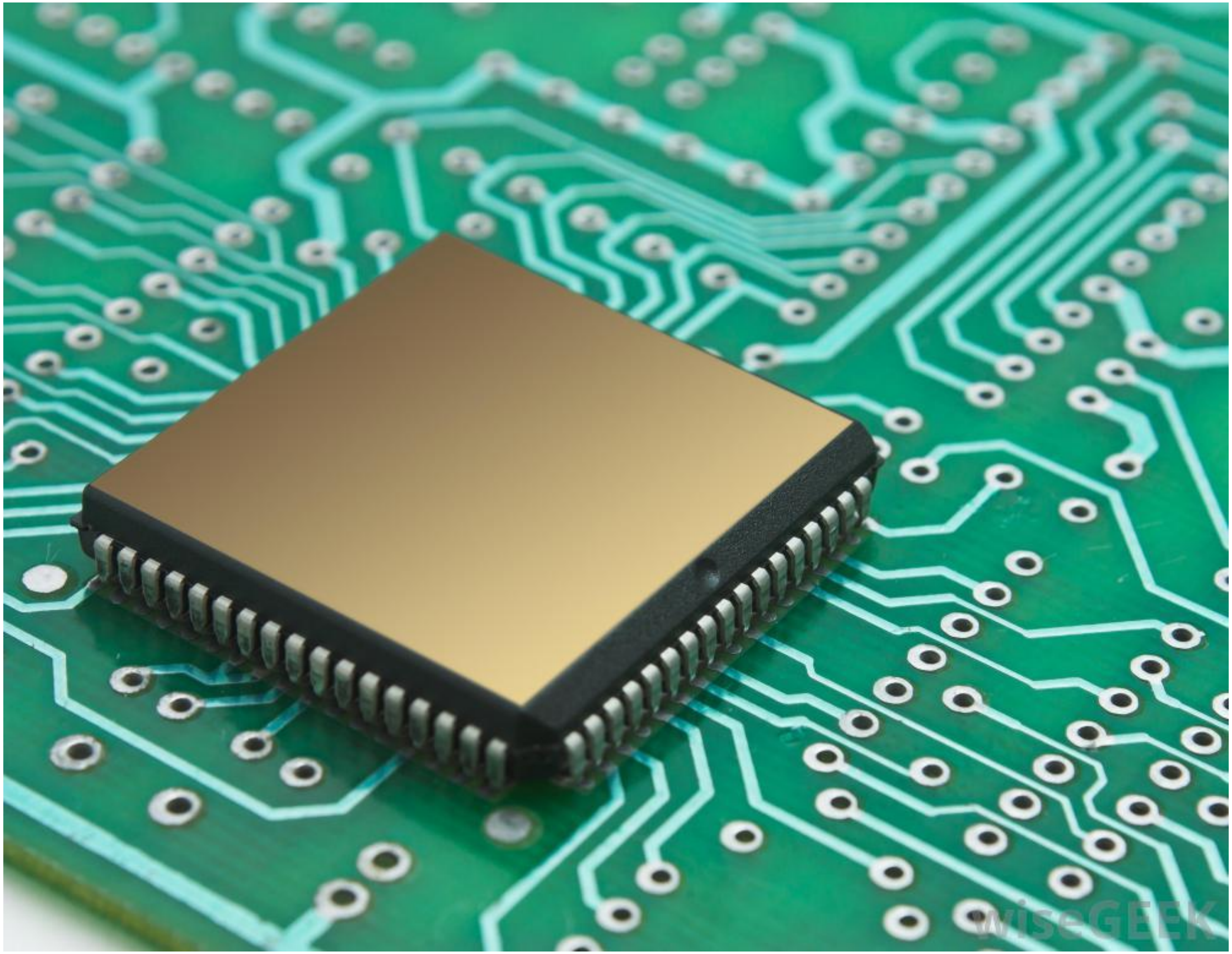
L1.2 HW/SW Platforms, V2X

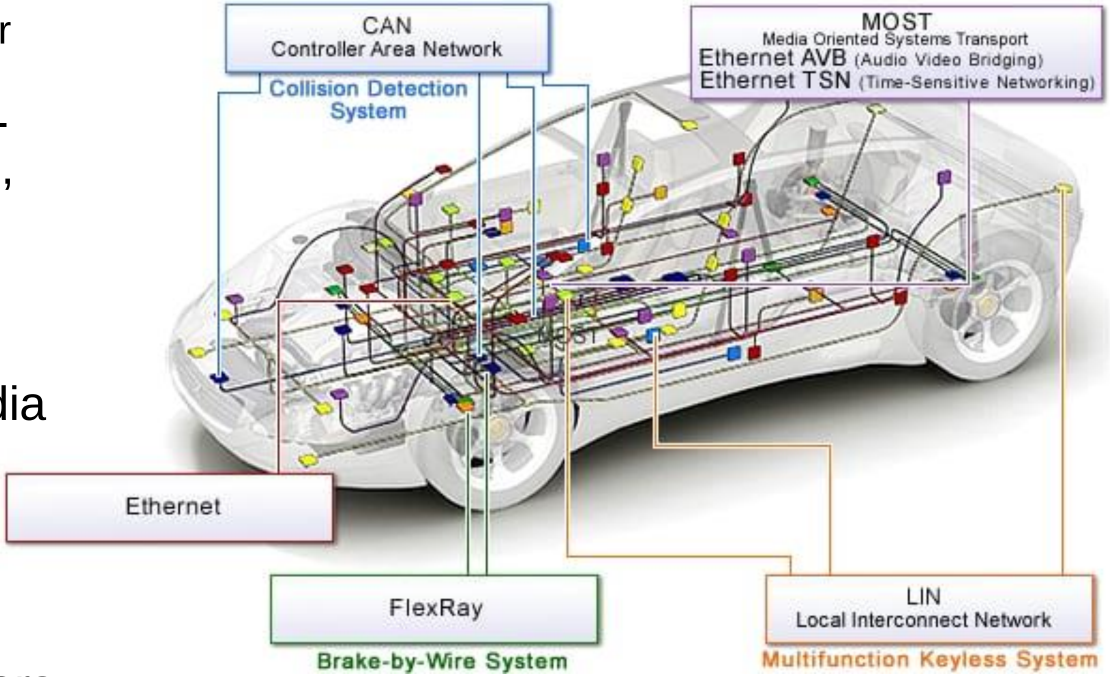
Zonghua Gu 2023

HARDWARE PLATFORMS



Typical Automotive E/E Architecture

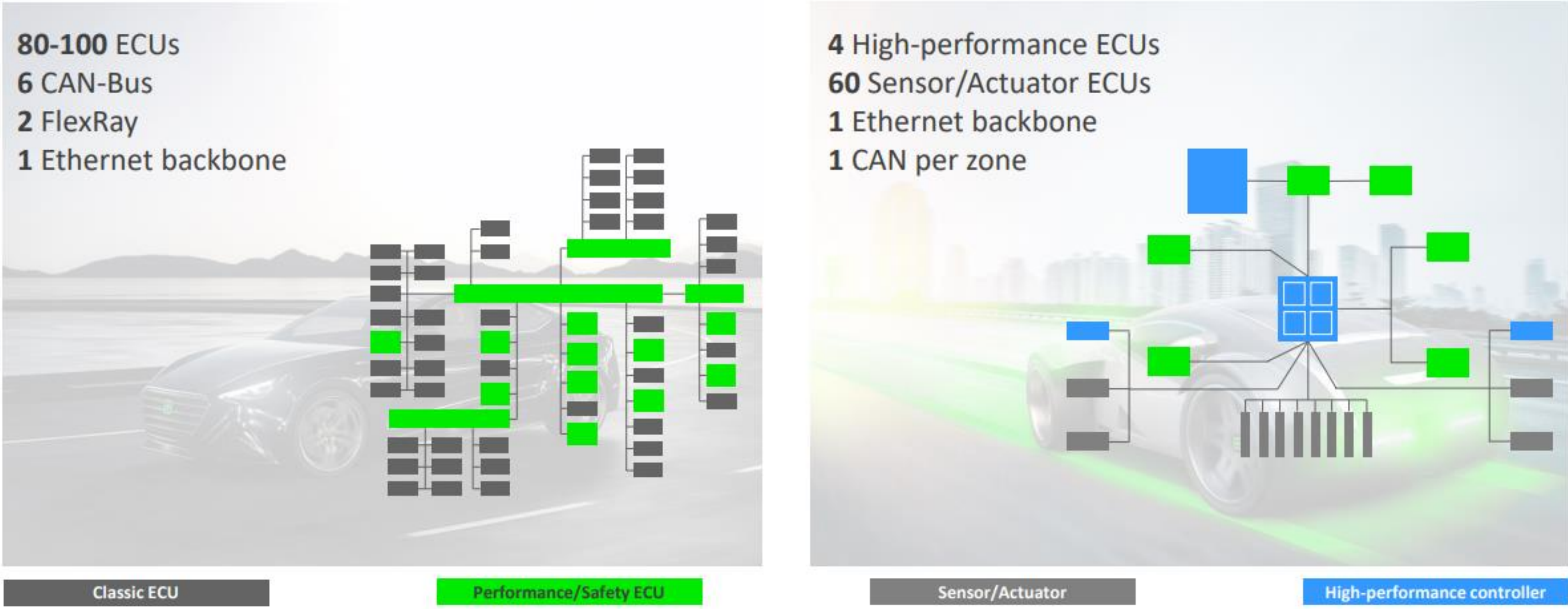
- Ethernet as high-bandwidth backbone network
  - Ethernet TSN posed to be the dominating standard protocol.
  - Regular Ethernet is also used for diagnostics
- FlexRay for safety-critical X-by-Wire, where X stands for brake, steer, drive…
  - Gradually being replaced by Ethernet
- Media Oriented Systems 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…



https://www.renesas.com/eu/en/solutions/automotive/technology/networking-solutions.html

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.



Trunk of an Experimental AV from Ford (2017)

Where do I put my groceries?

https://www.theverge.com/2017/10/10/16449416/nvidia-pegasus-self-driving-car-ai-robotaxi

AD Hardware Considerations

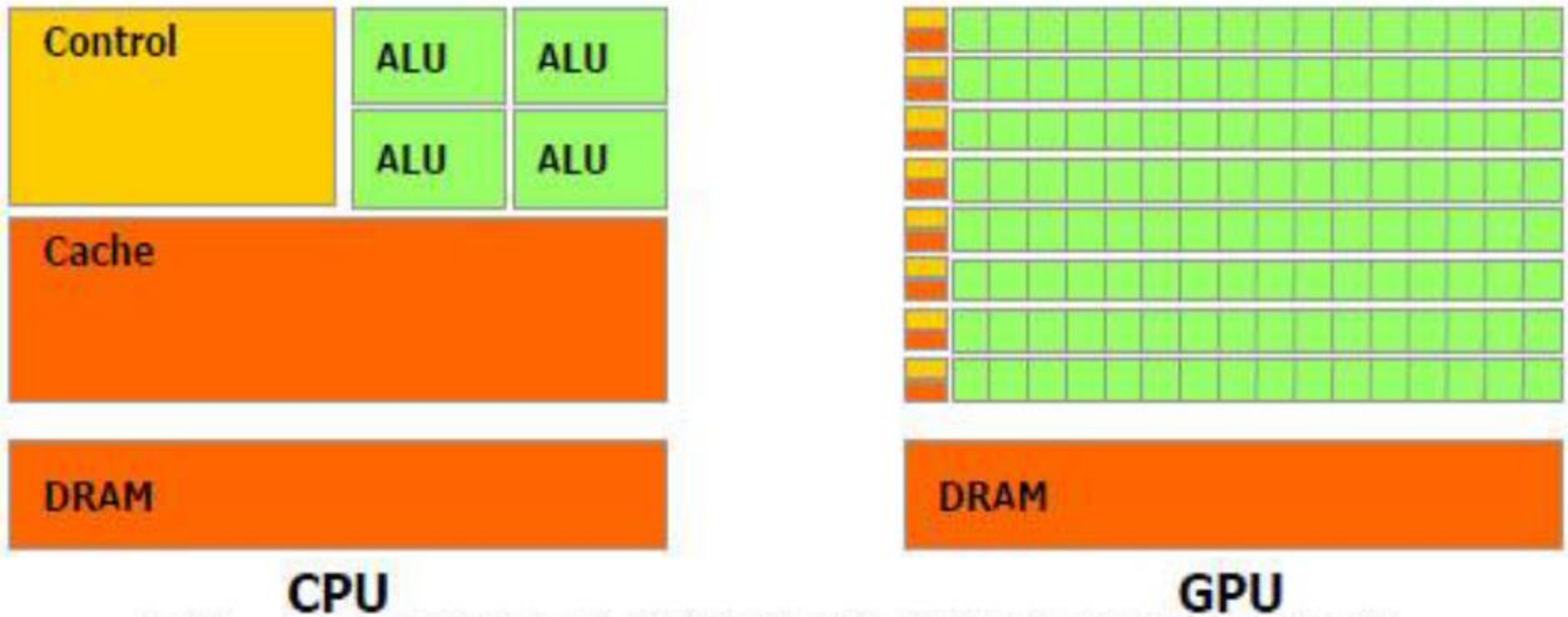
- Power consumption
  - Power consumption of electronics (sensors and computing hardware) for AD may be 100x that of a vehicle with regular ADAS. This drains battery and implies increased fuel consumption or reduced range for EVs
    - EV drivers often turn off air conditioning due to range anxiety; will they turn off AD and drive manually for this reason?
  - Waymo and Ford now focus on Hybrid Vehicles, while Uber uses a fleet of full gasoline SUVs.
- Cooling capacity
  - Fan or liquid cooling.
- Form factor
  - Must be compact and unobtrusive.
- 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
  - Mostly inference tasks, but may also perform training tasks in case of online-learning.
- 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 venders, 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.

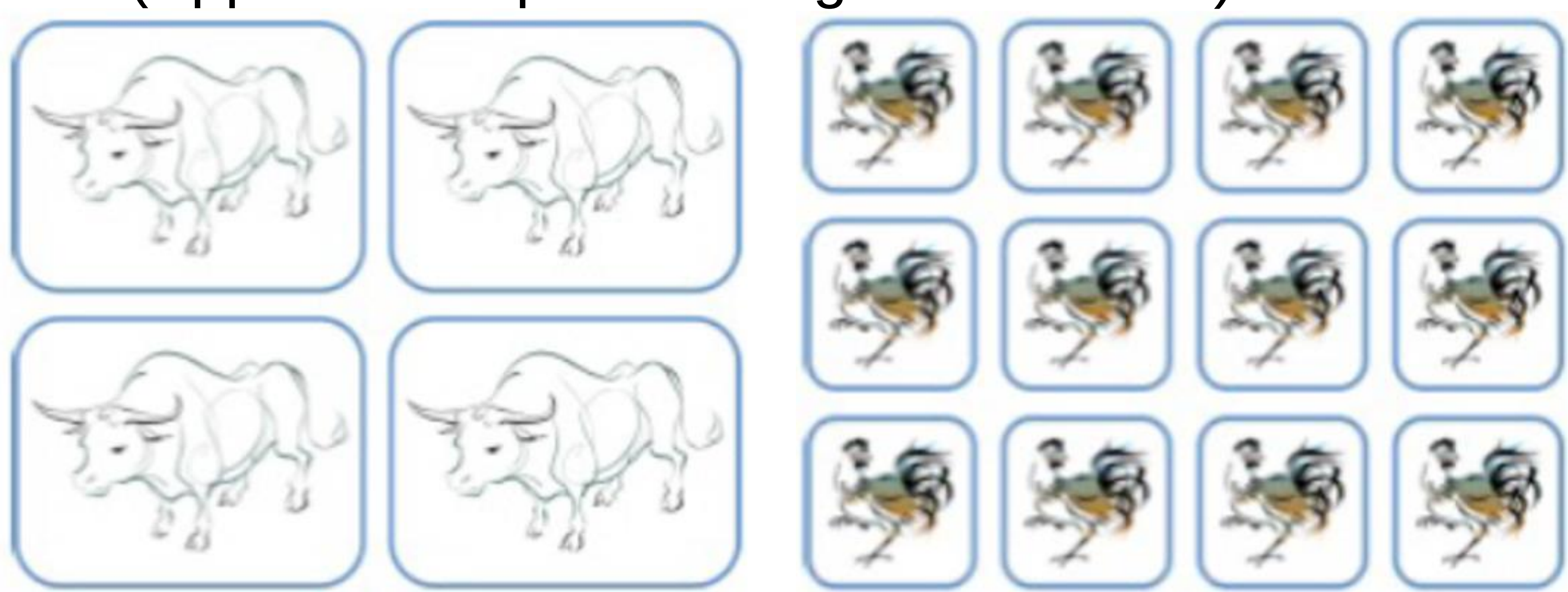
CPU vs. GPU

- GPU has much simpler control logic than CPU, hence has more computational elements (Arithmetic Logic Units)
- GPU is ideally suited for processing highly-parallel workloads
  – e.g. matrix-multiply, which is a core operation in Deep Learning algorithms



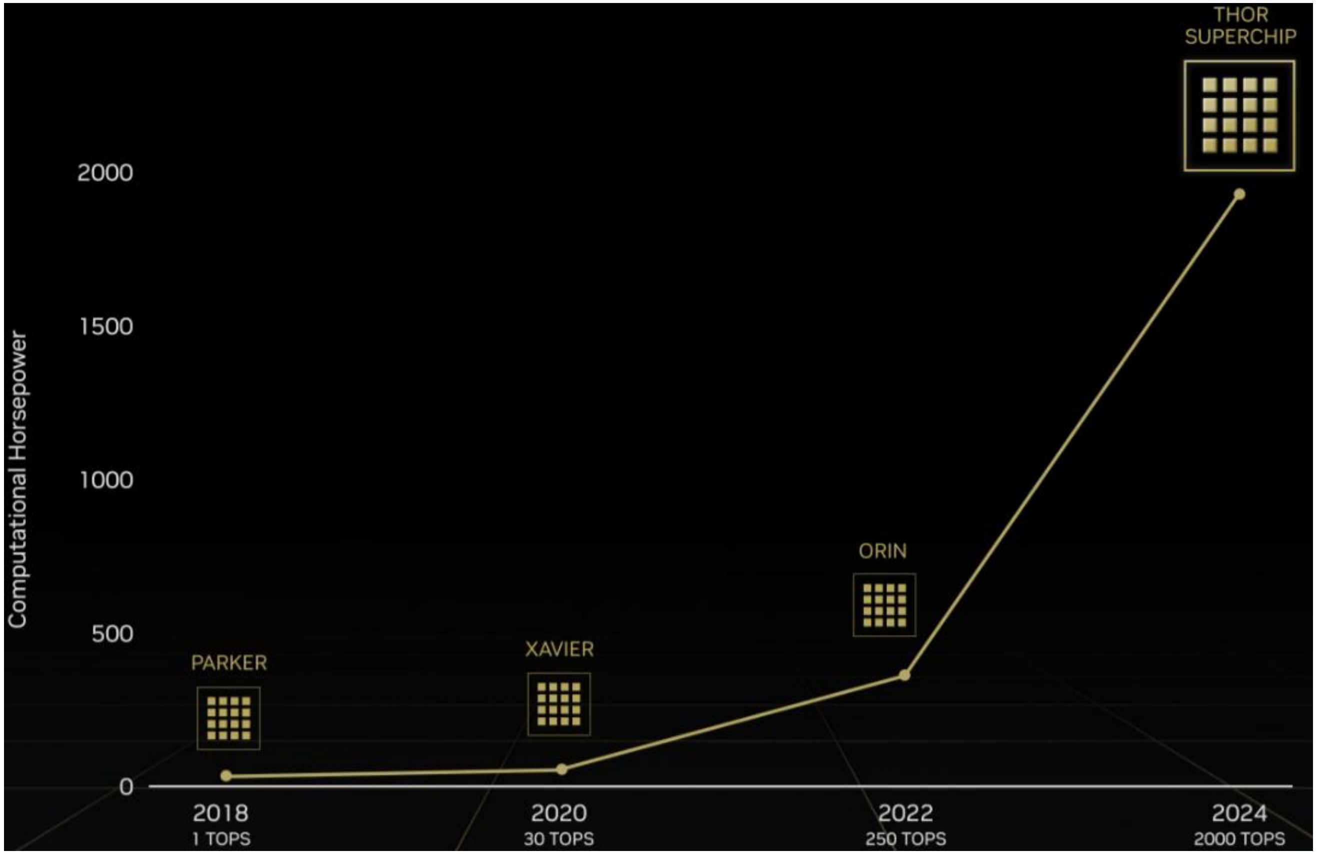
CPU vs. GPU

- When you are plowing a field, would you prefer 4 strong oxen (a multicore CPU), or 1024 chickens (a GPU)?
- Similar arguments for FPGAs (Field-Programmable Gate Arrays) and ASICs (Application-Specific Integrated Circuits)



NVIDIA DRIVE Hardware

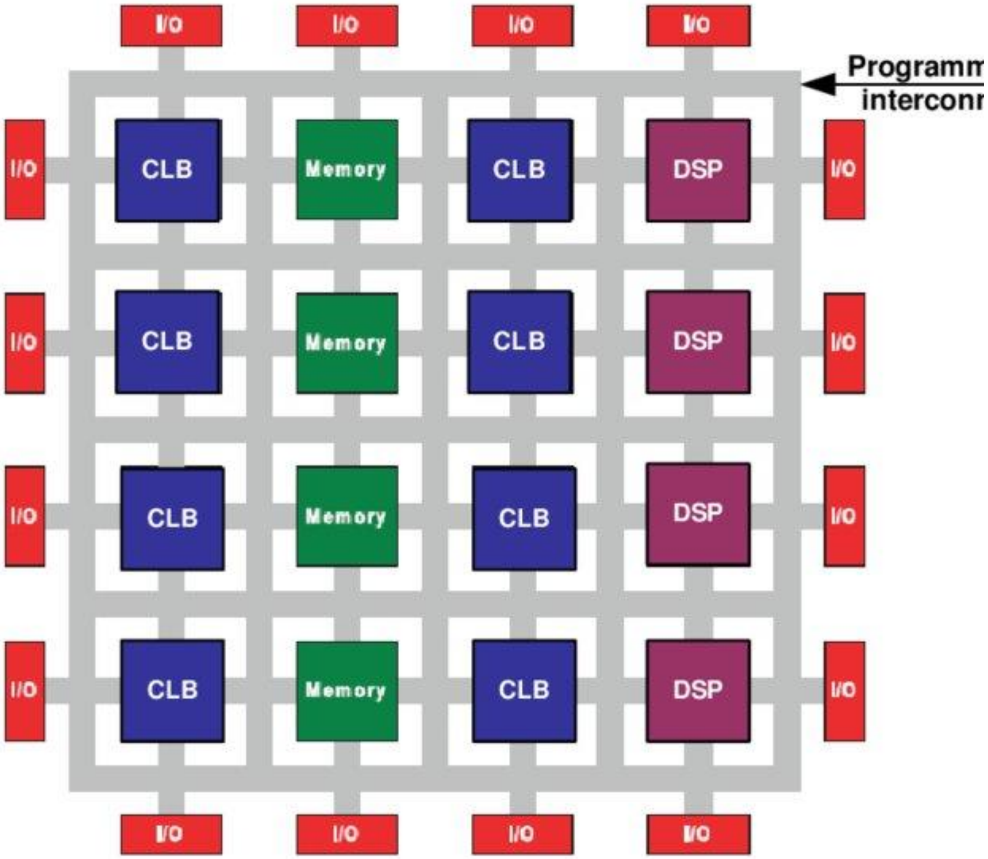
- A family of products, ranging from the low-end Parker to the latest high-end THOR with 2000 TOPS (Tera Operations Per Second)
  - Besides CPU and GPU cores, also includes NVDLA (NVIDIA Deep Learning Accelerator), an ASIC for Deep Learning inference.



https://wccftech.com/nvidia-intros-drive-thor-balancing-ai-performance-for-full-vehicle-autonomy/

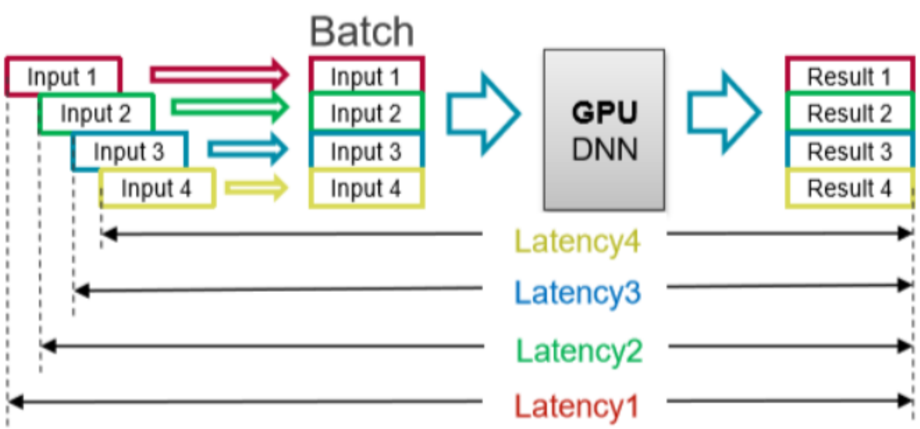
FPGAs

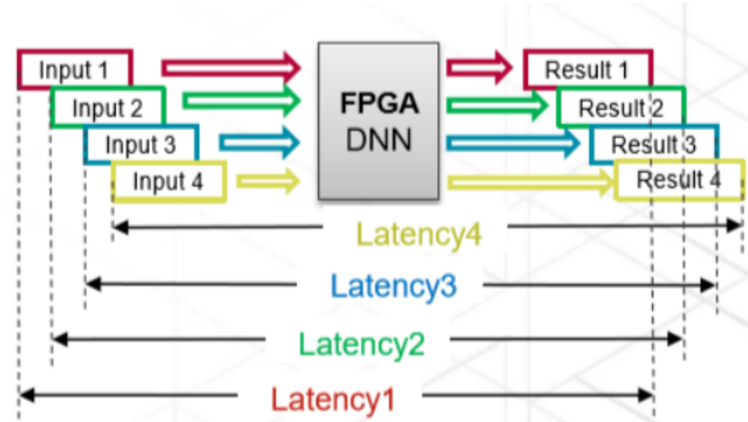
- FPGA is reprogrammable hardware, consisting of an array of Configurable Logic Blocks (CLBs) and interconnections which can be configured at design time or runtime.



FPGA vs. GPU

- 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.

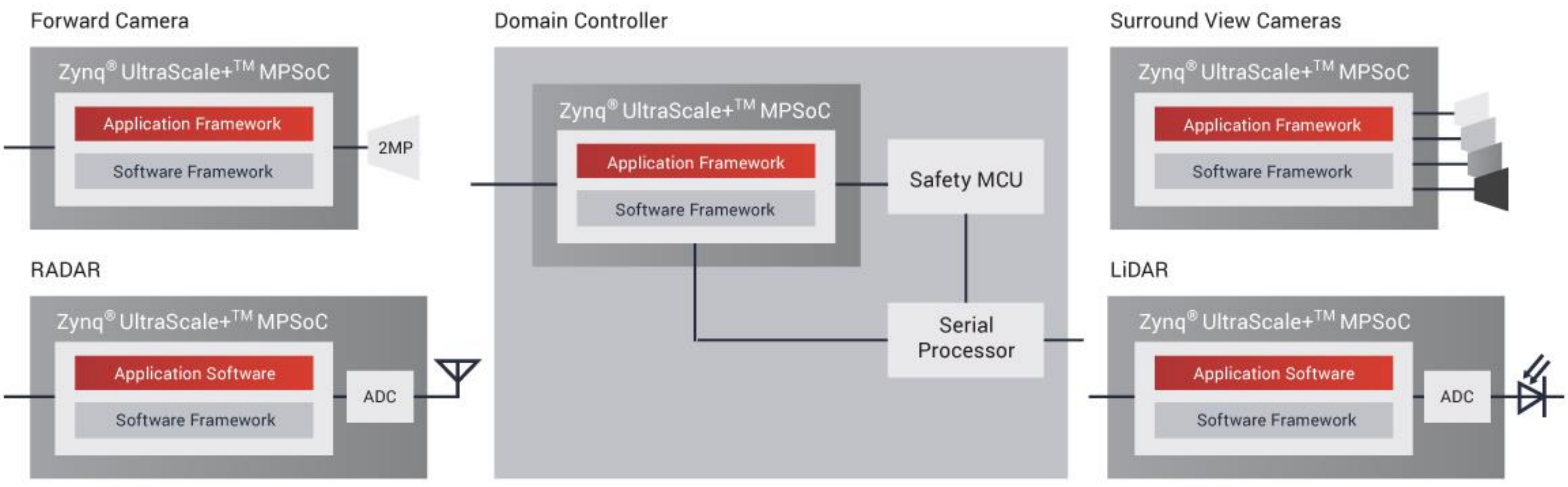




https://www.eetimes.com/where-do-fpgas-stand-in-auto-ic-race/#

FPGA for Automotive

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



https://www.xilinx.com/applications/automotive/automated-driving.html

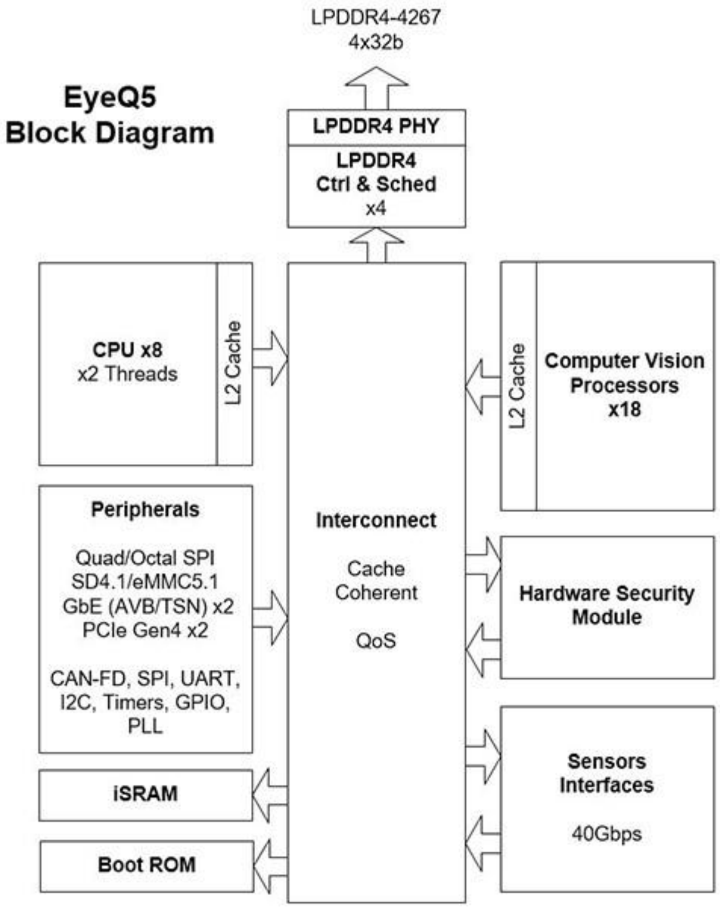
ASICs

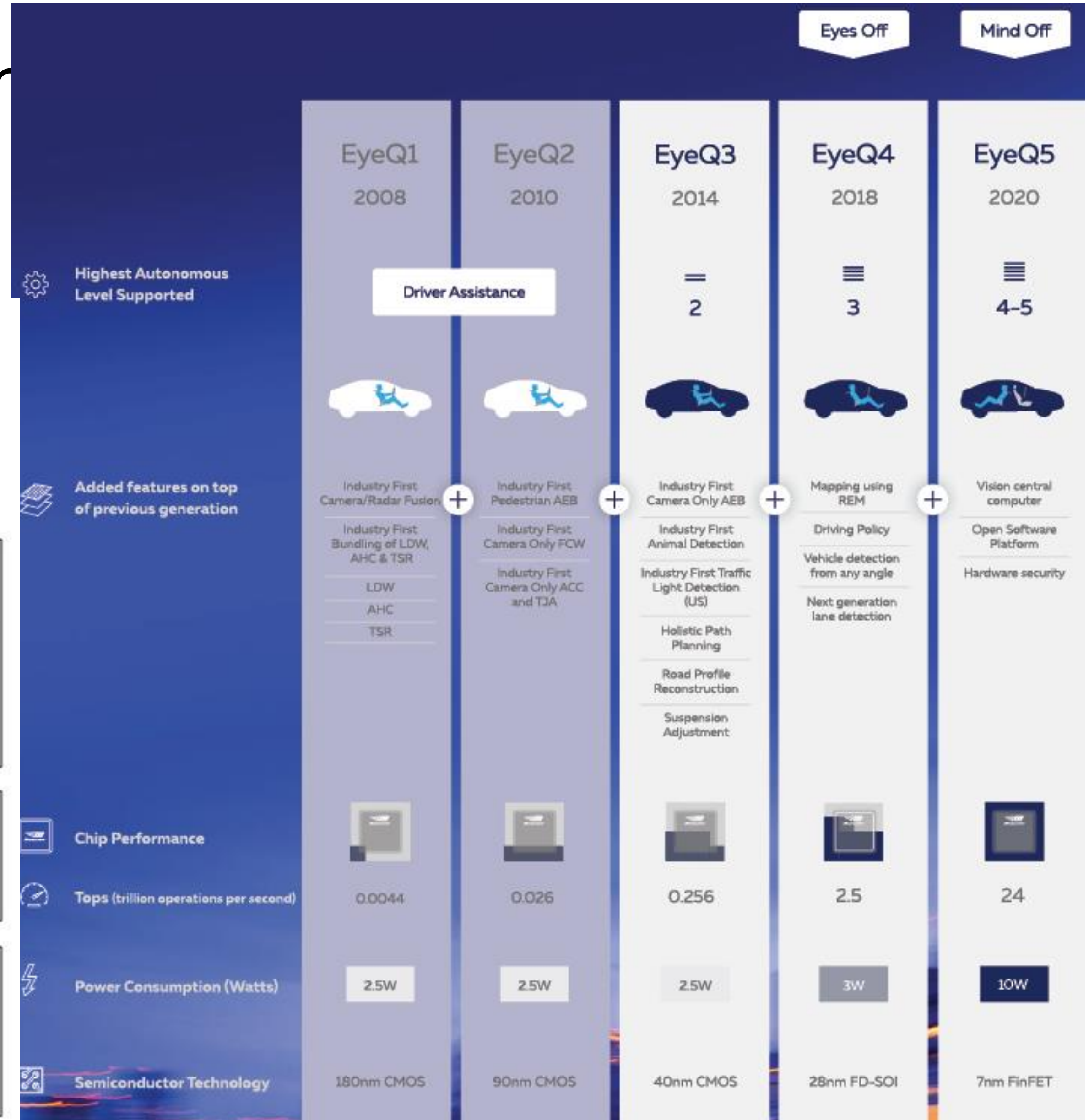
- ASICs for Deep Learning are often called Neural Processing Units or AI accelerators.
- ASICs, thanks to dedicated circuit design, may achieve up to 10x in computation efficiency and power consumption compared to CPU/GPU, and less dramatic, but still significant improvement compared to FPGA. The drawback is loss of programmability and flexibility.
  - Industry: almost every chip vendor provides some kind of AI accelerator, e.g. Google’s Tensor Processing Unit (TPU)
  - Academia: AI accelerators is a dominating topic in top conferences in computer architecture, including ISCA, MICRO and HPCA.

MobileEye

EyeQ Series

- “Computer Vision Processors” are ASICs

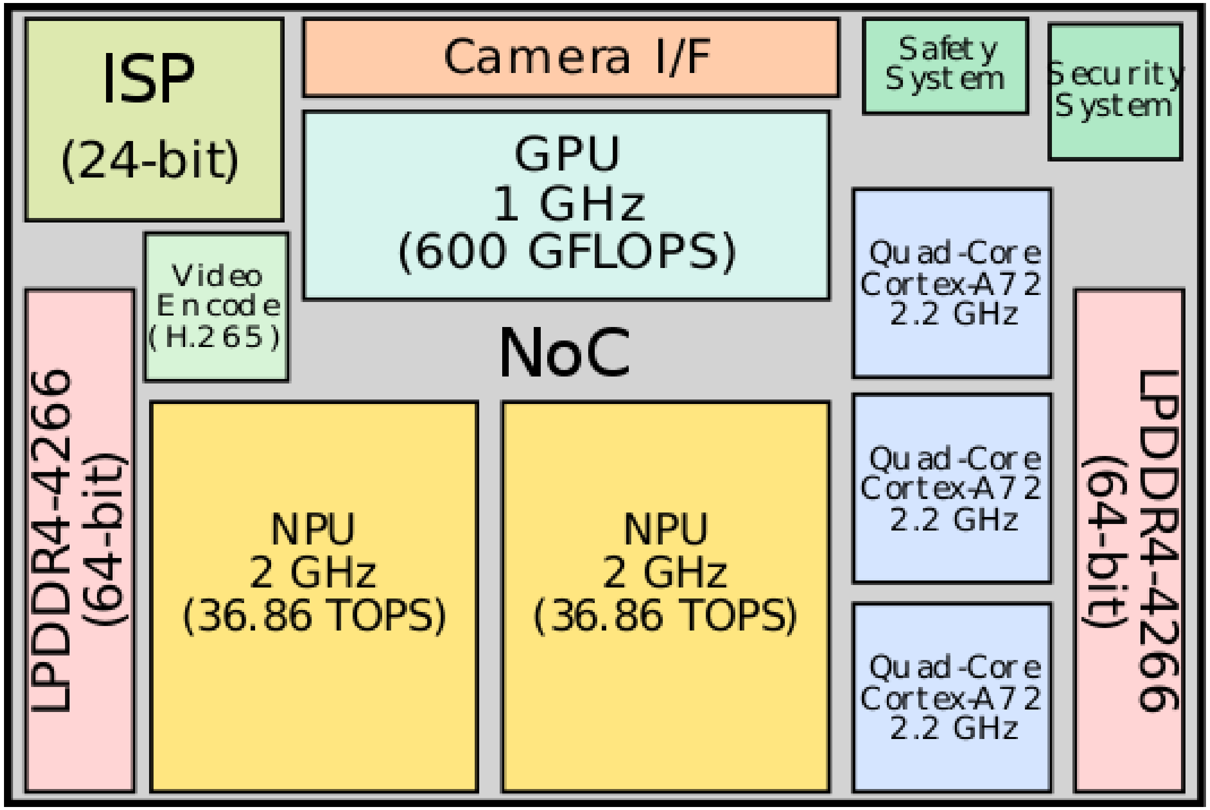




https://www.mobileye.com/our-technology/evolution-eyeq-chip/

Tesla FSD

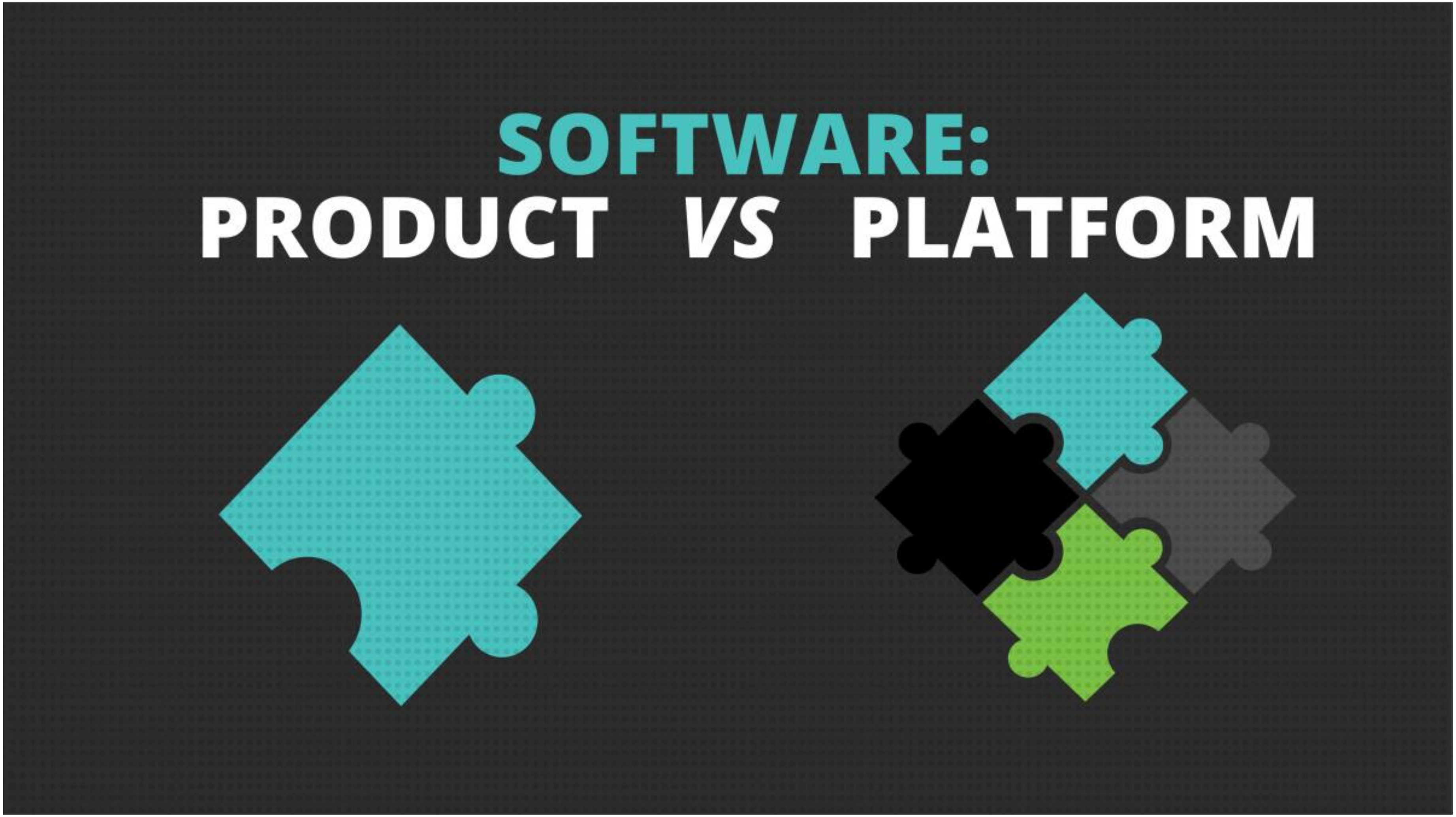
- Full Self-Driving Chip (FSD) is designed by Tesla and introduced in early 2019 for their own cars.
- It incorporates 3 quad-core Cortex-A72 clusters for a total of 12 CPUs operating at 2.2 GHz, a GPU operating 1 GHz, 2 Neural Processing Units (NPUs) operating at 2 GHz, and various other hardware accelerators.
  - NPU: ASIC for Deep Learning inference



Products from Automotive OEMs and Suppliers

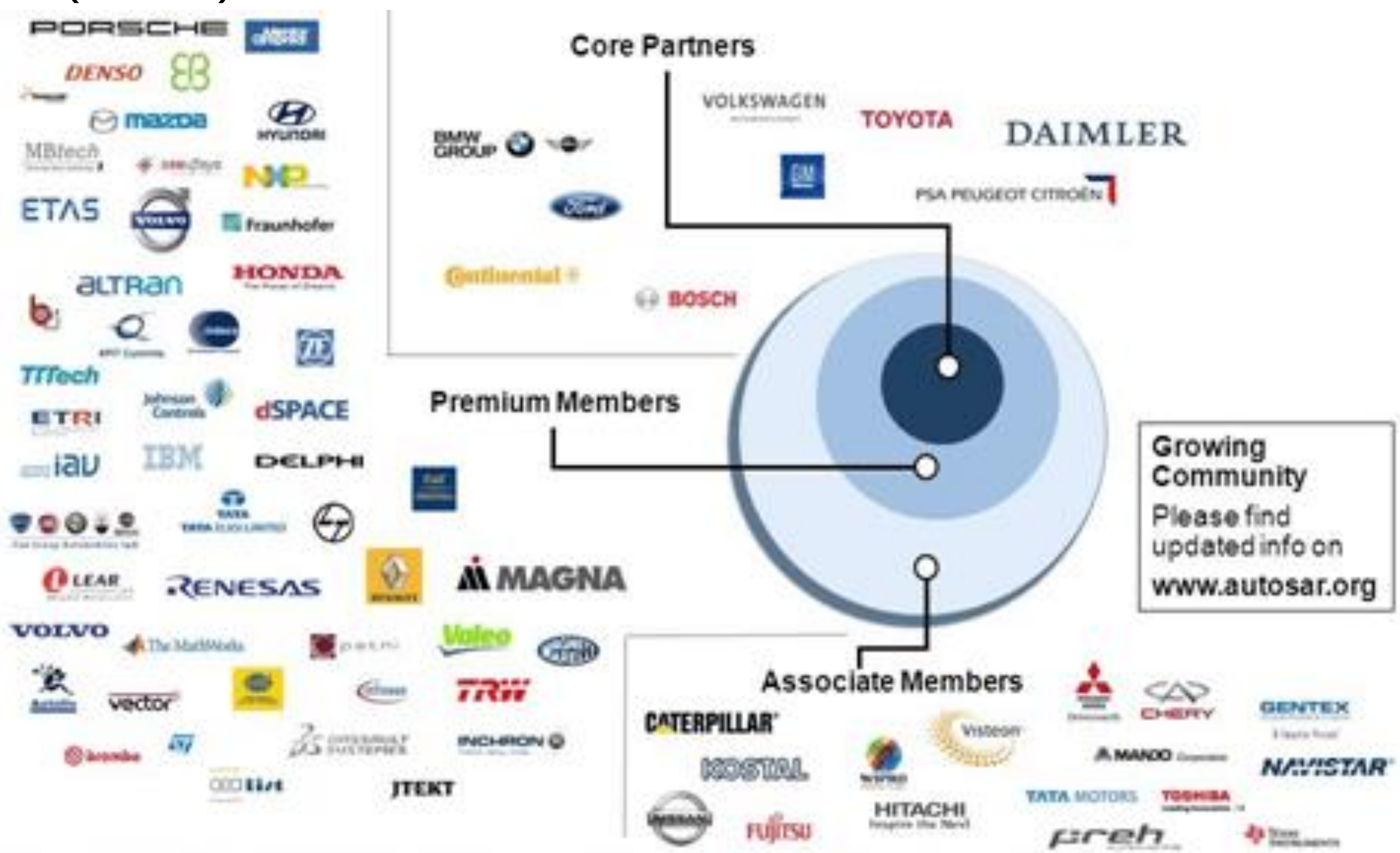
- These companies do not design chips. Rather, they offer integration solutions based on products from chip vendors like NVIDIA.
  - Delphi/Audi zFAS (zentrales Fahrerassistenz-Steuergeraet)
    - based on NVIDIA Tegra K1 and Mobileye EyeQ3
    - Hedging their bet on products from two mortal enemies ☺
  - ZF ProAI
    - based on NVIDIA DRIVE PX2
  - Bosch AI Car Computer
    - based on NVIDIA DRIVE AGX Xavier
  - Continental ADCU; Visteon DriveCore; NXP BlueBox; Renesas R-Car…

SOFTWARE PLATFORMS

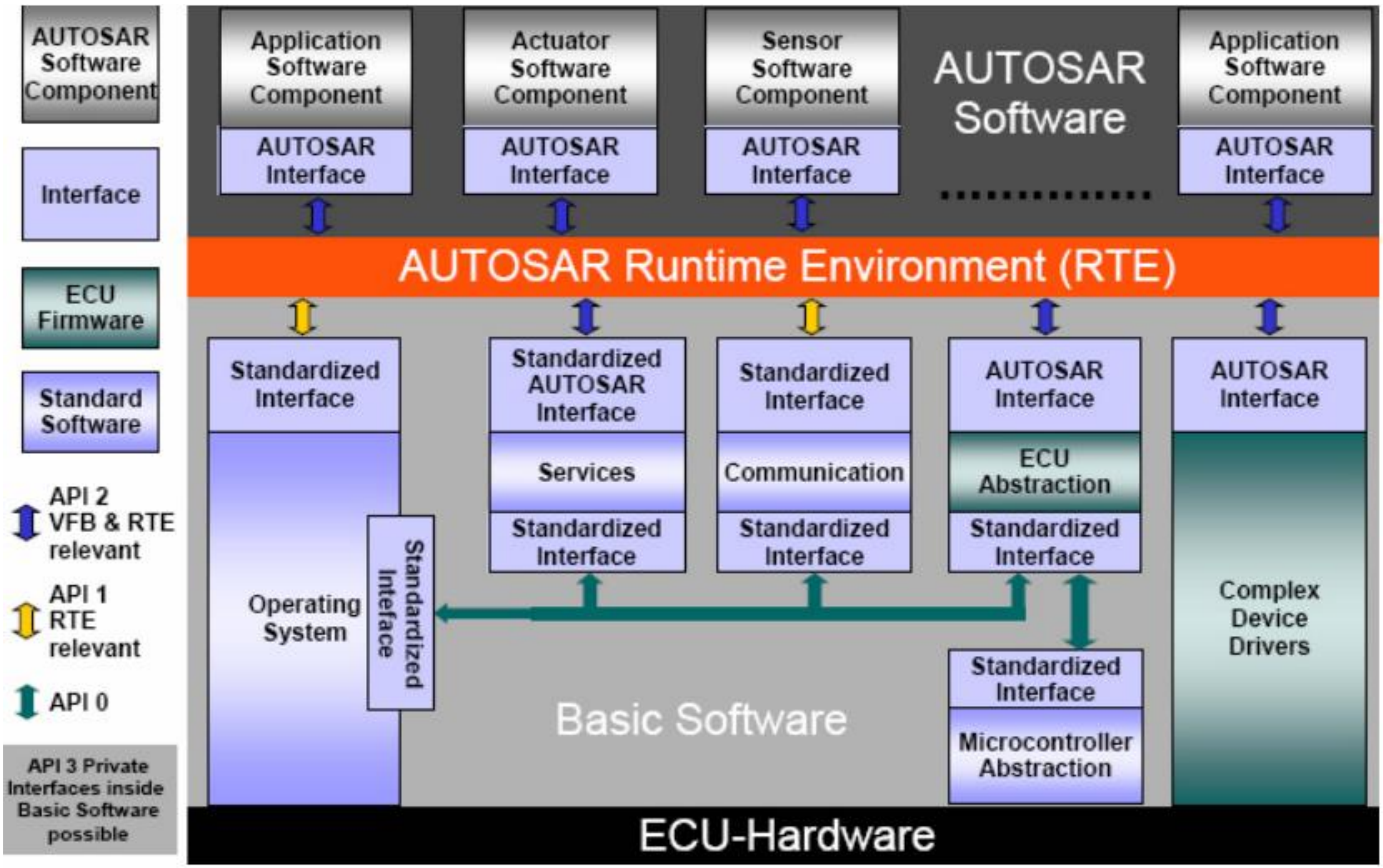


The AUTOSAR Consortium

- 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).

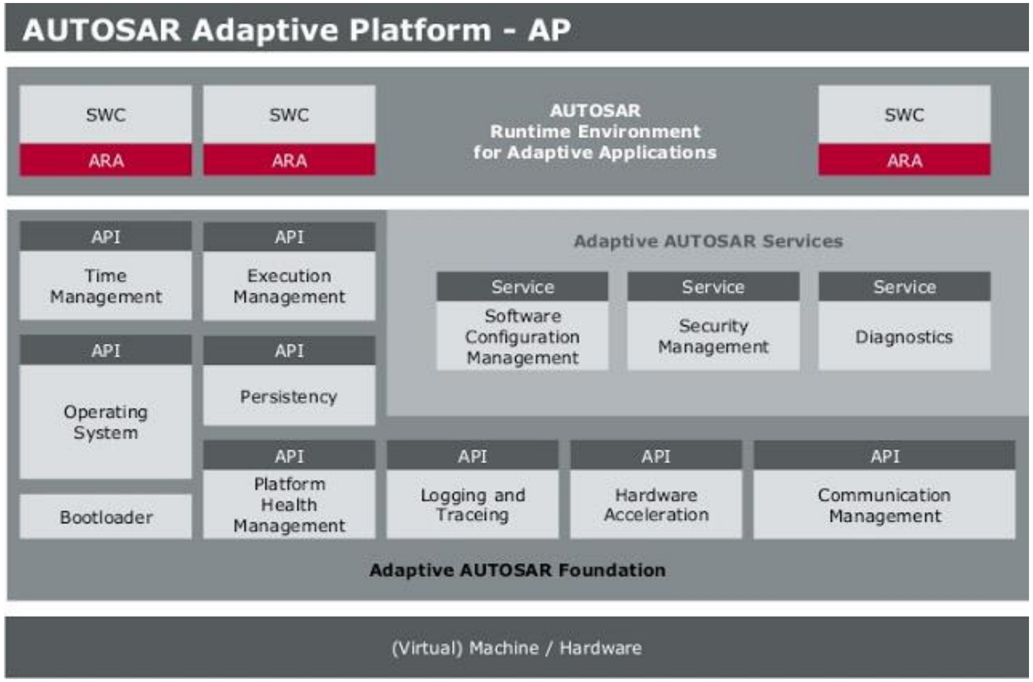


AUTOSAR Classic Platform



AUTOSAR Adaptive Platform

- AUTOSAR-AP is an industry standard that specifies standard interfaces required for developing future high-performance multicore automotive ECUs.
- Compared to Classic AUTOSAR for resource-constrained safety-critical ECUs, AUTOSAR-AP is designed for high-performance ECUs, and allows dynamic linking of services and clients during ECU runtime, which facilitates Over-the-Air (OTA) Update.



Integration of Multiple Software 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.

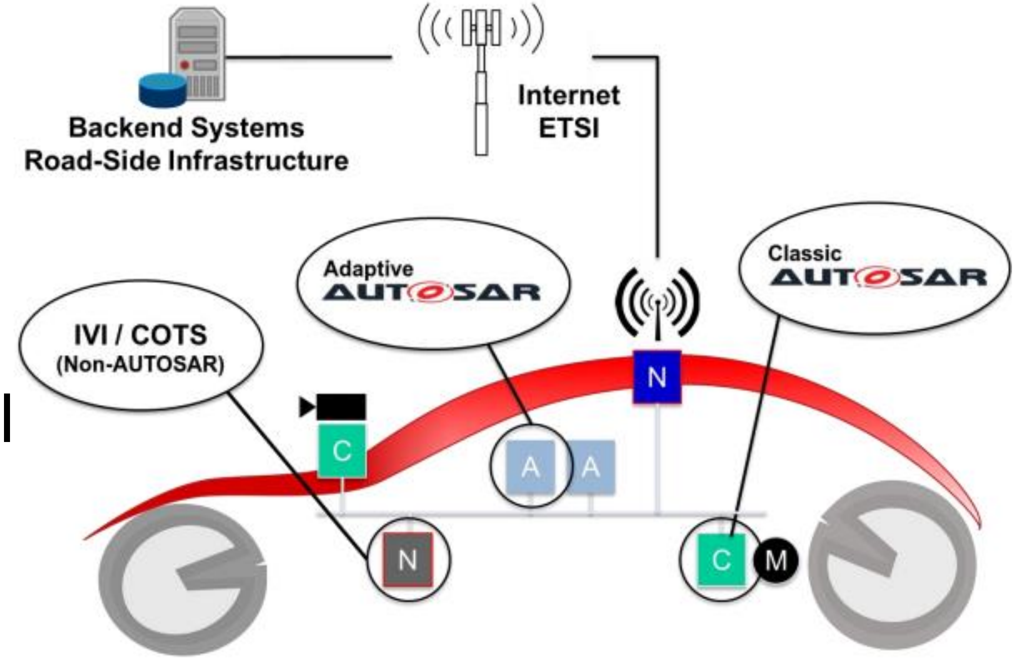


Figure 2-1 Exemplary deployment of different platforms

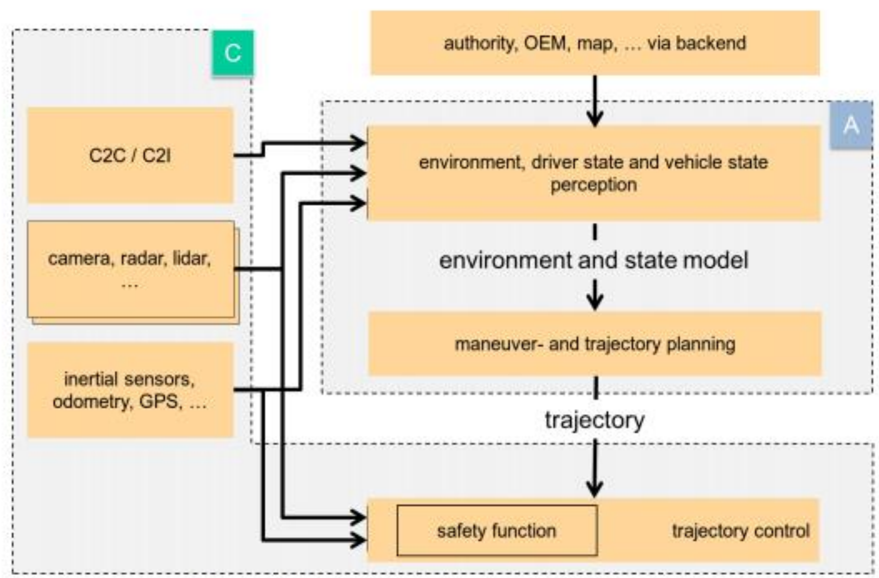
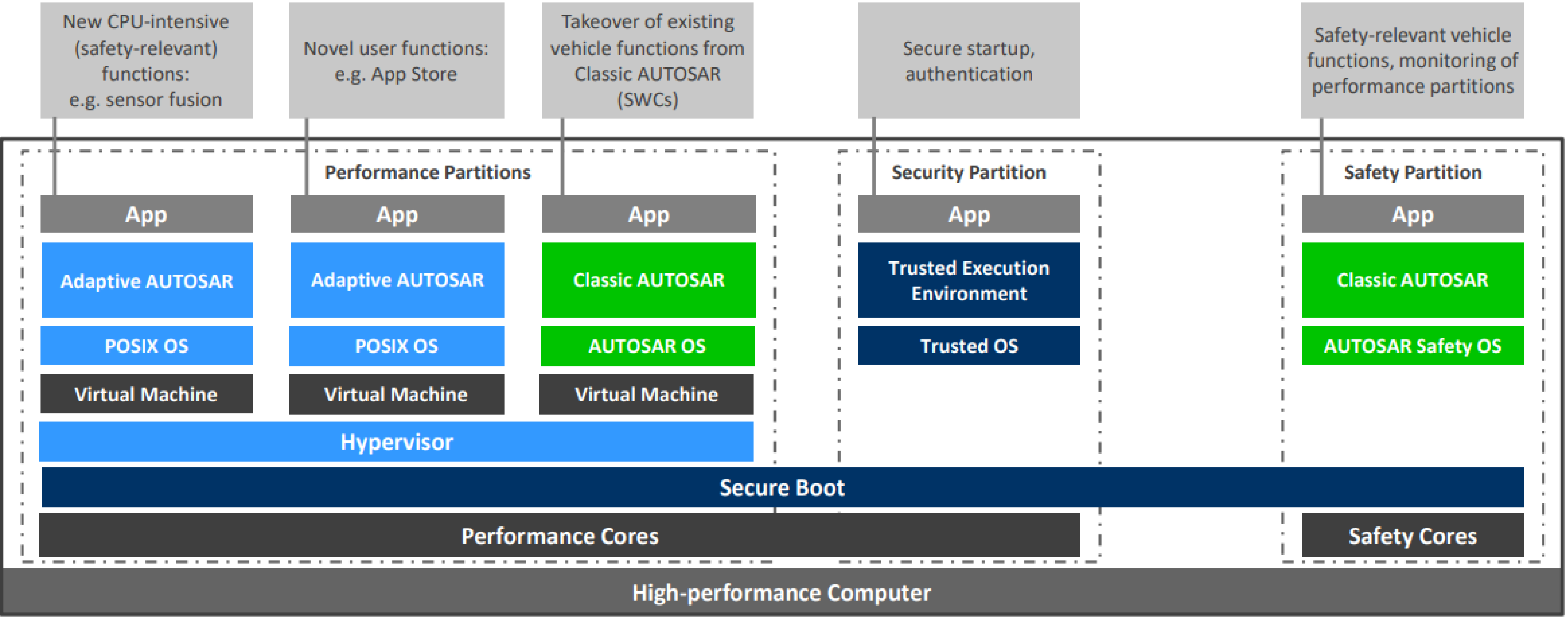


Figure 2-2 Exemplary interactions of AP and CP

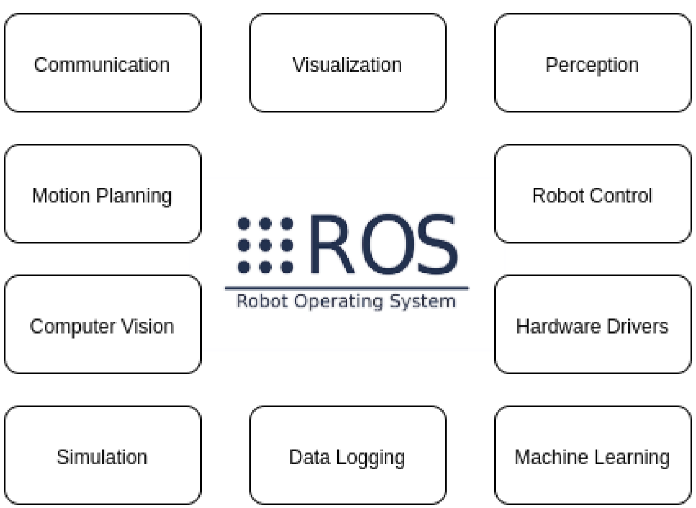
EB corbos

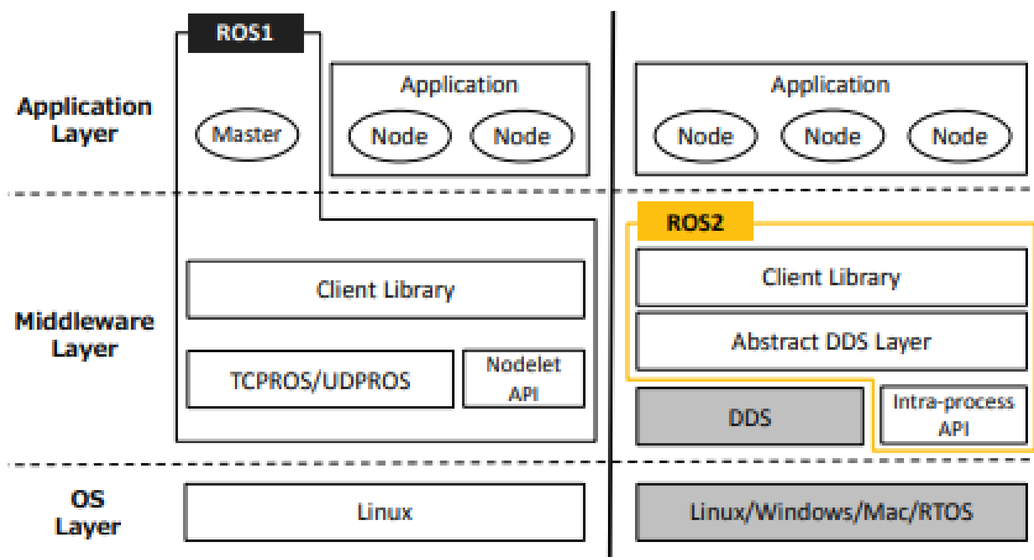
- EB corbos is a software platform that uses virtualization technology (hypervisor) to integrate AUTOSAR AP, CP and non-AUTOSAR OS on a single multicore ECU, while achieving high degree of isolation between different Virtual Machine partitions, including:
  - Performance Partitions with complex Performance Cores for high-performance, subject to low-levels of safety certification
  - Safety Partition with simpler Safety Cores for safety-critical functions, subject to high-levels of safety certification
  - Security Partition with processor security extensions (e.g., ARM TrustZone, Intel Software Guard Extensions (SGX)) for secure boot, crypto operations, etc.
- An example of a Mixed-Criticality System, where subsystems with different safety criticality levels are integrated on the same platform



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).
- A drawback of ROS compared to AUTOSAR:
  - Since ROS uses Linux as the OS, it is not possible to pass high-level of safety certification (ASIL-D).

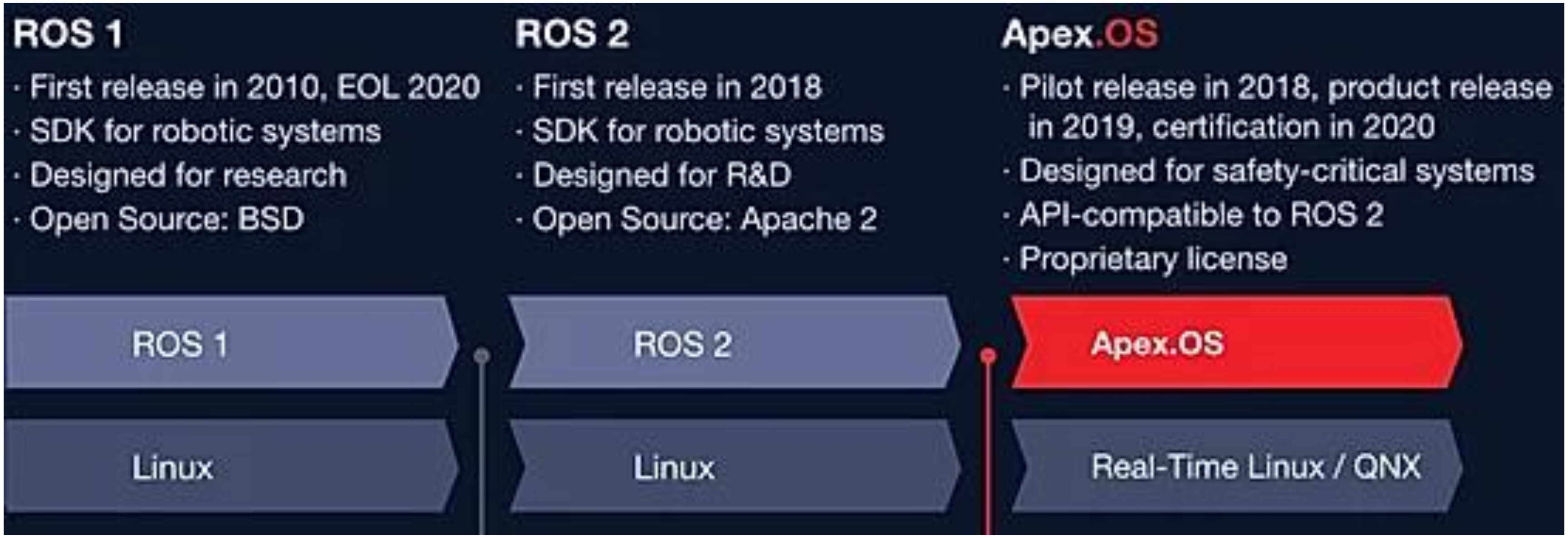




Maruyama Y, Kato S, Azumi T. Exploring the performance of ROS2[C]//Proceedings of the 13th International Conference on Embedded Software. 2016: 1-10.

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 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.



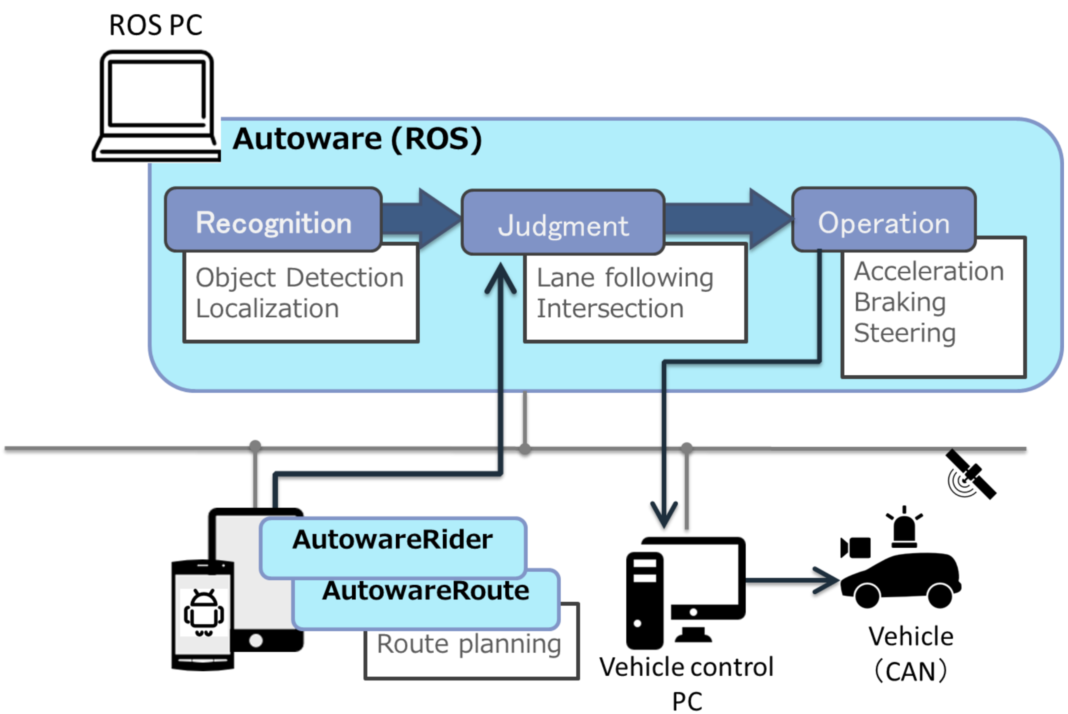
NVIDIA DRIVE Software Framework

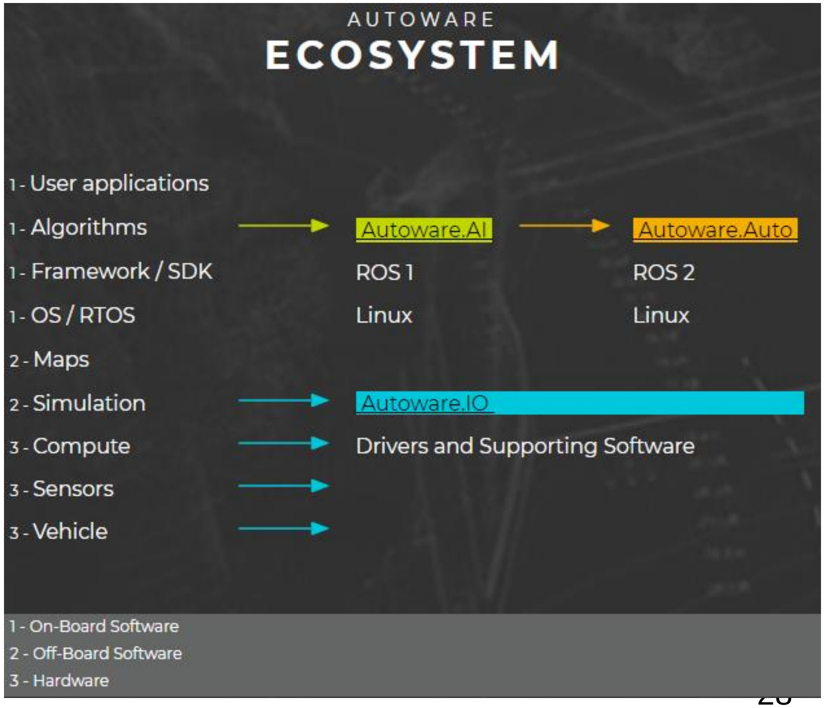
NVIDIA DRIVE Software Framework

- 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

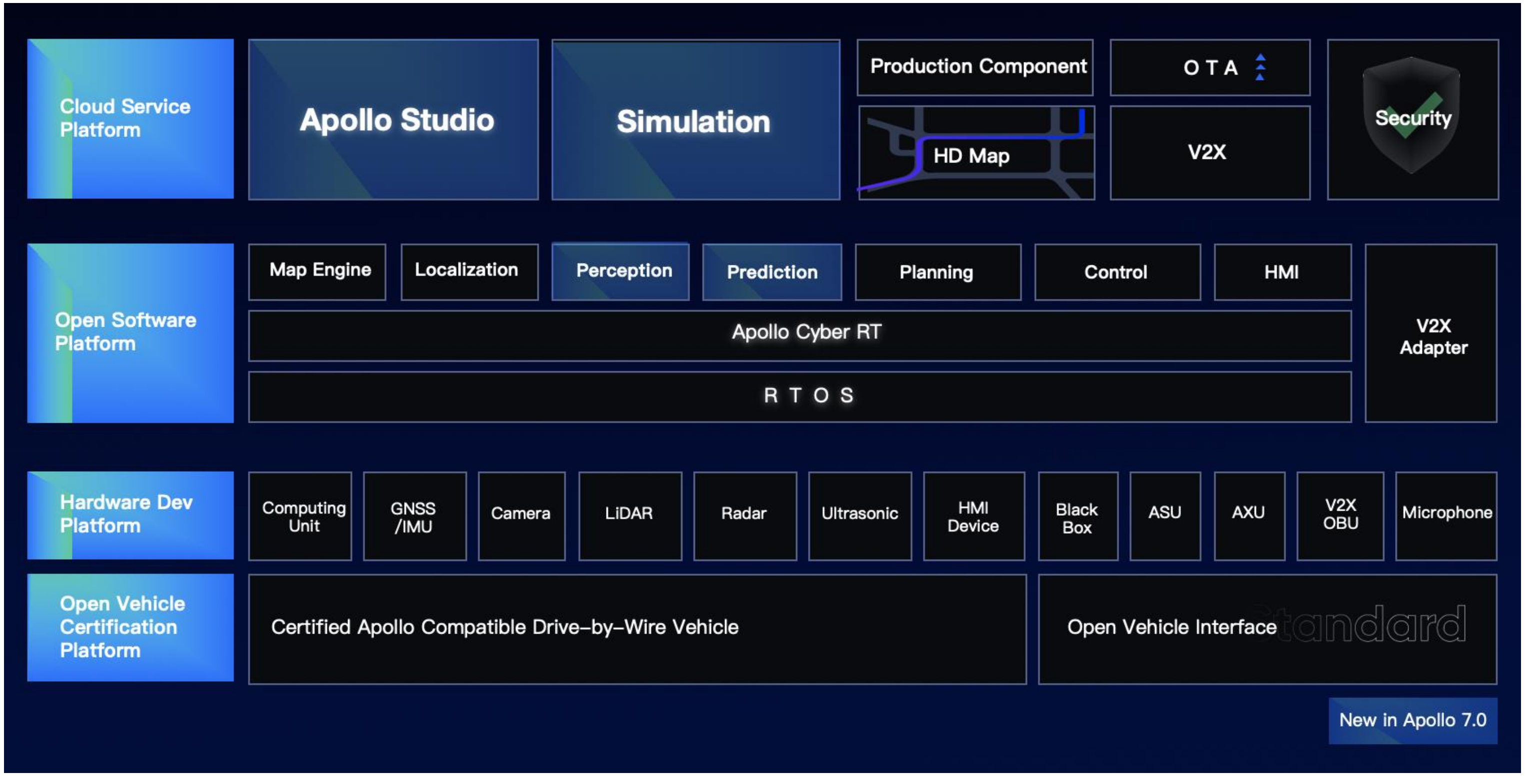
- Open-source AD platform from Japan.
  – Autoware.AI (https://www.autoware.ai) is based on ROS-1.
  – Autoware.auto (https://www.autoware.auto) is the new version based on ROS2.



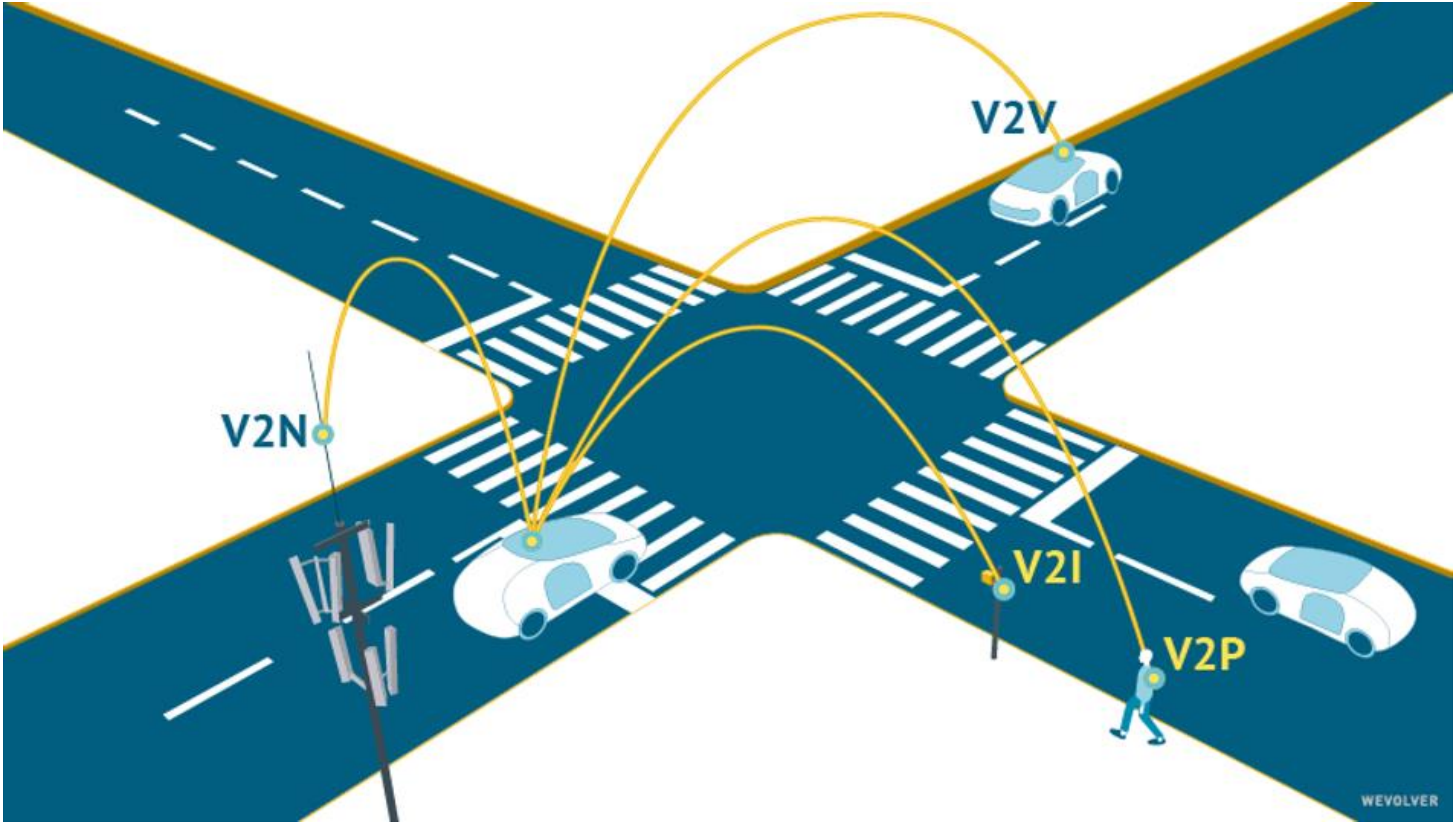


Baidu Apollo

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

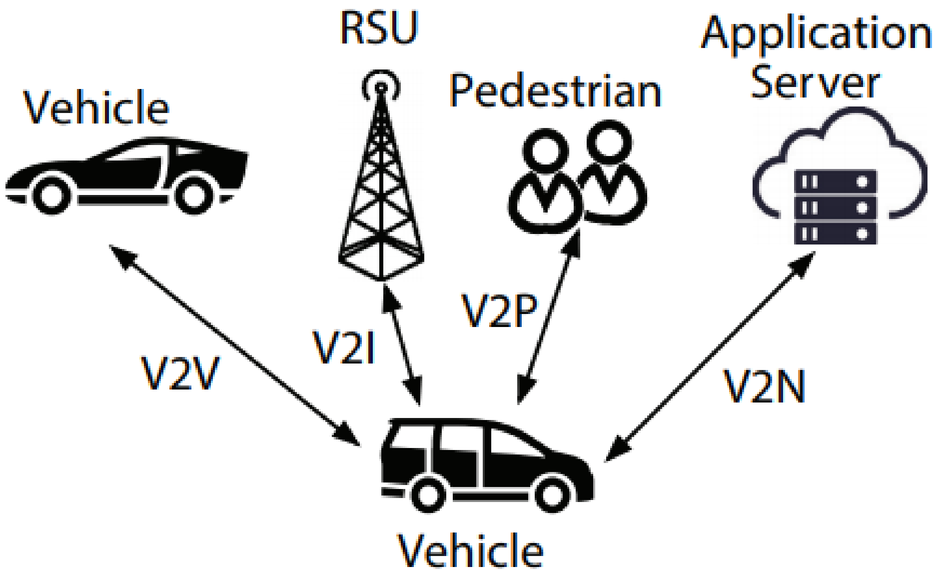


V2X



V2X Types

- Four types of V2X applications in 3GPP Standard:
  - Vehicle-to-Vehicle (V2V)
    - e.g., collision avoidance system
  - Vehicle-to-Pedestrian (V2P)
    - e.g., safety alerts to pedestrians and bicyclists
  - Vehicle-to-Infrastructure (V2I)
    - e.g., adaptive traffic light control, traffic-light optimal speed advisory
  - Vehicle-to-Network (V2N)
    - e.g., real-time traffic routing, cloud services
    - Also called Vehicle-to-Cloud



Two Camps

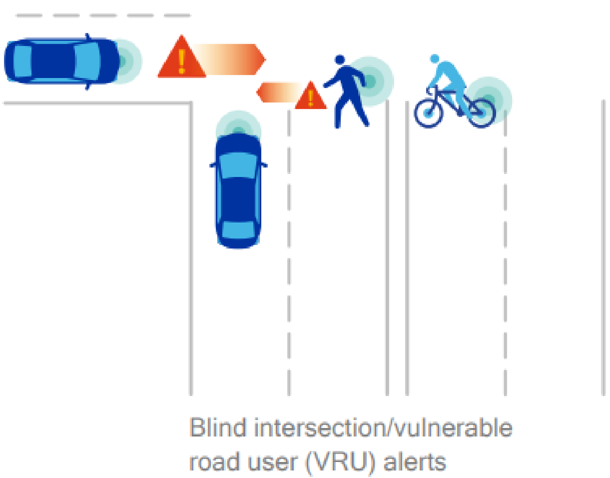
- DSRC (Distributed Short-Range Communication): Toyota, GM…
  – Based on WiFi, i.e., 802.11p at 5.9GHz
  – For latency-sensitive applications.
- C-V2X (Cellular V2X): Qualcomm/Ford…
  – Traditionally for latency-tolerant applications e.g., Over-the-Air (OTA) updates.
  – With the advent of 5G, C-V2X will become more prevalent, used also for latency-sensitive applications.
- C-V2X seems to be winning over DSRC in recent years. The following slides are based on Qualcomm’s C-V2X approach

https://www.qualcomm.com/media/documents/files/accelerating-c-v2x-commercialization.pdf

V2X Applications in AD

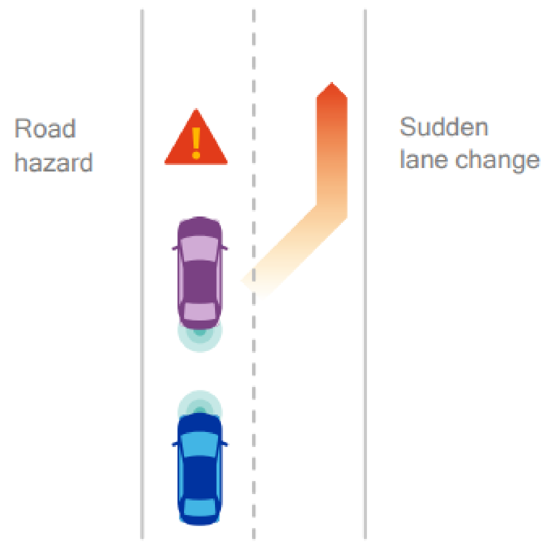
Non line-of-sight sensing

Provides 360° NLOS awareness, works at night and in bad weather conditions



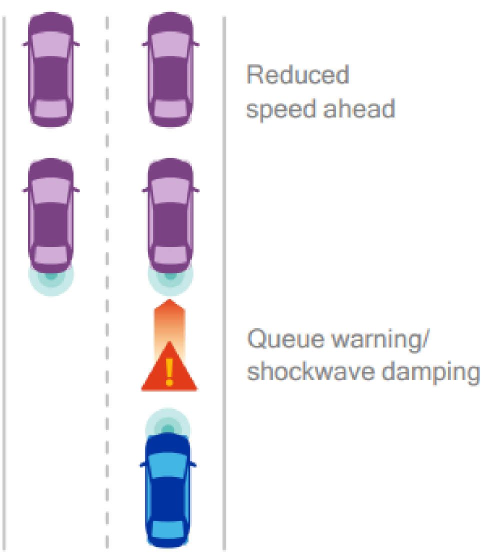
Conveying intent

Shares intent, sensor data , and path planning info for higher level of predictability



Situational awareness

Offers increased electronic horizon to support soft safety alerts and graduated warning



C-V2X Evolution towards 5G

- In 2020/07, 3GPP (3rd Generation Partnership Project) declared R16 to be frozen

	D2D communications R12/13	Enhanced safety C-V2X R14 (Ph. I) C-V2X R15 (Ph. II)	Autonomous driving C-V2X R16 5G NR support (Ph. III) (Advanced safety applications)
	Established foundation for basic D2D comm.	Enhanced communication’s range and reliability for V2X safety	Ultra-reliable, low latency, high throughput communication for autonomous driving
Network independent	No	Yes	Yes
Communications[1]	Broadcast only	Broadcast only	Broadcast + Unicast/Multicast
High speed support	No	Yes	Yes
High density support	No	Yes	Yes
Throughput		High throughput for enhanced safety	Ultra-high throughput
Latency		Low latency for enhanced safety applications	Ultra-low latency
Reliability		Reliability for enhanced safety application	Ultra-high reliability
Positioning	No	Share positioning information	Wideband ranging and positioning

C-V2X Defines 2 Transmission Modes

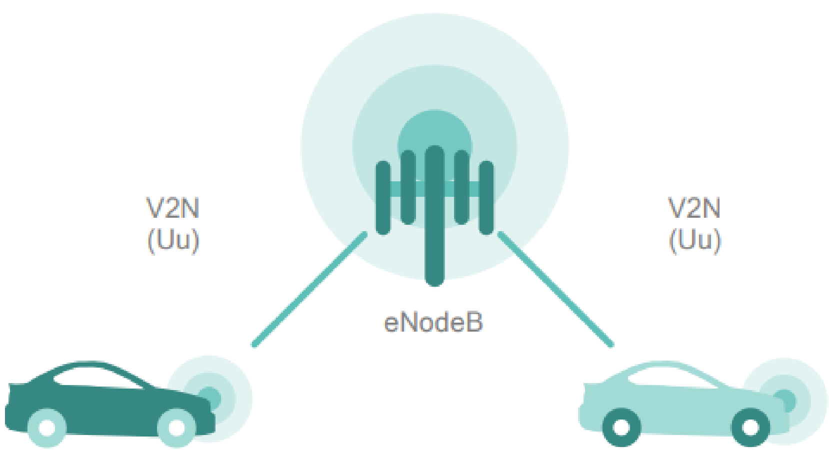
- V2X-Cellular: network communications going through base station (eNodeB)
- V2X-Direct: Device-to-Device direct communications without going through base station

Network communications

V2N on “Uu” interface operates in traditional mobile broadband licensed spectrum

Uu interface

e.g. accident 2 kilometer ahead

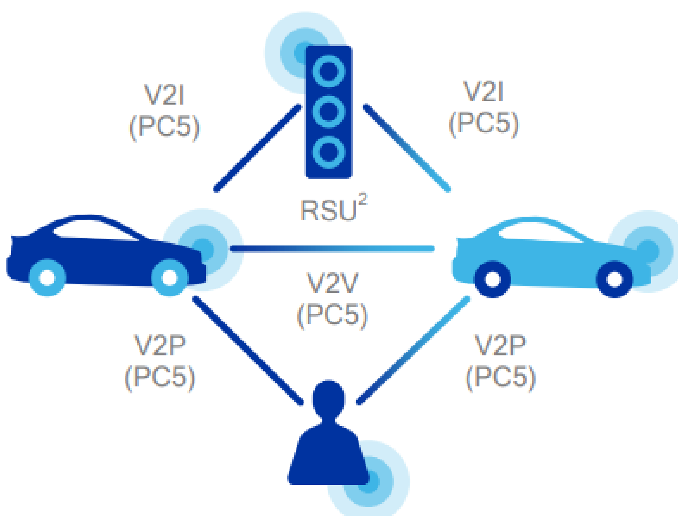


Direct communications

V2V, V2I, and V2P on “PC5” interface[1], operating in ITS bands (e.g. ITS 5.9 GHz) independent of cellular network

PC5 interface

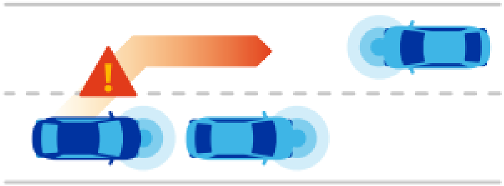
e.g. location, speed



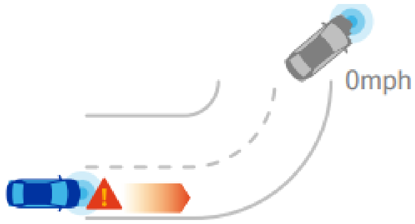
V2X-Cellular for Latency-Tolerant Use Cases

V2X-Direct for Latency-Sensitive Active Safety Use Cases

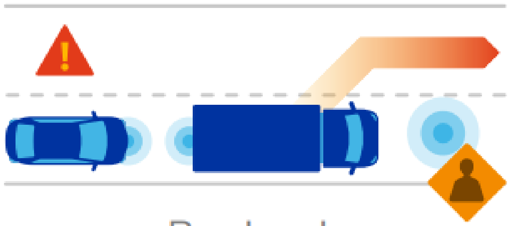
- Useful for NLOS (Non-Line-of-Sight) scenarios.



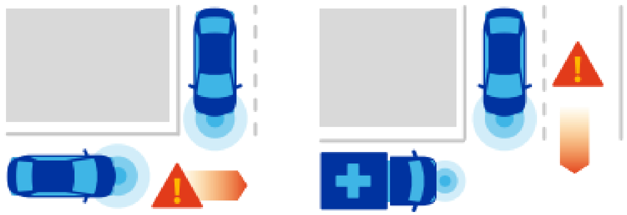
Do not pass warning (DNPW)



Blind curve/ Local hazard warning

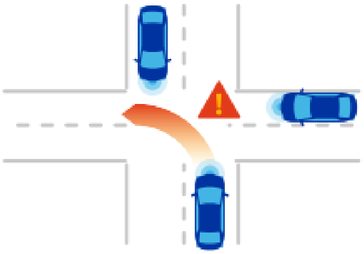


Road works warning



Intersection movement assist (IMA) at a blind intersection

Vulnerable road user (VRU) alerts at a blind intersection



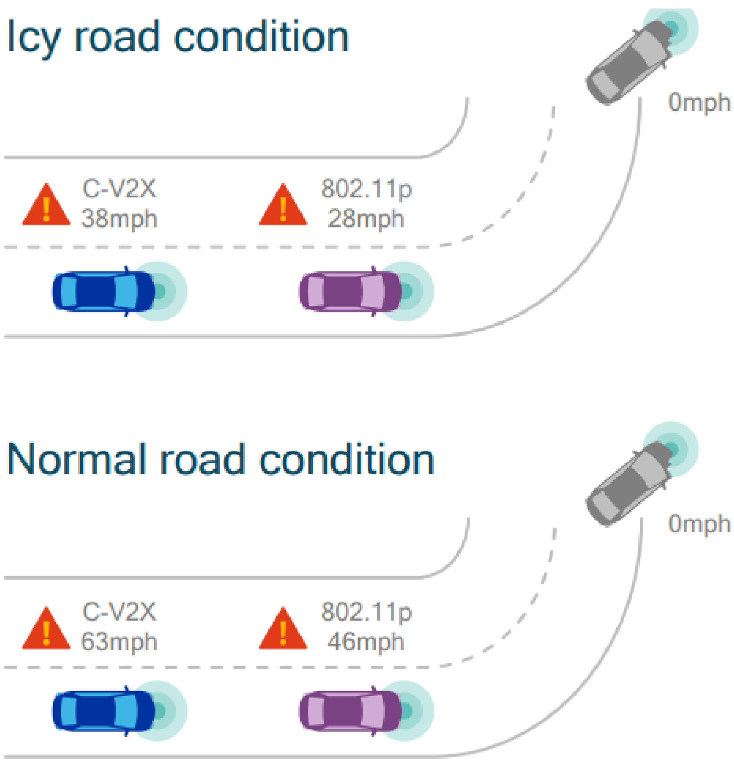
Left turn assist (LTA)

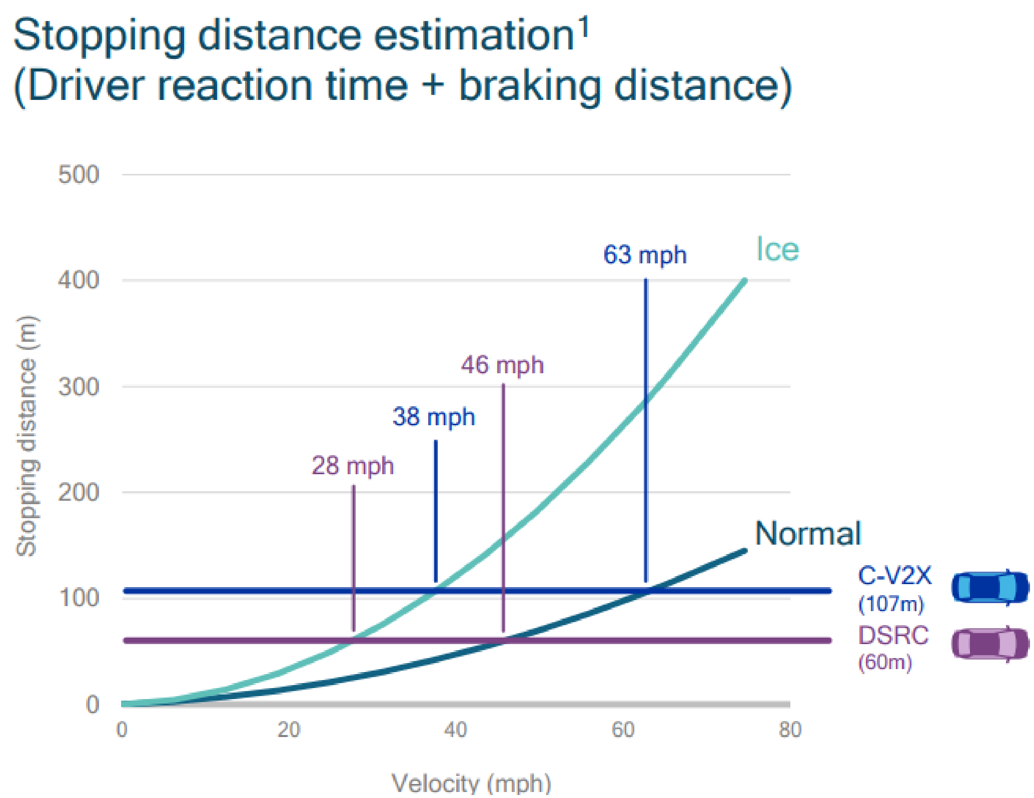
C-V2X Deployment

- Combined RSUs (Road-Side Units) with direct link (PC5) interface for V2X-Direct, and 4G/5G base stations for V2X-Cellular, benefiting from cellular network densification in 5G (smaller and denser cells)

Use Case: Disabled Vehicle after Blind Curve

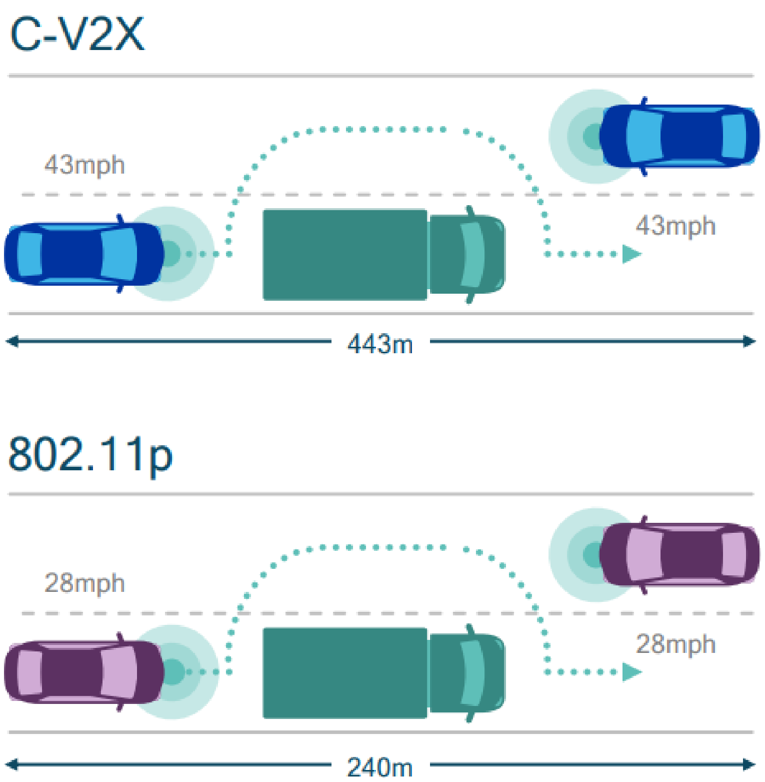
- C-V2X has (at least 2x) longer range than DSRC/802.11p, which enables the ego-vehicle to get warning message earlier, hence travel at higher speed while avoiding collision with the disabled vehicle



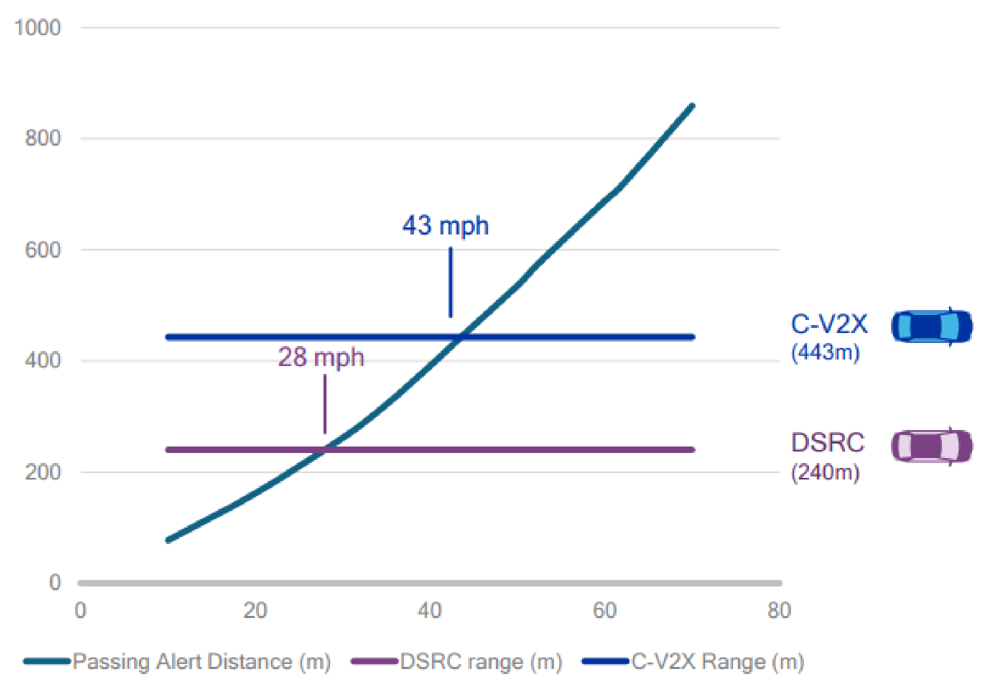


Use Case: Do Not Pass Warning

- C-V2X's longer range enables the ego-vehicle to get warning message earlier, hence travel at higher speed while avoiding collision with the disabled vehicle



Required passing alert distance (m) vs. speed (mph)[1]



Industry Consortium

- 5GAA is a cross-industry consortium that defines 5G V2X communications

5GAA
Automotive Association

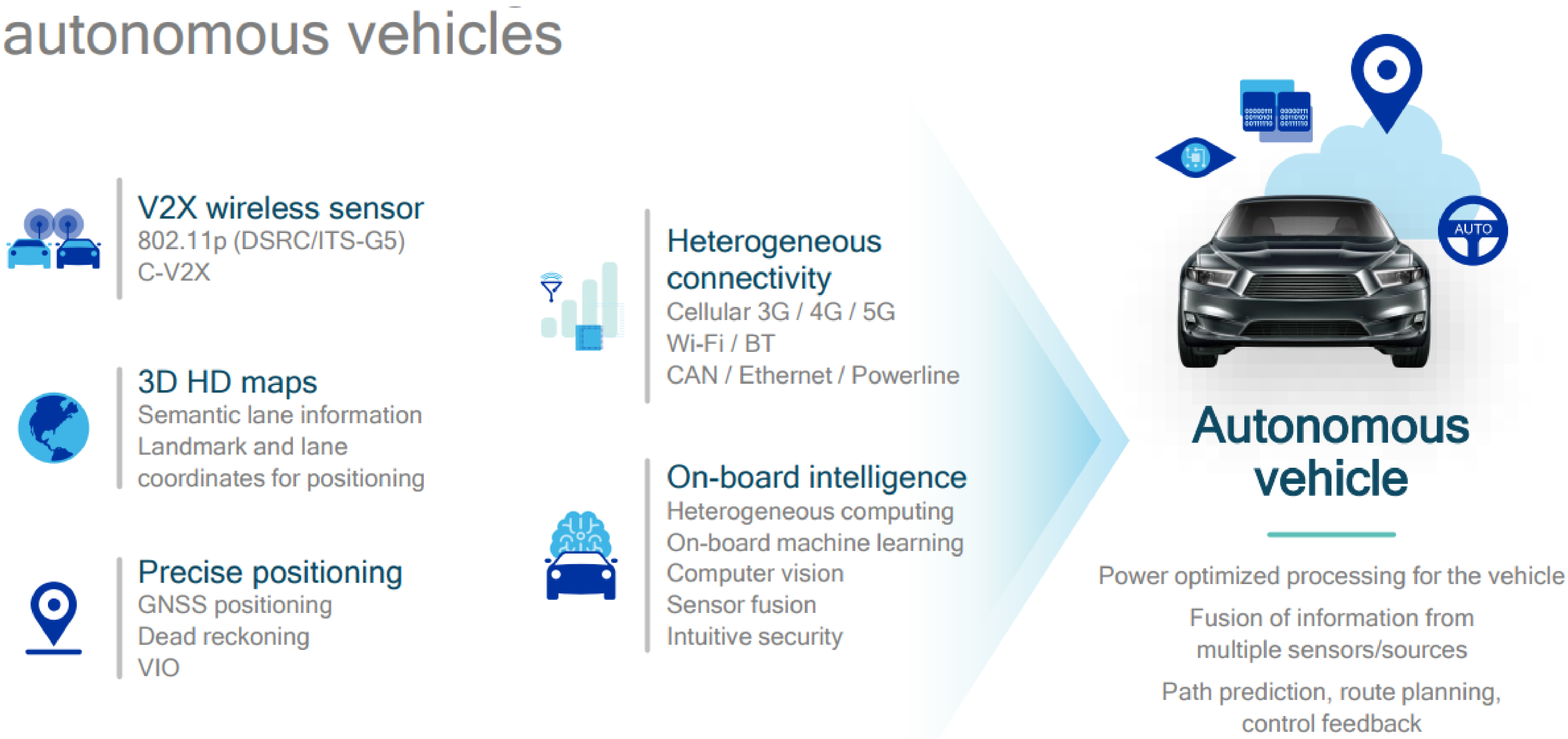
Automotive industry
Vehicle platform, hardware, and software solutions

Telecommunications
Connectivity and networking systems, devices, and technologies

End-to-end solutions for intelligent transportation mobility systems and smart cities

Analog Devices AT&T Audi BAIC BMW Bosch CAICT CETECOM China Mobile Continental Daimler
Danlaw Denso Ericsson FEV Ficosa Ford Gemalto Hirschmann Car Communication Huawei Infineon
Intel Interdigital Jaguar KDDI Keysight Technologies KT Laird Land Rover LG MINI muRata Nokia
NTT DoCoMo P3 Panasonic Qualcomm Rohde & Schwarz ROHM Rolls-Royce SAIC Motor Samsung Savari
SK Telecom SoftBank T-Mobile Telefonica Telstra TÜV Rheinland Valeo Verizon VLAVI Vodafone ZF ZTE

5G Accelerates AVs



- My thoughts: Massive deployment of V2X is necessary for AVs to benefit from V2X; It is a promising technology, but current AV players are not counting on V2X.