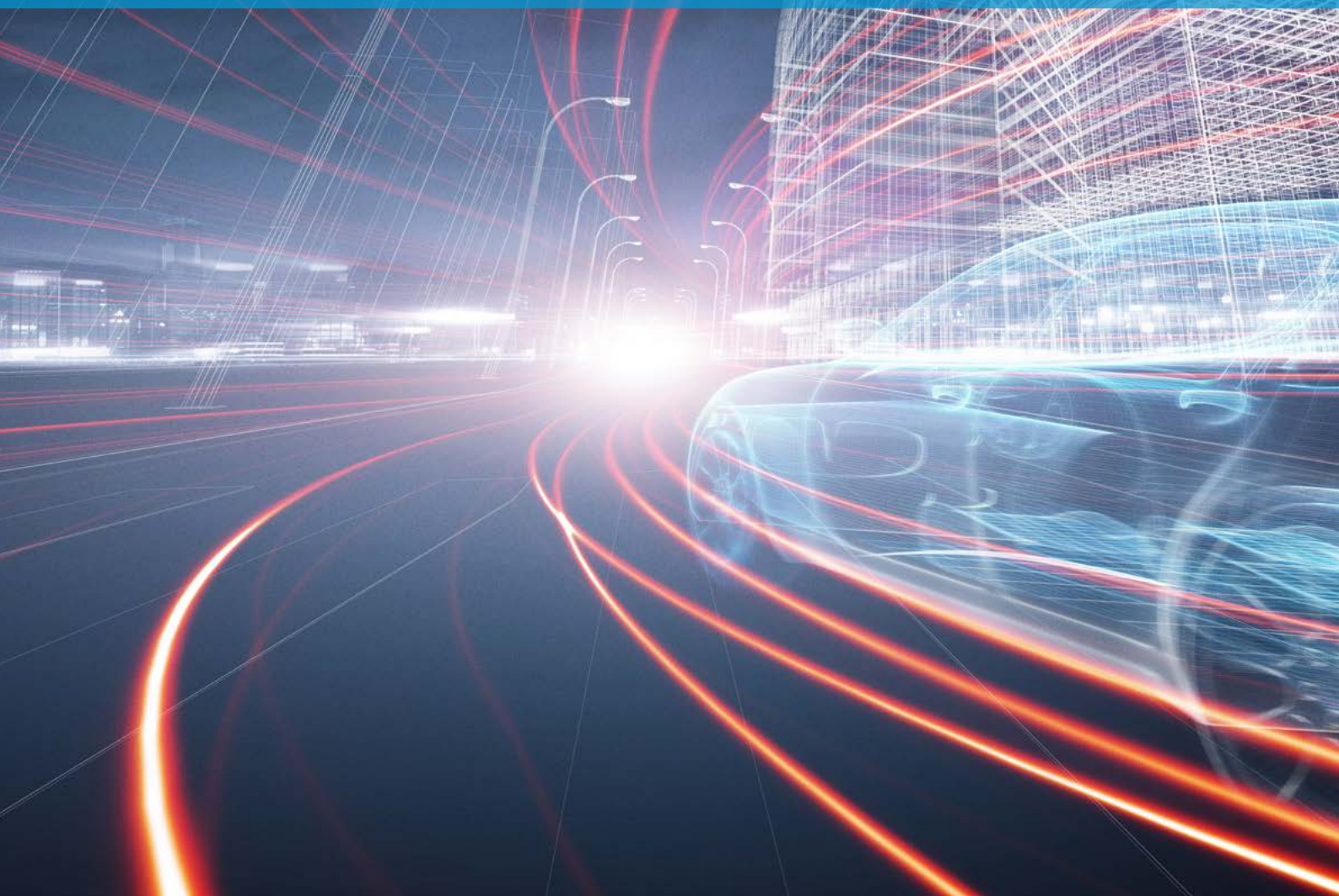
Cambridge Consultants

AI IN THE DRIVING SEAT:
KEY CHALLENGES FOR THE FUTURE OF ADVANCED DRIVER ASSISTANCE SYSTEMS (ADAS)

JUNE 2019



CONTENTS

1 EXECUTIVE SUMMARY

Advanced driver assistance systems (ADAS) is the fastest growing technology segment in the automotive market, worth an estimated $24 billion in 2018 and predicted to reach $92 billion by 2025.[1]

ADAS capabilities such as lane departure warning, blind spot detection and emergency braking have achieved widespread market penetration, significantly increasing driver safety and improving the driving experience. Meanwhile, more advanced semi-autonomous ADAS (S-ADAS) technologies such as GM's Super Cruise, Volvo's Pilot Assist and Tesla's Autopilot are paving the way for self-driving vehicles in the future.

Artificial Intelligence (AI) is one of the most critical components in ADAS. AI technologies such as deep learning vastly outperform traditional software techniques when it comes to real-time perception, prediction and decision-making tasks. As future ADAS systems tackle more complex self-driving scenarios, even more AI will be required.

However, the current approach to S-ADAS architecture, which is highly centralized, makes it prohibitively expensive for vehicles in the mid-range to adopt. These centralized architectures rely on powerful AI-enabled compute platforms, which are only available from a small number of vendors, and require high-bandwidth, low-latency vehicle networks to carry large volumes of sensor data to the central compute module.

An alternative S-ADAS architecture that distributes AI across multiple electronic control units (ECUs) would remove this barrier to entry while also enabling mid-range original equipment manufacturers (OEMs) to offer individual S-ADAS features to their consumers as modular, value-added options. However, this will require some technical innovation in order to optimize AI algorithms for resource-constrained computers.

In this whitepaper, we will examine the commercial and technological challenges to wider adoption of self-driving technology and set out our roadmap to delivering the S-ADAS features currently offered by premium OEMs into the mid-range vehicle segment.

"The current approach to S-ADAS architecture, which is highly centralized, makes it prohibitively expensive for vehicles in the mid-range to adopt."

2 ADAS – AN OVERVIEW

ADAS are electronic systems designed to assist the driver and, in some instances, assume control to increase the safety of the driver, passengers and other road users.

ADAS is enabled by technologies that allow on-board computers to perceive the external environment, through an array of on-board sensors and data processing. Cameras, radar, ultrasonic sensors and, in some instances, LiDAR collect vast amounts of data from the vehicle's surroundings. These distributed on-board sensors help provide the vehicle and the driver with 360° awareness.

As the range of many of these sensors, including long range radar, extends beyond that of human vision, today's ADAS has the ability to not only react to external events outside the driver's current line of sight, but also anticipate them much sooner than a human alone.

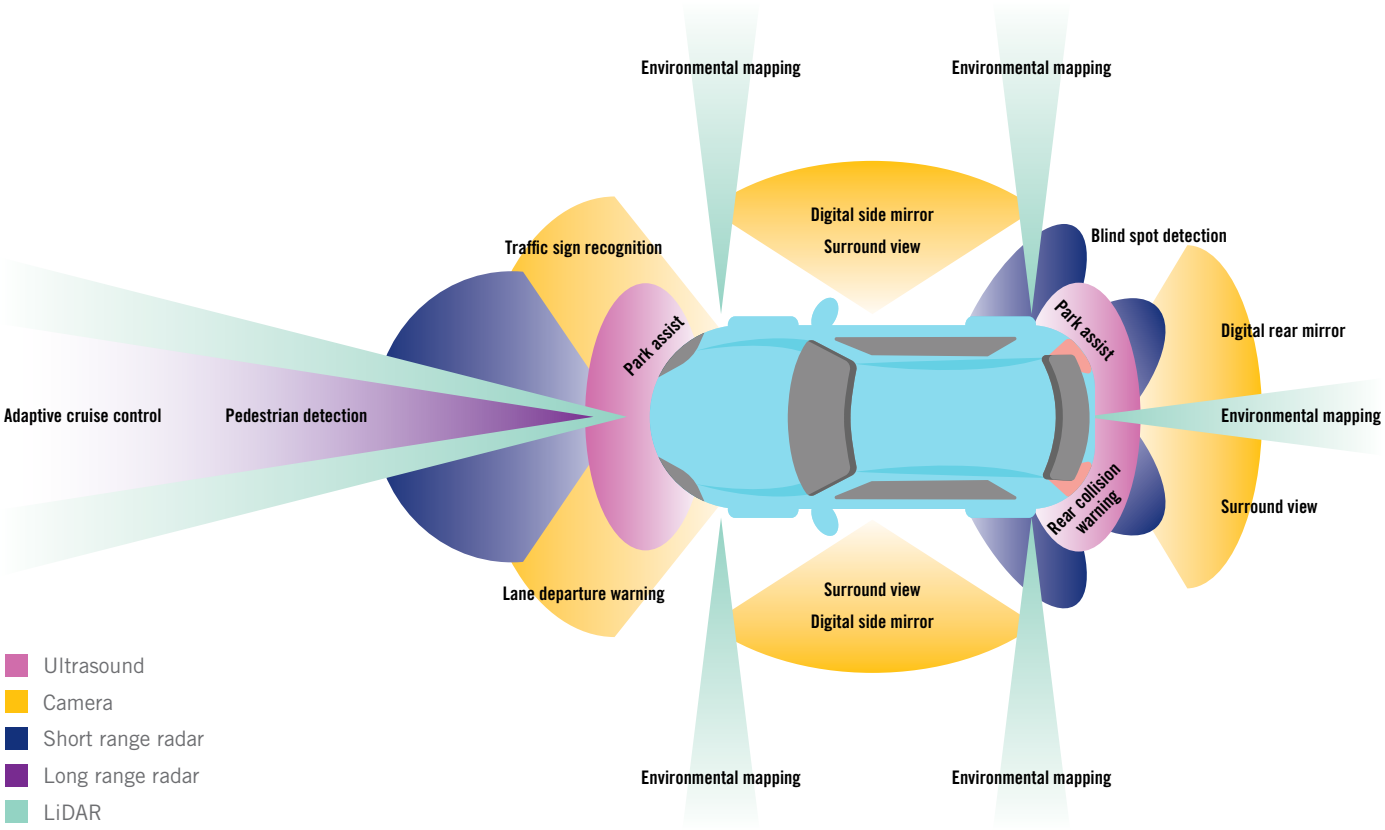


FIGURE 1: The array of distributed sensors on a modern vehicle and their fields of view

"ADAS is enabled by technologies that allow on-board computers to perceive the external environment, through an array of on-board sensors and data processing."

2.1 UNDERSTANDING THE FIVE LEVELS OF VEHICLE AUTONOMY

ADAS vary greatly in complexity. A useful way to classify ADAS is by the level of autonomy they provide, ranging from basic Informational ADAS, where the driver remains in full control, through current and next-generation S-ADAS, to the fully autonomous vehicles of the future.

"Despite its benefits to driver safety, Informational ADAS only facilitates Level 0 autonomy as the driver maintains control of all driving functions."

LEVEL	0	1	2	3	4	5
VEHICLE	Vehicle only provides warnings and momentary assistance	Vehicle provides steering OR brake/ acceleration support to the driver	Vehicle provides steering AND brake/acceleration support to the driver	Vehicle can operate autonomously under specific conditions	Vehicle can operate autonomously under limited driving conditions	Vehicle can operate autonomously under all driving conditions
DRIVER	Driver controls all functions	Driver must constantly supervise ADAS functions and resume control immediately if required	Driver must constantly supervise ADAS functions and resume control immediately if required	Driver must resume control immediately if required	No driver required during limited driving conditions	No driver required
ADAS FUNCTIONS	• Blind spot monitor • Lane departure warning • Forward collision warning • Parking sensors • Pedestrian detection	• Adaptive cruise control • Lane centering • Semi-automated parking assistance • Intersection assist	• Automatic lane change • Collision avoidance system • Dynamic driving assistance	• Autopilot • Automatic parking		

FIGURE 2: The ADAS functions enabling increasing levels of autonomy

2.1.1 INFORMATIONAL ADAS

So-called Informational ADAS is now equipped on most modern vehicles. These systems provide information to the driver and alert them to potential hazards, but do not assume control of the vehicle under any circumstances. Applications utilize cameras, radar and ultrasonic sensors to capture data from the vehicle environment. Microcontrollers (MCUs) and ECUs then help the vehicle to understand these sensory inputs to perceive its surroundings.

Informational ADAS are typically independent subsystems in which ECUs process information from individual sensors associated with the application. For example, the parking assist ECU will process ultrasonic information from the sensor and alert the driver if they are approaching an obstacle.

Despite its benefits to driver safety, Informational ADAS only facilitates Level 0 autonomy as the driver maintains control of all driving functions.

The useful sensor data for Level 0 vehicles will be very basic – for example, a distance sensor for proximity warning. This differs from the amount of information necessary to enable more advanced functions like automatic parking, where advanced data processing techniques and AI would also be required to provide object detection, identification and decision-making capabilities.

ADAS applications such as adaptive cruise control (in which the vehicle can autonomously maintain headway distance with the vehicle in front), collision avoidance systems and lane centering are required to combine data from multiple sensors and ECUs before signals are communicated to steering or braking ECUs. These systems can be considered Level 1 autonomous, as the vehicle can independently operate steering or braking in specific use cases.

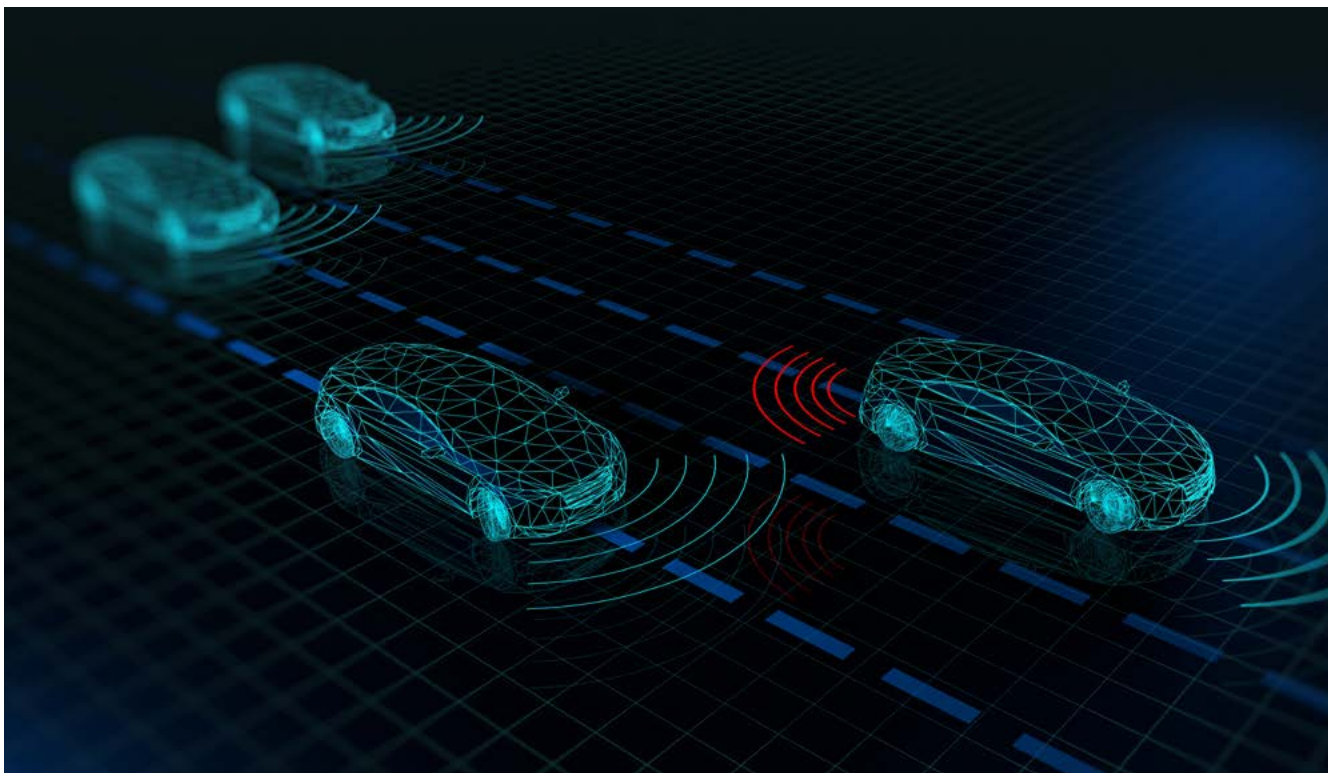
2.1.2 S-ADAS

S-ADAS currently deployed by companies, such as Audi, Tesla, GM (Cadillac) and Volvo, apply advanced algorithms and AI to data from multiple sensors to perform complex tasks, such as automatic lane changes, turning assist, advanced collision avoidance and automatic parking.

In controlled driving situations, such as slow-moving traffic or certain highway driving scenarios, these systems can perform lateral as well as longitudinal control and operate steering and braking functions autonomously. These applications reach Level 2 to 3 autonomy.

Tesla's Autopilot utilizes data from its forward radar, its array of forward, side and rear-view cameras together with its ultrasonic sensors to deliver Level 2 autonomy. These systems allow the vehicle to autonomously navigate complex environments, whilst Tesla's 'Enhanced Summon' feature enables the vehicle to locate and navigate to the driver in a parking lot.[2]

GM's Super Cruise provides capabilities such as lane keep assist and adaptive cruise control. This Level 2 feature is currently only available on Cadillac's top-end CT-6 model.[3]



Volvo's Pilot Assist works together with their vehicles' Adaptive Cruise Control (ACC) to offer lane keep assistance, provided there are clear lane markings on the road. The ACC feature maintains a set speed or time interval with respect to the vehicle in front, which is detected by a set of cameras and radar sensors.[4]

All S-ADAS systems rely on a continuous flow of accurate data from the sensor subsystems to control the vehicle safely. Emerging real-time AI approaches (see case study 1) can be used to mitigate the effects of distortion caused by dirt, dust, adverse weather and even sensor damage.

CASE STUDY 1 – AI MOVES BEYOND HUMAN VISION

Deep learning has the potential to transform the performance of ADAS systems, greatly increasing reliability, performance and safety. A recent breakthrough in this area is Cambridge Consultants' DeepRay™ technology, which mitigates the effects of distortion in sensor data. Based on advances in deep learning, DeepRay learns what real-world scenes and objects look like, from both an optically perfect viewpoint and with various image distortions applied.

When presented with a distorted real-time video feed it has never seen before, the technology can form a judgement in real time of the 'true' scene behind the distortion. This can prevent degradation in the performance of object detection, classification and tracking algorithms as a result of sensor wear and tear, dirt, poor weather, and other real-world conditions.

Future ADAS sensor subsystems, such as camera, radar and LiDAR, could utilize this technology to maintain the quality of sensor data, regardless of external factors.

DeepRay can remove the effects of distortion (left image) to drastically improve the performance of object detection and classification of vehicles (right image)

"The S-ADAS systems currently deployed by companies such as Audi, Tesla, GM (Cadillac) and Volvo apply advanced algorithms and AI to data from multiple sensors to perform complex tasks."

3 COMMERCIAL AND TECHNOLOGICAL CHALLENGES TO WIDER ADAS ADOPTION

3.1 CONSUMER CONFIDENCE

S-ADAS systems are a pre-cursor to a future where drivers will increasingly rely on self-driving ADAS technology. ADAS will have the ability to not only improve the individual driver experience, but also to reduce traffic congestion and tackle pollution. By co-ordinating and synchronizing with infrastructure and other vehicles, traffic flow can be optimized and made more efficient.

Despite S-ADAS being relatively new and only available from premium OEMs, we are seeing rising driver acceptance of self-driving technology. For example, a recent study by MIT's Human Centered AI Institute on Tesla journey data found drivers relied on Tesla's Autopilot for over 30% of their total miles driven. Indeed, Tesla states that over one billion miles have been completed by this feature.

To enable a self-driving future, the automotive market must adopt a multi-faceted approach. OEMs must demonstrate to the consumer the efficiency, convenience and safety of self-driving ADAS technology features in passenger vehicles. At the same time, they must continuously improve the sensing and AI technologies that are required to enable these features.

This means adopting a step-by-step approach, gradually increasing self-driving capabilities over several successive vehicle models, while at the same time collecting operational datapoints from installed ADAS systems to further improve the efficiency, reliability and performance of their core sensing and AI technologies.

As consumer confidence and trust in self-driving ADAS technology gradually increases, this continuous advancement in ADAS technologies will help to form a virtuous circle, enabling the next phase of autonomy. In this way, OEMs can begin to roll out their fleets of fully autonomous vehicles with confidence in consumer acceptance, as well as the reliability of their self-driving technology.

3.1.1 CONSUMER DEMAND FOR SELF-DRIVING FEATURES ON THE RISE

Currently, tech-savvy consumers show the greatest interest in self driving features, with 69% interested in using autonomous vehicles when they become available, compared to just 49% of all surveyed drivers.[5] For ADAS and autonomous vehicles to gain widespread adoption, consumer interest in autonomy must extend beyond tech-savvy early adopters and those in the premium segment who are willing to pay for such features.

Consumer interest varies by ADAS feature and their associated autonomy levels. For example, two of the most popular ADAS functions today are blind spot warning (Level 0) and parking assist (Level 1), which achieve 84% and 81% customer satisfaction respectively.[6] Informational ADAS features are now widely accepted whilst newer, more advanced ADAS features are gaining acceptance, with Level 1 features such as ACC and automatic emergency braking achieving 75% and 69% customer satisfaction respectively.[7]

As ADAS technology improves and such features become more cost-effective, consumer interest in the mid-range market segment will increase – 79% of drivers state they would choose an ADAS-enabled vehicle if it was available at no additional cost when compared with a conventional vehicle.[8]

"For ADAS and autonomous vehicles to gain widespread adoption, consumer interest in autonomy must extend beyond tech-savvy early adopters and those in the premium segment who are willing to pay for such features."

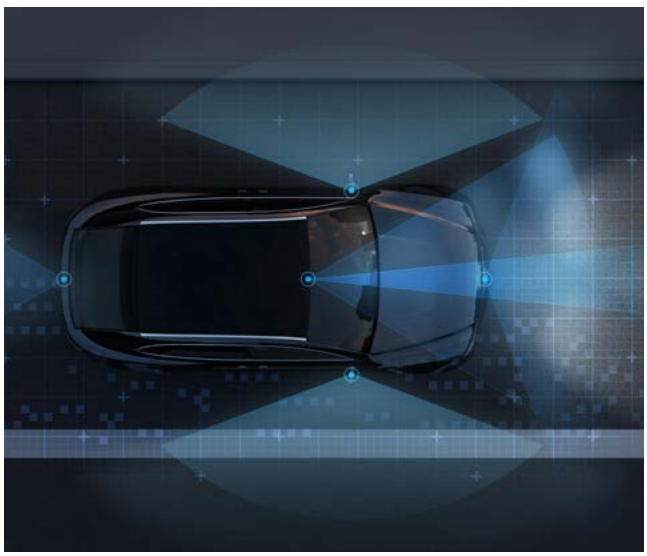
3.2 SUPPLY CHAIN, DEVELOPMENT AND TESTING

Automotive OEMs are accustomed to sourcing the required hardware and mechanical components from preferred Tier 1 suppliers. This is changing as more advanced technologies are integrated into new vehicle models.

As the ADAS content in vehicles increases, so does the software content. This is changing the automotive supply chain, with software focused entrants challenging more established suppliers. As a result, established suppliers are re-aligning their businesses to meet the challenges of integrating new technology and software content, whilst also establishing faster development lifecycles.

In future, the need for software testing is expected to be reduced as standards and regulations become established. Automotive safety standards, such as ISO 26262, should broaden their definition of hazards and include scenarios where harm is caused by behavioural interactions between humans and the vehicle.[9] The automotive industry also needs to incorporate new safety standards, such as ISO PAS 21448, into their design and development processes for the AI based algorithms required for self-driving ADAS systems. ISO PAS 21448 seeks to reduce the safety threat associated with increasing AI content in ADAS systems through analysis and verification.

In the immediate term, these challenges have forced OEMs to shift their strategy for in-vehicle software innovation to directly collaborating with disruptive software companies and start-ups.



Another trend for premium-range vehicles is the establishment of OEM partnerships to accelerate the development and launch of self-driving technologies. Companies, such as Audi, BMW Group and Daimler AG, are actively working together to develop next-generation technologies for ADAS. The collaboration between these companies shows the extent to which OEMs are willing to disrupt the supply chain to accelerate autonomous driving capabilities. As demand for self-driving ADAS grows in mid-range vehicles, OEMs will become increasingly likely to follow similar strategies and integrate S-ADAS technologies as cost effectively as possible.

3.3 POWER AND PROCESSING REQUIREMENTS

AI approaches, such as deep learning, are central to S-ADAS, but with them come significant additional requirements. AI adoption is already driving significant change in vehicle electrical architecture in order to handle the vast amounts of multimodal sensor data required, as well as the processing demands of the AI algorithms themselves.

The current generation of S-ADAS on the market follow a similar architecture, with centralized compute platforms running AI-based algorithms connected to high-bandwidth networks, which are feeding them a continuous stream of camera, LiDAR, radar, GNSS and INU data. These architectures are complex, power-hungry and, as a result, currently restricted to the premium automotive segment.

An alternative to centralized ADAS architecture will be needed to realize self-driving functions more widely in automotive, specifically in mid-range vehicles, enabling consumers to experience some of these capabilities at a cheaper price point. To facilitate an alternative ADAS architecture, more power-efficient AI-enabled processing platforms will be required. In addition, AI approaches, such as the deep learning neural networks required for ADAS perception, prediction and planning will need to be optimized to meet the memory, processing power, power consumption and cost requirements of the sensor subsystem ECUs that make up this ADAS architecture approach.

In the following sections, we examine technology strategies for delivering S-ADAS capabilities to the mid-range vehicle market.

4 ADAS ARCHITECTURE: CENTRALIZED VS DISTRIBUTED

4.1 COMPUTE ARCHITECTURE REQUIREMENTS FOR AUTOMOTIVE AI

Typical modern vehicles may contain up to 90 ECUs which process sensor data from individual subsystems.[10] For example, an ECU responsible for ultrasonic parking assist will process the ultrasound data to determine, at a basic level, the immediate surroundings of a vehicle during parking. These useful, but relatively simple, Informational ADAS functions are facilitated by local processing at the sensor node. These subsystems operate independently from each other and the central computer.

Increased levels of autonomy, where the ADAS has limited control of steering and braking functions, require communication between various different subsystems and ECUs. Whilst the majority of processing is still performed locally at the sensor node, individual subsystems must communicate with each other to enable vehicle automation.

For example, an ACC system may use cameras and radar to detect headway distance, and local processing to transform this raw data into object level data before communicating with the central ECU that governs steering and braking. This is typically performed via a Controller Area Network (CAN) which facilitates low-speed transfer of object level data.

For vehicles to execute S-ADAS functions, such as automatic lane changing (Level 2) and automatic parking (Level 3), the vehicle must be able to not only detect but also model and perceive the surrounding environment. Consequently, S-ADAS requires greater collaboration between the central decision taking ECUs and individual subsystems.

Two predominant architectures have emerged – centralized and distributed. Each architecture demonstrates unique advantages and disadvantages which must be considered from technical, commercial and regulatory perspectives.

4.1.1 THE EXPECTED STANDARD FOR NEXT GENERATION ADAS FOR PREMIUM OEMs

Today's S-ADAS architectures centralize the compute required to enable self-driving capabilities. As we have already seen, these architectures are complex, power-hungry and prohibitively costly for all but the premium automotive segment.

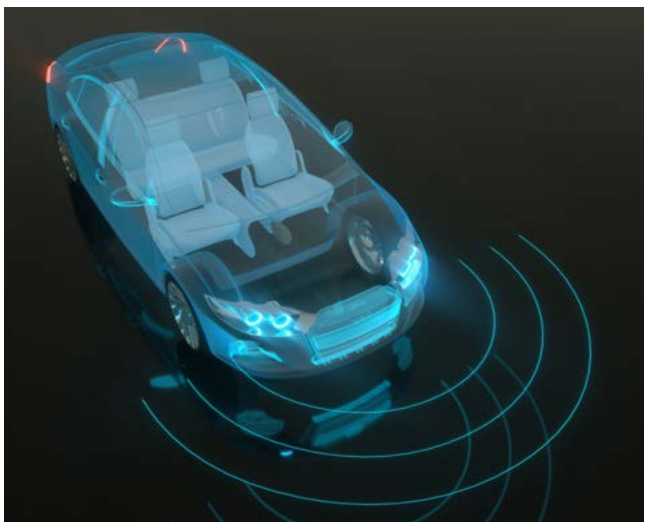
For vehicles to execute S-ADAS functions, the vehicle must combine information from an array of complimentary sensors and subsystems. For example, the automatic lane changing feature will combine inputs from forward-facing cameras that detect lane markings with rear- and side-view cameras, as well as radar to detect vehicles in the blind spot, before issuing commands to the central ECU to operate steering and braking functions.

As S-ADAS operate an increasing number of driving functions, computationally intensive sensor fusion processing is required, merging inputs from multiple sensors to increase the reliability and robustness of the system. Sensor fusion processing can be performed on raw sensor data or on object level data that has been pre-processed locally.

Centralized architectures are commonplace for premium vehicles that offer high levels of autonomy. Tesla, Audi and BMW utilize a centralized computing architecture to facilitate complex signal processing, sensor fusion and decision-making.[11] OEMs such as Volvo are integrating NVIDIA's Drive PX 2 computing platform into the central ECU.[12] The platform features 12 CPU cores, providing 8 teraflops (TFLOPs) of performance, equivalent to 150 MacBook Pros. The platform is capable of processing 2,800 images per second, enabling image recognition and object detection on camera feed data. Tesla, on the other hand, has recently announced it will develop its own chip platform for its Autopilot Hardware system. Another major automotive chipmaker, NXP, has announced its own Bluebox centralized platform solution, capable of up to 90,000 Dhrystone million instructions per second (DMIPS).[13]

Centralized architecture offers many advantages over the distributed architecture found in less technically advanced vehicles. Centralizing computing operations results in a straightforward system architecture, reducing design complexity. Somewhat counterintuitively, centralizing computations also reduces system latency as more powerful central computers can process data faster than MCUs located at the sensor node. Sensor fusion of raw data captured by the sensor nodes ensures no potentially useful data is lost during pre-processing. Moreover, centralized architectures are agnostic to sensor vendor and therefore components from different suppliers can be integrated easily into the same platform.

Despite the many advantages offered by centralized computing, there are a number of disadvantages that should be carefully considered. Powerful centralized computing platforms are expensive due to silicon area and are unlikely to be adopted by mid-range vehicles in the short term. Furthermore, centralizing all computing operations means that vehicles must be equipped with a high-bandwidth, low-latency communication infrastructure, which is also costly and power hungry. Additionally, centralized systems are far less scalable than modular, distributed architectures, making them potentially more expensive to repair or replace.



"Distributing some or all intelligence throughout the vehicle enables individual ADAS functions to be integrated incrementally, offering greater cost and power scalability."

4.1.2 DISTRIBUTED COMPUTE AND HYBRID ARCHITECTURES – A WAY FORWARD FOR MID-RANGE OEMs

The high cost of centralized ADAS architecture means an alternative is required for mid-range vehicles if consumers are to experience some of these self-driving capabilities at a cheaper price point. This alternative approach involves distributing some or all of the intelligence to the sensor nodes to facilitate local processing of raw sensor data. This results in a hybrid or fully distributed system architecture. For example, a camera could perform its own object detection locally, only sending the relevant information to the central computer for further processing and decision making.

While this approach has the potential to increase latency, careful hardware and software architecting can ensure the system still meets the latency requirements for safety-critical applications.

Distributed and hybrid architectures in which some or all sensor data is pre-processed locally can also reduce bandwidth requirements because only important object level data must be transferred to the central ECU. This reduces the requirement for expensive, high-power centralized computers and high-bandwidth communication infrastructure. Distributing processing to multiple, decentralized processors can also significantly reduce power consumption, reducing the need for expensive cooling systems required by centralized computing architectures.

So, despite increasing system complexity, distributed and hybrid architectures can nevertheless reduce the overall cost of integrating S-ADAS functionality, lowering the barrier to entry for mid-range OEMs that wish to improve vehicle capability.

Furthermore, distributing some or all intelligence throughout the vehicle enables individual ADAS functions to be integrated incrementally, offering greater cost and power scalability when compared to the centralized architecture. This approach can enable mid-range OEMs to offer individual ADAS features as value-added functions, allowing consumers to customize vehicles with the capabilities they desire in the most cost-effective manner.

In the near to mid future, we foresee mid-range OEMs focusing on hybrid architectures for S-ADAS. This will allow them to exploit some of the technology development from, and driven by, premium OEMs today, as well as benefit from some of the advantages of distributed S-ADAS architectures.

5 ENABLING COST-EFFECTIVE AI IN HYBRID ADAS ARCHITECTURES

In addition to the topological question of *where* signal processing should take place within the system (centralized or distributed), the computationally demanding AI features that enable higher levels of vehicle autonomy will increasingly force ECU designers to also reconsider *how* that processing is performed.

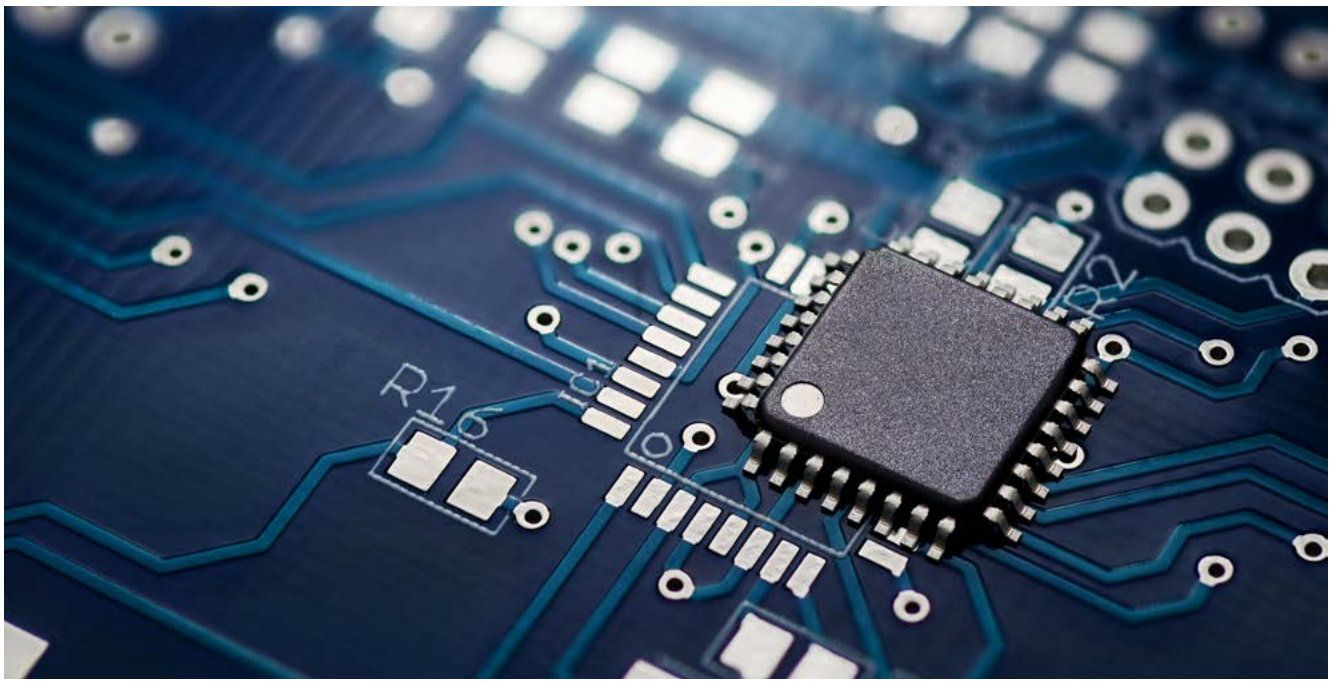
In the following sections we explain how specialized system-on-chip (SoC) designs and optimized AI algorithms can be used to reduce the computational demands, development costs and deployment costs associated with S-ADAS and fully autonomous ADAS in hybrid architectures.

5.1 OPTIMIZING SYSTEM-ON-CHIP (SOC) TO RUN AI APPLICATIONS

An SoC is a microchip or integrated circuit (IC) that combines all the required electronic circuits and components of an electronic system on a single chip. A desktop computer, for example, may have a central processing unit (CPU), video card, and sound card that are connected by different buses on the motherboard. This is an example of a system with many distinct components. An SoC would combine all these components and more into a single chip.

Traditional CPU and microcontroller architectures, ubiquitous in current ECUs, are not well suited to AI tasks like the advanced machine vision and deep learning that are necessary in S-ADAS functions. Automotive SoC devices for S-ADAS must process large quantities of raw signal data in real time, especially when running AI applications, and must therefore incorporate architectures with very high degrees of parallelism – in other words, architectures that are highly optimized for completing compute tasks in parallel rather than sequentially.

Where applications are still in the rapid early stages of development, graphics processing units (GPUs) can provide powerful parallel computation while maintaining general purpose programmability. As applications mature, opportunities arise to reduce the power and cost of automotive SoCs by replacing GPUs with dedicated vision processors and neural network accelerators.



5.1.1 PARALLEL COMPUTATION

AI applications are typically computationally intensive and data hungry and autonomous vehicles are no exception. The safety and efficacy of highly automated driving systems depends on their ability to extract high-level information from raw sensor data, in real time and with low latency.

The computational complexity of this task is the focus of much debate and continuing innovation, but it is typically measured in the hundreds of trillions of operations per second (TOPS). The processing power of an individual CPU is measured in the billions of operations per second. However, this does not mean that highly automated vehicles must contain hundreds of thousands of CPUs.

The majority of this computational complexity stems from the number of mathematical operations required, rather than the state space of the decisions to be made. Three 1080p 60fps video cameras would produce over a billion samples for processing per second. This sizeable demand on computation is then compounded by the requirements of the deep neural networks, which make crucial machine learning features such as object recognition possible.

The good news is that, for large sections of the computation, the same mathematical operations can be applied in parallel to each data point. This is what makes AI applications ideal for densely packed parallel processing architectures, such as GPUs or single input multiple data (SIMD) processing arrays.

By taking advantage of parallel processing, and without having to lay down hundreds of thousands of microcontrollers, several vendors have already brought to market platforms that boast performance in the tens or hundreds of trillions of operations per second range. NVIDIA, for example, offers both a 30 TOPS platform (Drive AGX Xavier) and a 320 TOPS platform (Drive AGX Pegasus).

Note that microcontrollers and CPUs in automotive SoCs are supplemented by parallel processors, not replaced by them. CPUs are still the backbone of decision-making, communication and orchestration, but they can be augmented by devices better suited to highly parallel digital signal processing.

Further advantages of parallel platforms include their high computational speed, which would allow the operation of more sensors simultaneously. On the other hand, platforms such as NVIDIA's Drive PX Pegasus can require up to 500W of power.[14] What is more, the maximum speed of GPU to CPU data transfer is likely to place constraints on overall system performance.[15]

PROCESSING PLATFORM	COMPANY	HEADLINE PERFORMANCE (TOPS)
Jetson AGX Xavier	NVIDIA	32 TOPs
Jetson TX2	NVIDIA	1.3 TOPs
Snapdragon 845	Qualcomm	2 TOPS
EyeQ4	ST / MobilEye (Intel)	2.5 TOPS (3W)
S32V234	NXP	0.1 TOPS (estimated)
TDA2Px	TI	0.1 TOPS (estimated)
Movidius	Intel	4 TOPs

TABLE 1: Automotive semiconductor vendors offering AI accelerated devices today

5.1.2 SPECIALIZED ACCELERATION

More often than not, what SoC designers regularly refer to as "*optimization*" is really a trade-off between silicon area (which governs cost and power) and application flexibility. The more confident you are in the maturity of your algorithm, the happier you might be to carve elements of it into silicon, giving away your option to change your mind in exchange for a cheaper and lower-power device.

In the case of deep learning inference applications, particularly convolutional neural networks, this trade-off can be mitigated by replacing general purpose programmable parallel processing architectures (such as GPUs or SIMD arrays) with dedicated neural network accelerators. This is because the structure of the different computation layers is well established and can be mapped easily to dedicated hardware modules. The intelligence that is continuously being improved upon is in the models themselves, which remain programmable.

5.2 OPTIMIZING NEURAL NETWORKS TO RUN IN COST-OPTIMIZED SOCS

Deep neural networks have achieved great success in tasks such as perception and decision-making, surpassing human-level performance for image recognition.[16] Neural networks are a specific set of algorithms that are inspired by our own biological neural networks and rely on millions or even billions of parameters.

Deep neural networks are typically organized into layers. These layers are made up of interconnected nodes which are formed of weights and biases. These weights are updated when the neural network is being trained, to optimize the system against a given task. As the field has matured, different designs of layers have proven more adept at different types of tasks. One commonly used layer in visual processing tasks is the convolution layer, and convolution neural networks have made significant progress in image classification, object detection and many other applications.

However, most high-performing examples of deep learning are computationally and memory intensive. This is at odds with the demand to run in a distributed architecture where low memory resources or strict latency requirements exist.

In the area of image classification (assigning an image to a predefined class), some of the best known neural networks have significant floating number multiplication requirements.

ARCHITECTURE	IMAGE INPUT SIZE (PIXELS)	FLOPS
Alexnet	227x227	0.7 GFLOPS
Googlenet	224x224	2 GFLOPS
Resnet-152	224x224	11 GFLOPS
VGG-16	224x224	16 GFLOPS

TABLE 2: Floating point operations per second (FLOPS) for well-known deep learning implementations

One area of active research is compressing and accelerating deep neural networks – i.e. optimizing AI algorithms to run in highly embedded resource-constrained computers (see case study 2). While necessitating a trade-off between efficiency and flexibility, this approach has the potential to greatly reduce compute requirements. In a breakthrough paper, the SqueezeNet neural network demonstrated accuracy comparable to a network 50 times its size when tested on a benchmark dataset.

In addition to the compression and acceleration of existing neural networks, it is also possible to architect a more compact neural network from the very start, creating an optimized network capable of performing the desired task at a similar level of accuracy. Neural architecture searches can automate the task of designing an architecture optimized to be as small as possible. In place of a human designer, it searches for the optimal architecture, using techniques such as reinforcement learning, genetic algorithms, differentiable searches, or other learning algorithms.

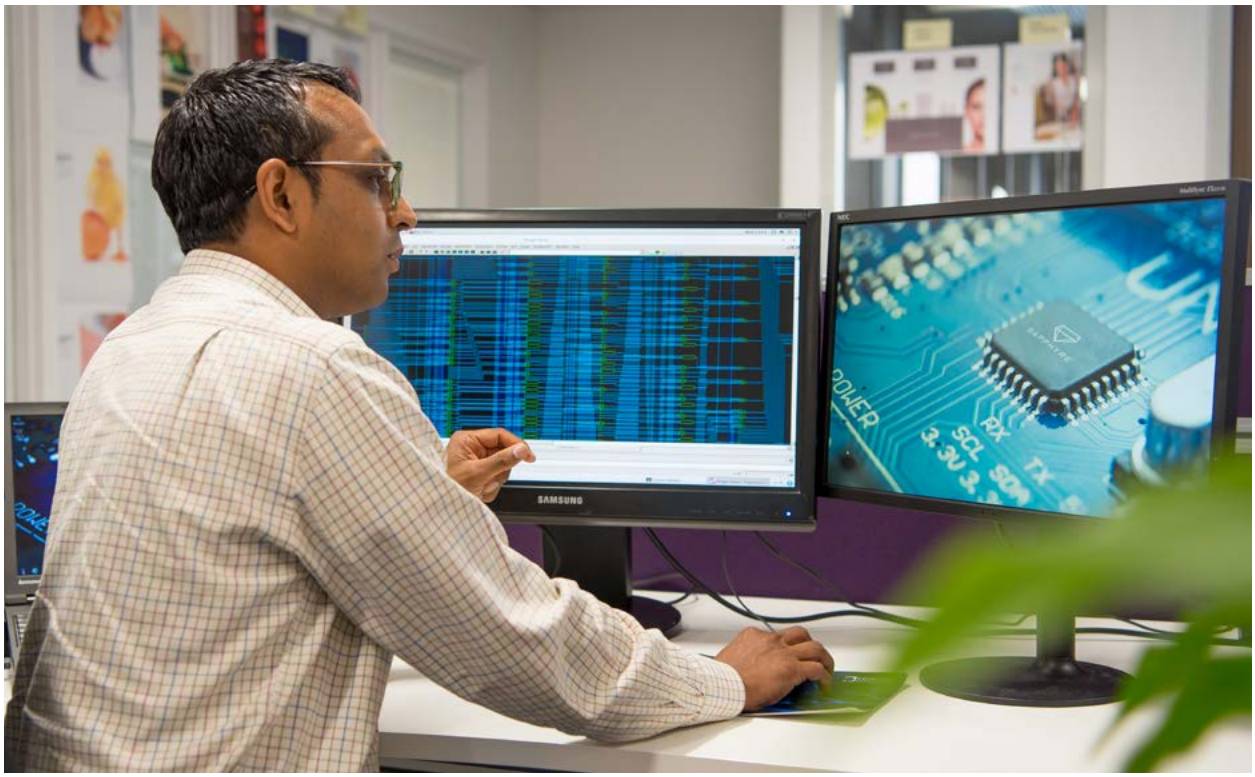
"In addition to the compression and acceleration of existing neural networks, it is also possible to architect a more compact neural network from the very start."

CASE STUDY 2 – ULTRA-LOW POWER AI

Due to the number of embedded ECUs used to manage ADAS sensors, there is significant scope for neural networks to improve the flexibility, capability and performance of these sensing subsystems.

Reducing power consumption of neural networks to enable integration in embedded subsystems has been an active area of research by Cambridge Consultants, leading to the development of the Sapphyre™ neural network demonstration.

The Sapphyre design flow for custom cores and accelerators, has enabled machine learning neural networks to run at a power reduction of up to 1,000x compared to an off-the-shelf processor, without significant loss of accuracy.



In order to investigate the levels of power consumption and accuracy that could be achieved, we focused on an industry standard problem – image classification using the CIFAR-10 dataset. This consists of 32 x 32 pixel colour images, which are to be identified as one of 10 different image categories. We took two small neural network architectures that have previously been used for this classification task, the VGG network and the CIFAR-10 tutorial network and we deployed these models on our Sapphyre platform.

On top of the inherently low-power nature of the Sapphyre architecture itself, using cutting-edge AI algorithm techniques allowed our team to trade-off final accuracy of the network and the energy required to execute the model. By leveraging the XNOR-net algorithm for one or more layers of the neural networks running in Sapphyre, we achieved a sliding scale of accuracy and power consumption.

For the lowest-power network, consuming just 86 µJ of energy per classification, it is possible to run 25 million classifications from a single coin cell battery. This opens the door to applications running for weeks from batteries, or from energy harvesting systems. The same calculation for the 7J of energy of the VGG reference network yielded an unimpressive 308 classifications.

6 CONCLUSIONS

S-ADAS penetration in the premium market will continue to accelerate as more premium OEMs adopt these systems to increase passenger comfort and safety. The systems themselves will continue to become more and more advanced, as premium OEMs push towards Level 4 and 5 autonomy.

While the consensus is that this eventual goal is still five to ten years away, steadily increasing the level of self-driving capabilities offered in passenger vehicles will allow these OEMs to monetize their ADAS technology investments in the shorter term. This in turn will allow designers and engineers to iterate the technology to make it safer, more reliable and more capable.

At the same time, cities, transport infrastructure and road regulations will need to be overhauled to enable more pervasive use of Level 4+ vehicles. Widespread use of this technology has the potential to deliver large-scale societal benefits, reducing pollution and traffic congestion through close co-ordination of vehicles and infrastructure to optimize traffic flow.

In the near future, as consumer awareness and acceptance increase, mid-range OEMs will demand S-ADAS features to differentiate their vehicles. This will drive demand for a more modular and cost-effective alternative to the centralized approach to ADAS architecture employed by the premium OEMs, with the focus on cost-optimizing ADAS sensors, systems, networks and compute components. Technology innovation will be required in all three areas to realize S-ADAS capabilities at a mid-range price point.

As the adoption of AI increases throughout the vehicle, innovations in algorithm optimization and on-chip AI acceleration, which can reduce the resources required for this powerful technology, will be of pivotal importance. Mid-range OEM manufacturers will also need to move away from the incumbent architectures that enable S-ADAS at the premium end and embrace emerging AI acceleration silicon devices to run their resource-optimized algorithms.

By focusing on these key areas to enable an effective hybrid S-ADAS architecture, mid-range OEMs will soon be able to offer the semi-autonomous features currently seen in premium vehicles.

To discuss strategies to leverage AI, sensors and connectivity for the automotive industry, contact:

Thomas Carmody, Head of Transport and Infrastructure
thomas.carmody@cambridgeconsultants.com

REFERENCES

1 https://www.marketsandmarkets.com/PressReleases/driver-assistance-systems.asp

2 https://www.tesla.com/en_GB/autopilot

3 https://www.cadillac.com/world-of-cadillac/innovation/super-cruise

4 http://volvo.custhelp.com/app/answers/detail/a_id/9731/~/pilot-assist

5 https://www.here.com/sites/g/files/odxslz166/files/2018-11/Consumer_Acceptance_of_Autonomous_Vehicles_white_paper_0.pdf

6 http://article.images.consumerreports.org/prod/content/dam/cro/Consumer%20Clarity%20and%20Safety%20for%20Todays%20Advanced%20Driving%20Systems

7 http://article.images.consumerreports.org/prod/content/dam/cro/Consumer%20Clarity%20and%20Safety%20for%20Todays%20Advanced%20Driving%20Systems

8 https://www.mckinsey.com/~/media/mckinsey/industries/automotive%20and%20assembly/our%20insights/how%20carmakers%20can%20compete%20for%20the%20connected%20consumer/competing_for_the_connected_customer.ashx

9 https://arxiv.org/pdf/1709.02435.pdf

10 https://www.continental-corporation.com/resource/blob/118106/deafe75b7e11426dabcc785c0e0316ab/2018-01-09-strategy-key-figures-data.pdf#page=18

11 https://www.electronicdesign.com/automotive/bmw-and-audi-want-separate-vehicle-hardware-software

12 https://www.theverge.com/2017/6/27/15878900/volvo-nvidia-self-driving-car-partnership

13 https://www.nxp.com/products/processors-and-microcontrollers/arm-based-processors-and-mcus/qoriq-layerscape-arm-processors/nxp-bluebox-autonomous-driving-development-platform:BLBX

14 https://www.wired.com/story/self-driving-cars-power-consumption-nvidia-chip/

15 https://www.autonews.com/article/20180827/OEM01/180829843/tesla-prefers-its-self-driving-hardware-over-that-of-other-makers

16 Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., 2015. Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), pp.211-252.

AUTHORS

Thomas Carmody, Head of Transport and Infrastructure

Sally Epstein, Senior Machine Learning Engineer

Jiahui Lu, AI and Strategy Lead, Asia

Sam Sturgess, Consultant, Technology Strategy

James Peet, Senior Engineer, Digital Design

Ivan Petrov, Consultant, Technology Strategy

GLOSSARY

Advanced driver assistance system (ADAS) — *Electronic systems designed to assist the driver and, in some instances, assume control to increase the safety of the driver, passengers and other road users*

Artificial intelligence (AI) — *The simulation of natural (human) intelligence by machines, including performing tasks such as visual perception, prediction and decision making*

Controller area network (CAN) — *A serial bus network allowing the controllers, sensors and actuators in a sub-system to communicate with each other to enable real-time control applications*

Centralized architecture / compute — *A popular ADAS compute architecture where all, or almost all, the computing is done at a central unit, which is typically owned by an OEM*

Compute architecture — *A specification/description of how a set of software and hardware technologies interact to form a computer system*

Central processing unit (CPU) — *The primary processor unit of a computer that processes instructions and coordinates the other units that make up the compute architecture*

Deep learning — *Also known as deep neural networks, this is a subset of machine learning where algorithms with multiple (deep) artificial neural network layers learn to perform tasks using large amounts of data*

Distributed architecture / compute — *An alternative ADAS compute architecture where some data processing is done at certain sensor nodes and object level information is passed to a central compute for decision making*

Dhrystone million instructions per second (DMIPS) — *A measure of computer performance*

Electronic control unit (ECU) — *An embedded system in automotive electronics that controls one or more electrical systems or subsystems in a vehicle*

Floating point operations per second (FLOPS) — *A measure of computer performance*

Graphical processing unit (GPU) — *Used for parallel processing*

Global navigation satellite system (GNSS) — *Includes GPS, GLONASS, Galileo, Beidou*

Informational ADAS — *Used for Level 0 vehicles. These systems provide information to the driver and alert them to potential hazards, but do not assume control of the vehicle under any circumstances*

Inertial navigation system (INU) — *A navigation system that uses a computer and sensors, like accelerometers and gyroscopes, to continuously calculate the position, orientation and velocity of an object (vehicle), without the need for external references*

Level 0-5 — *A term used to refer to autonomous vehicles with different levels of autonomy. Level 0 refers to an unconnected vehicle. Level 5 refers to a fully autonomous vehicle*

Light detection and ranging (LiDAR) — *A detection system that measures distance by illuminating a surface with a pulsed laser beam and measuring the reflected pulses*

Machine learning — *An application of AI that provides systems with the ability to automatically learn and improve from experience without being explicitly programmed*

Machine vision — *An application where hardware and software combine to provide operational guidance to devices (e.g. vehicles) in the execution of their functions based on the capture and processing of images*

Microcontroller unit (MCU) — *An integrated circuit designed to control a specific task in an embedded system. A typical MCU includes one or more CPUs, memory and programmable input/output peripherals on a single chip*

Neural network — *A set of algorithms which aims to recognize underlying relationships in a dataset through a process that mimics the way the human brain operates*

Object level data — *A high level interpretation of the raw sensor data – e.g. the raw sensor data from a camera would be pixel intensities, whereas the object level data could be "person", "parked car", etc.*

Original equipment manufacturer (OEM) — *For automotive, this refers to car manufacturers like Tesla, Volvo, Ford, etc*

Radar — *A detection system that sends out an intermittent beam of radio waves and measures the reflected echos to determine the range, angle and velocity of objects*

Semi-autonomous ADAS (S-ADAS) — *ADAS used in semi-autonomous vehicles, e.g. Level 2-3*

Sensor node — *A node in a sensor network capable of performing some processing, gathering sensory information and communicating with other connected nodes in the network. This is usually a microprocessor*

Single input multiple data (SIMD) — *A class of parallel computers with multiple processing elements that perform the same operation on multiple data points simultaneously*

System-on-chip (SoC) — *A microchip or integrated circuit that combines all the required electronic circuits and components of an electronic system onto a single chip*

Tier 1-3 supplier — *Suppliers within the automotive supply chain. Tier 1 are module or system level suppliers. Tier 2 are component suppliers. Tier 3 are parts and raw or close-to-raw material suppliers*

Trillions operations per second (TOPS) — *A measure of computer performance*

About Cambridge Consultants

Cambridge Consultants is a world-class supplier of innovative product development engineering and technology consulting. We work with companies globally to help them manage the business

impact of the changing technology landscape. With a team of more than 850 staff in the UK, the USA, Singapore and Japan, we have all the in-house skills needed to help you – from creating innovative concepts right the way through to taking your product into manufacturing. Most of our projects deliver prototype hardware or software and trials production batches. Equally, our technology consultants can help you to maximize your product portfolio and technology roadmap.

We're not content just to create 'me-too' products that make incremental change; we specialize in helping companies achieve the seemingly impossible. We work with some of the world's largest blue-chip companies as well as with some of the smallest, innovative start-ups that want to change the status quo fast.

Cambridge Consultants is part of the Altran Group, a global leader in innovation. www.Altran.com

To discuss strategies to leverage AI, sensors and connectivity for the automotive industry, contact:

Thomas Carmody, Head of Transport and Infrastructure (Cambridge, UK)
thomas.carmody@cambridgeconsultants.com

Jiahui Lu, AI and Strategy Lead, Asia (Singapore)
jiahui.lu@cambridgeconsultants.com

Eric Olasen Director, Industrial & Mobility Business (Seattle)
eric.olason@cambridgeconsultants.com

Oli Qirko, SVP, US Division Head, Industrial Consumer and Energy (Boston)
oli.qirko@cambridgeconsultants.com

Cambridge Consultants

UK ▪ USA ▪ SINGAPORE ▪ JAPAN

www.CambridgeConsultants.com