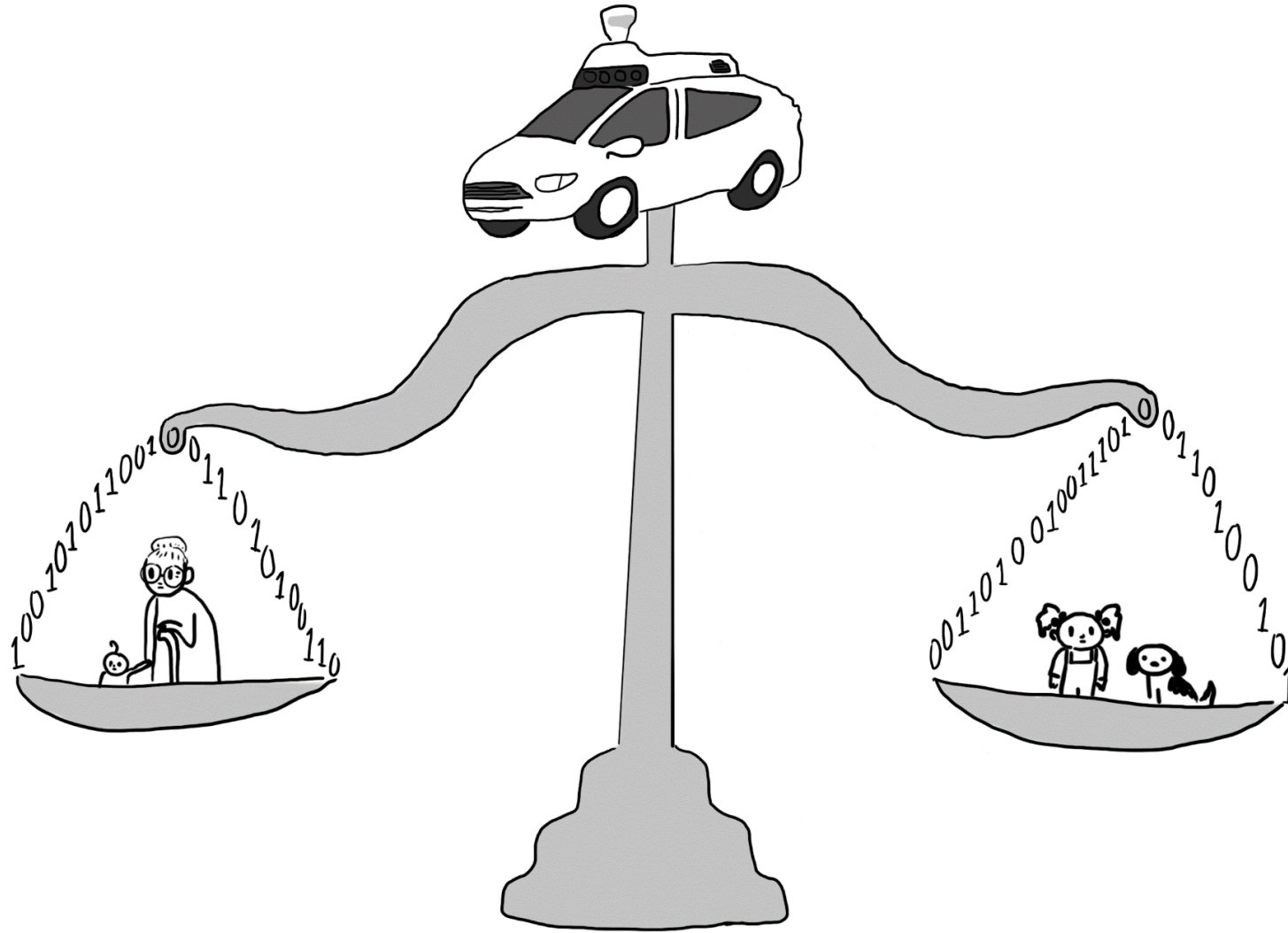


Baidu Apollo

- An open-source, hardware-neutral AD platform from China
- Initially based on ROS, but later replaced ROS with their own components
 - Real-Time OS: Linux kernel with real-time patch
 - Cyber RT: lightweight, high-performance communication middleware

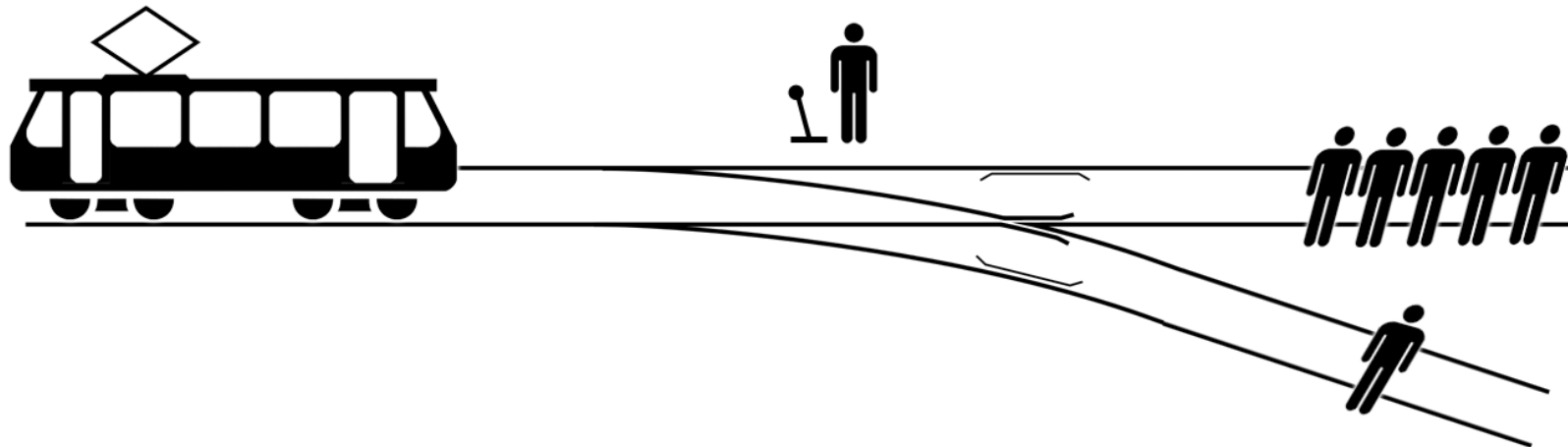


Ethical Issues



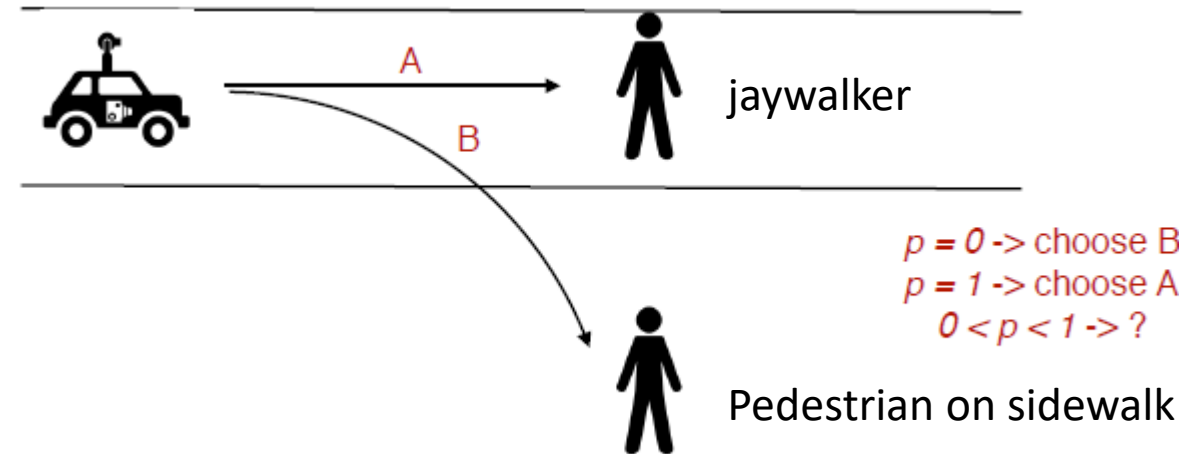
The Trolley Problem

- There is a runaway trolley barreling down the railway tracks. Ahead, on the tracks, there are five people tied up and unable to move. The trolley is headed straight for them. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a different set of tracks. However, you notice that there is one person on the side track. You have two options:
 - Do nothing and allow the trolley to kill the five people on the main track.
 - Pull the lever, diverting the trolley onto the side track where it will kill one person.
- What is the right thing to do?



Variant of Trolley Problem with a Probability Threshold

- You are in a situation where:
 - A. you kill a pedestrian with probability **1**, but it's **not your fault**
 - B. you kill a different pedestrian with probability p , and it is **your fault**
- What is your threshold value p_{th} for making the choice?
 - if($p \geq p_{th}$) choose A; otherwise choose B



MIT Moral Machine Experiment

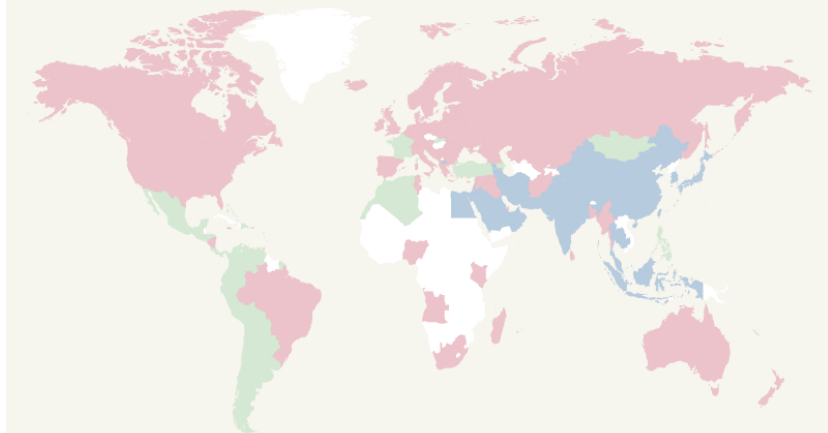
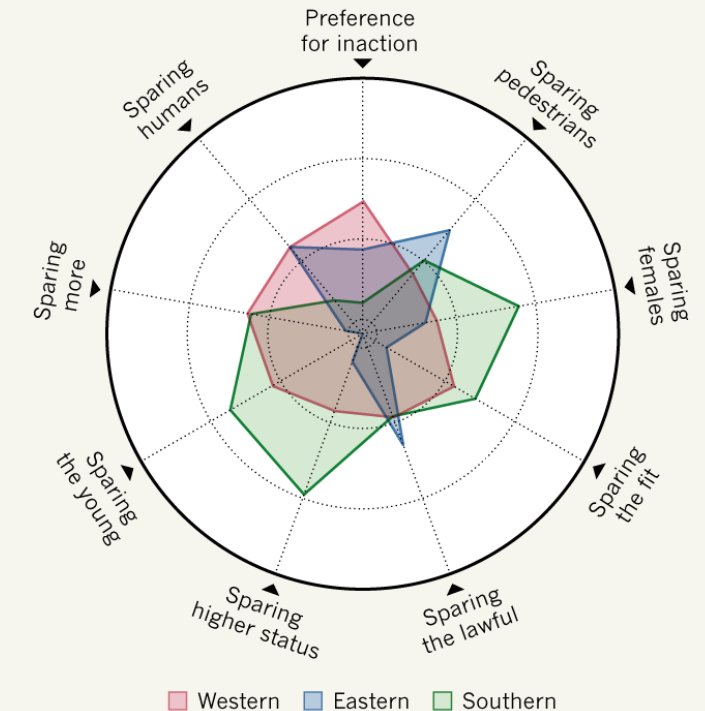
- A 2016 survey indicates that people wanted an autonomous vehicle to protect pedestrians even if it meant sacrificing its passengers — but also that they wouldn't buy self-driving vehicles programmed to act this way. This prompted the MIT Moral Machine Experiment, a platform for gathering a human perspective on moral decisions made by machine intelligence (<http://moralmachine.mit.edu/>)
 - One scenario: an AV must choose between killing two passengers or five pedestrians. An AV experiences a sudden brake failure. Staying on course would result in the death of two elderly men and an elderly woman who are crossing on a 'do not cross' signal (left). Swerving would result in the death of three passengers: an adult man, an adult woman, and a boy (right)
 - You can also design other scenarios. Accident scenarios are generated with nine factors: sparing humans (versus pets), staying on course (versus swerving), sparing passengers (versus pedestrians), sparing more lives (versus fewer lives), sparing men (versus women), sparing the young (versus the elderly), sparing pedestrians who cross legally (versus jaywalking), sparing the fit (versus the less fit), and sparing those with higher social status (versus lower social status)
- This platform gathered 40 million decisions in ten languages from millions of people in 233 countries
- Moral Machines: How culture changes values
 - <https://www.youtube.com/watch?v=jPo6bby-Fcg>

Cultural Clusters

- Three large clusters
 - Western: Protestant, Catholic, and Orthodox countries in Europe and North America
 - Eastern: Islamic and Confucian (Asian) cultures
 - Southern: Central and South America, as well as France and former French colonies
- The preference to spare younger characters rather than older characters is much less pronounced for countries in the Eastern cluster, and much higher for countries in the Southern cluster.
- The same is true for the preference for sparing higher status characters.
- Countries in the Southern cluster exhibit a much weaker preference for sparing humans over pets, compared to the other two clusters.
- Only the (weak) preference for sparing pedestrians over passengers and the (moderate) preference for sparing the lawful over the unlawful appear to be

MORAL COMPASS

A survey of 2.3 million people worldwide reveals variations in the moral principles that guide drivers' decisions. Respondents were presented with 13 scenarios, in which a collision that killed some combination of passengers and pedestrians was unavoidable, and asked to decide who they would spare. Scientists used these data to group countries and territories into three groups based on their moral attitudes.



AV Ethical Issues: is it Worth the Time?

- Many argue that ethical issues are just a distraction from the real problem of AV safety and security, esp. in the presence of ML/DL algorithms.
 - None of the AV accidents in recent years involved any ethical decisions similar to the Trolley Problem. They are due to failures in sensors or perception algorithms.
- Sebastian Thrun (former head of Google's SDC project, former professor at Stanford who led the development of Stanley, winner of DARPA Grant Challenge in 2005):
 - "I think it's a great thing for philosophers to discuss these kind of problems. They can get tenure at their universities, but it's not of practical relevance. If we manage with certain car technology to halve the traffic deaths in the world, which means if we are able to have 500,000 fewer deaths in total, then for this extremely rare, purely hypothetical trolley problem that might occur once in a hundred years. I think whatever the outcome is, the mental energy that philosophers have spent on discussing it is completely out of proportion to the benefit of others on one problem. I will leave it at that."

AV Testing Legislation (USA)

- It is necessary to test AVs on public roads for technology development, but is it ethical?
- Legislation regulating AV testing differs widely across states. Several states have no proposed legislation, meanwhile states like Nevada, California, Texas, and Arizona are hotbeds for testing AVs (partly due to sunny weather)

