Counterpoint
Technology Market Research

Sponsored by: Western Digital

WHITEPAPER

Storage Capacity Requirement for Autonomous Vehicles to Balloon Over 2TB in the Next Decade

Table of Contents

EXECUTIVE SUMMARY

Automotive innovation across the last decade or two has largely been about hardware platform evolution. This can be seen in the improved efficiency of internal combustion engines. However, during the last few years, the automotive industry is starting to undergo a significant transformation. We are now seeing glimpses of what cars will become in the coming decades. Vehicles of the future will be electric or hybrid, be connected, and increasingly be capable of making decisions. This will lead, eventually, to the emergence of fully autonomous cars.

To become connected and autonomous, vehicles need to incorporate advanced driving assistance systems (ADAS), that ultimately will be capable of taking control of the vehicle from the human driver. These vehicles will be designed to continuously accumulate, process, and share data. This data will be used to inform everything from training artificial intelligence systems to developing new business models. Counterpoint's research estimates that over the next decade, the storage capacities needed will be ranging from a minimum of 2TB to as much as11TB to support the needs of in-vehicle storage at different autonomy levels.

However, there is a challenge. Both the volume of data accumulated, and the potential costs associated with transmitting it are unlike anything that has been seen in the automotive sector before. Vehicles will be increasingly designed to gather, process, and store data locally as well as selectively upload data at an appropriate time, creating a need for data storage and computing infrastructures within the car, as well as in the cloud.

While automotive OEMs will develop in-vehicle sensor systems and powerful data processing capabilities, investment in automotive-grade, high-performance, high-capacity onboard storage will also be critical.

To make real-time, split-second decisions, the 'brain' of an autonomous vehicle will demand several types of storage to accommodate the data from multiple in-car sensors. Efficient and robust processing capability is key to ensuring the industry successfully reaches greater levels of autonomy. Therefore, the data stored locally also needs to be extracted and processed quickly. Counterpoint Research expects that in-vehicle storage systems will move from SLC/MLC NAND to UFS/embedded SSD for Level 3 to Level 5 autonomous vehicles to enable the speed and performance required for data management.

In this paper, we analyze the roadmap for the growth of autonomous vehicles, the amount of data generated by the growing sensor system at each level of autonomy, and the corresponding rise in the amount of in-vehicle storage needed. It aims to help automakers understand the importance of the data generated and what can be stored in-vehicle versus what can be sent to the cloud for further analysis.

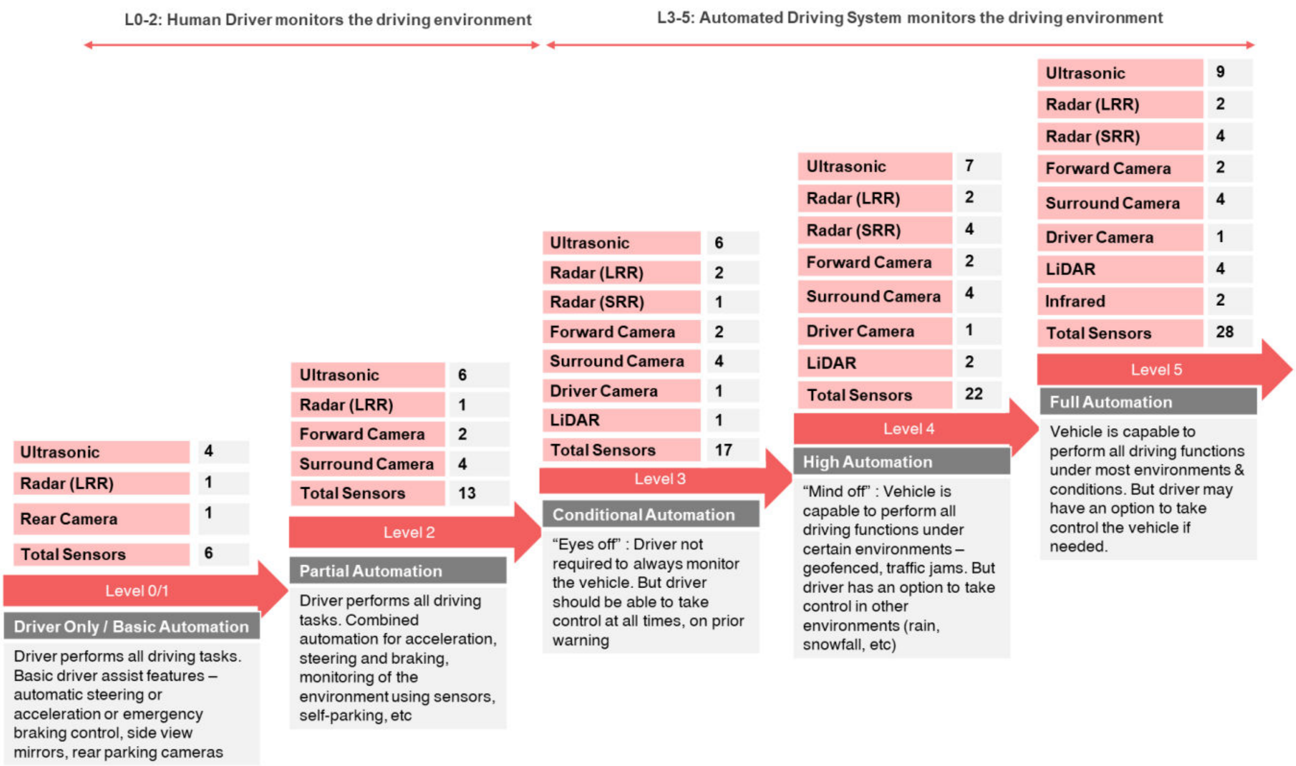
"Vehicles will be increasingly designed to gather, process, and store data locally as well as selectively upload data at an appropriate time, creating a need for data storage and computing infrastructures within the car, as well as in the cloud"

THE EVOLUTION OF AUTONOMOUS VEHICLES

Sensor Network: Building Blocks of an Autonomous Vehicle

The foundation of an autonomous vehicle is built on an Advanced Driver Assistance System (ADAS). This is a collection of sensors, software, and artificial intelligence (AI) capabilities. ADAS captures and processes data from the surrounding environment, drivetrain, motion sensors, and road conditions to support a human driver. Eventually, ADAS will be sufficiently capable that it will be able to take over all driving tasks from the human driver. The number and type of sensors mandated at each autonomy level will increase progressively to meet the driving requirements of the system. The exhibit below highlights the increasing number of advanced sensors required for the goal of achieving highly or fully automated driving. The key combination includes Ultrasonic, Radar, LIDAR, Image and other sensors to emulate human senses. In addition to the vehicle becoming a real-time sensor network, advanced computing, and processing capabilities at both silicon and machine-learned algorithmic level will also be key to build a robust ADAS system and self-driving capabilities.

Exhibit 1: Sensor Installation Potential – by Autonomy Levels



Market Penetration Roadmap for Autonomous Vehicles

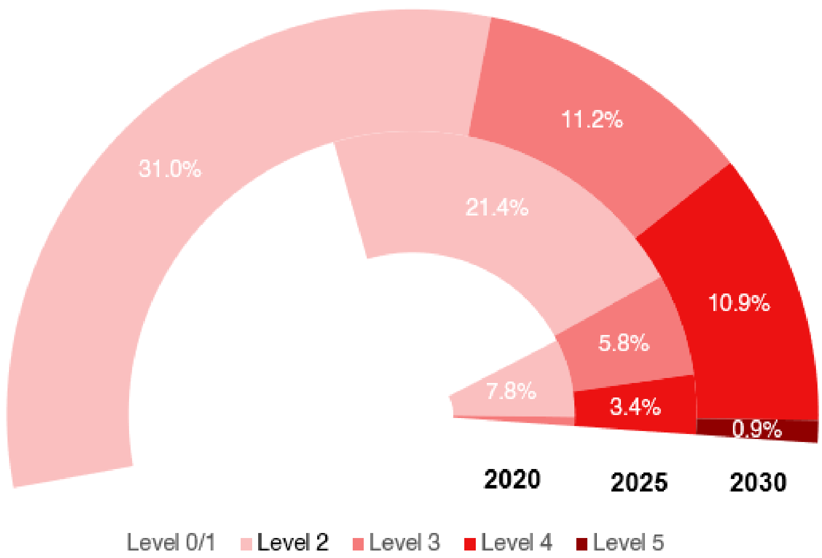
Automakers started to integrate Level 2 capabilities in 2015. Level 2 is defined by a combination of several hardware sensors, including multiple cameras, ultrasonics, and radars, together with software capable of supporting ADAS features such as adaptive cruise control, lane centering and autosteering. Some automakers offer these features as part of their standard package as a point of differentiation, others due to regulatory pressure, such as Euro NCAP[1] and NHTSA[2]. For example, Ford is planning to roll out Co-Pilot360 as a standard feature for most of its new vehicles in North America.

> **"By 2025 we expect around 30% of cars sold globally will support Level 2 or above. This proportion will exceed 50% by 2030, with fully autonomous vehicles exceeding 10% by 2030"**

We expect around 8% of all vehicles sold globally, including commercial vehicles, will feature Level 2 or above capability in 2020. By 2025 we expect around 30% of cars sold globally will support Level 2 or above. This proportion will exceed 50% by 2030, with fully autonomous vehicles – those capable of almost all driving functions without any human intervention – exceeding 10% by 2030.

Businesses that intend to use robo-taxis will likely drive the development of Level 5 capability. Robo-taxis will be deployed on predictable routes, such as from airports to city centers. The attraction of operational cost savings and uninterrupted working will spur companies to invest in their development. Further, Level 5 vehicles for industrial, logistics, agricultural and similar commercial use-cases, driven in controlled environments, will also be on a rapid rise over the second half of the next decade.

Exhibit 2: % Share of Autonomous Vehicles Sales by Autonomy Level



[1] Euro NCAP - The European New Car Assessment Programme

[2] NHTSA - The National Highway Traffic Safety Administration (under Department of Transportation, the US)

THE ROLE OF DATA IN AUTONOMOUS VEHICLES

Autonomous vehicles will use data to accomplish their driving tasks. Meanwhile, they will generate huge amounts of data. The data accumulated from the network of driving cars – both autonomously and guided by human drivers – will be used as the base training dataset to help all other cars improve in their decision-making ability.

In addition to both using and generating data, the ability to grow autonomy to free drivers from the task of driving will potentially enable them and other passengers to engage in more data-intensive infotainment, productivity, leisure, and communication applications. Several other data-driven services will blossom depending upon the type of autonomous vehicle-passenger, commercial, or robo-taxi.

Exhibit 3: Key Applications in a Vehicle Using Data

1 Autonomous Driving

- Autonomous driving systems to enable features like object detection, mapping and path planning
- The majority of applications to run on-board at the edge, rather than cloud
- On-board computing to decide if the data is stored locally, pushed to the cloud (for V2X and other applications), or discarded
- V2X (DSRC[1], C-V2X)
- HD Map
- ADAS
  - ACC (adaptive cruise control)
  - AEB (autonomous emergency braking)
  - ALCA (Active Lane Change Assist)
  - DMS (driver monitoring system)
  - FCW (forward collision warning)
  - LKA (lane keeping assist)
  - PDS (passenger detection system)
  - RSR (road sign recognition)

1. DSRC - Dedicated Short Range Communications
2. C-V2X - Cellular vehicle-to-everything

2 In-Vehicle Infotainment (IVI)

- Navigation (Maps, Point of interest)
- Biometric authentication
  - Voice recognition/command
  - Gesture control
- Information
  - HUD (heads-up display, AR)
  - Audio/video streaming & storage
  - Location-based recommendations
- Mobile Office
  - Enterprise video calling
  - Content creation apps
- Real-time gaming

3 Additional Platform & Services

- Robo taxi platform (or subscription management)
- OTA (Operating system & apps)
- Maintenance & fleet management
- Insurance services
- Blackbox
- Other services

Connected and autonomous vehicles will generate data primarily via ADAS sensors such as ultrasonic, cameras, LiDAR, and radars. This will create a need for local data storage in the vehicle. Moreover, different entities within the vehicle ecosystem to the smart city will likely use the data generated by all the vehicles, whether they are moving or idle. Data aggregated from vehicles and the roads will help alert other users of dangerous road conditions, or for the automaker to update training datasets, or for fleet managers to monitor usage and performance, or dealers to schedule maintenance, or any of myriad and, as yet, undreamt of use cases.

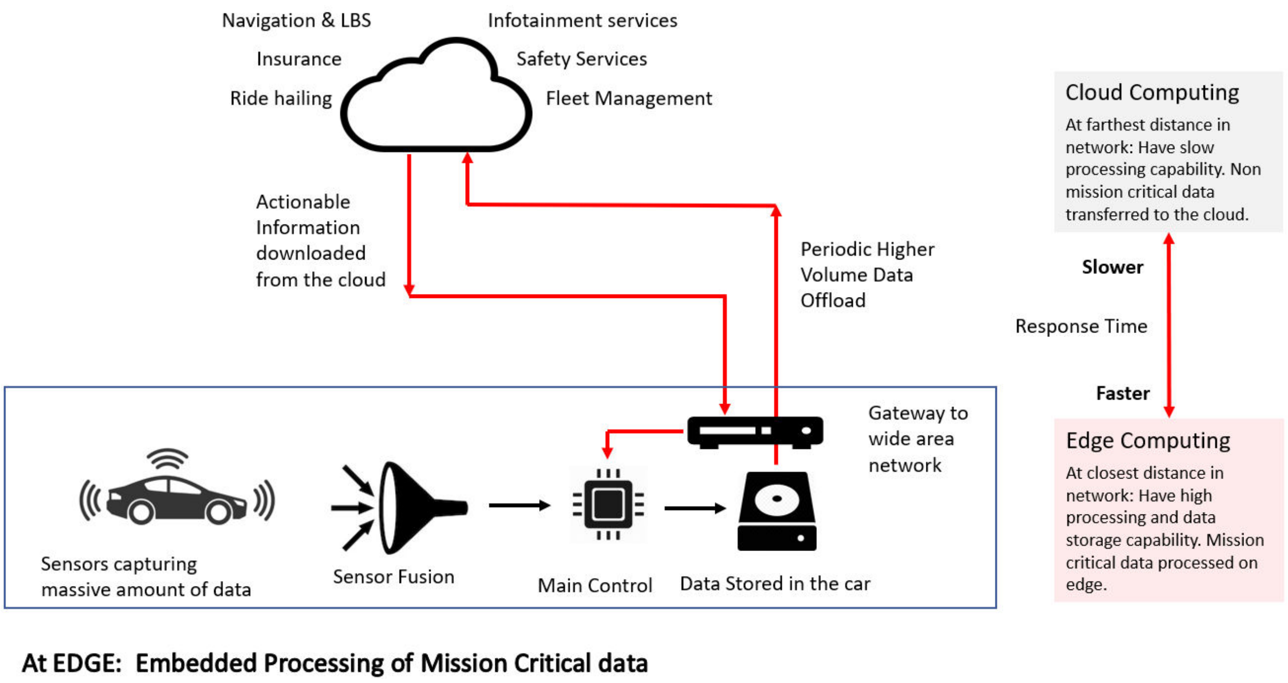
Exhibit 4 shows the journey of the data captured by these sensors. The on-board AI accelerator first processes the data. Once processed, the curated data is stored in the vehicle's embedded storage. Any dynamic information, such as local traffic or road conditions, or weather anomalies, are uploaded either to local edge servers or to the central cloud, depending on whether the information is relevant only to other users in the immediate vicinity, or to the wider ecosystem.

Vehicles will need more refined training data as they evolve from Level 2 to higher levels. Training data will come from both the accumulated miles driven by customers' vehicles as well as extensive OEM test fleets combined with simulated driving environments. The data includes number of miles driven by all contributing vehicles, the power of the autonomous driving computers, and the design of the algorithms, which are critical to the improvement of autonomous driving systems.

Separately, car manufacturers modify their software and services through all the data they collect and analyze from their own customers and test fleet and simulators. Once thoroughly tested, updates are made available to owners' cars via over-the-air (OTA) directly by car manufacturers, who continue to rely on training data to improve performance.

Any aberrant behaviors, such as an incident demanding human drivers to immediately take over, will be prioritized. Such anomalies will form special cases alerting urgent upload to central servers where the data will be simulated. Correct responses to the situation will also be developed and incorporated into the training data for future reference.

Exhibit 4: Journey of Data Captured by Vehicles



How Much Data Will Autonomous Vehicles Generate?

Sensors will be the primary data source in autonomous vehicles. The number and type of sensors will vary from automaker to automaker based on many factors, including preference, cost, and use cases. Data generated per sensor type also varies depending on the application. For example, data generated by a front camera, depending on the camera resolution, color depth, frame rate, and compression level, can typically range between 70-300 GB/hr per vehicle.

The amount of data generated will also vary by the type and autonomy levels of a vehicle. Robo-taxis and OEM test vehicles are likely to install greater numbers of higher-resolution cameras compared to passenger vehicles, as they are less sensitive to hardware cost. A Level 4 OEM test vehicle will likely generate around 80% more data compared to a passenger car of the same autonomy level.

With more than 20 different types of sensors that can be integrated into a Level 4 autonomous vehicles' ADAS, the vehicle can generate data in the range of 1-2TB per hour, depending on whether it's a passenger vehicle, commercial vehicle, or robo-taxi. Exhibit 5 outlines a range of being data generated currently by the key sensors used in an autonomous vehicle. Camera Image and LIDAR sensor data will be the main contributors to the amount of in-vehicle data generated.

Exhibit 5: Data Generated (GB per hr) by ADAS Sensor Type

Sensor Type*	Key Applications	Comments	Data Generated per Vehicle(GB/hr)
Radar (SRR)	BSD, FCW	+ Robust in tough environments (rain/snow) - Concerns regarding people detection and angular resolution	0.03 - 0.4
Radar (LRR)	ACC, FCW		
Forward Camera	LDW, FCW, TSR	+ Overall a versatile sensor for many applications - Limited depth perception - Performance affected by environment (rain/fog) - Needs proper illumination	70-300*
Surround Camera	Blind Spot Monitoring 360-degree Video APS		
LiDAR	Environment mapping, BSD, FCW	+ Better accuracy and precision compared to other sensors - Performance concerns in tough environments (rain/snow) - Higher price compared to other sensors	36-252

* In the case of L3 to L5

ACC – Adaptive Cruise Control
BSD – Blind Spot Detection
FCW – Forward Collision Warning
LDW – Lane Departure Warning
LRR – Long Range Radar
SRR – Short Range Radar
TSR – Traffic Sign Recognition
APS – Automated Parking system

"Camera Image and LIDAR sensor data will be the main contributors to the amount of in-vehicle data generated"

Data Storage in Autonomous Vehicles

The autonomous vehicles' sensor fusion data stored post-processing coupled with the platform, infotainment, OS, and applications, will account for the majority of data storage requirements.

Sensor Data

- Sensor data will account for the largest share of vehicle onboard data storage in the future.
- Sensor data mostly comes from ADAS systems and vehicle-to-everything communication (V2X).

Autonomous Vehicle Platform, Infotainment, OS & Applications

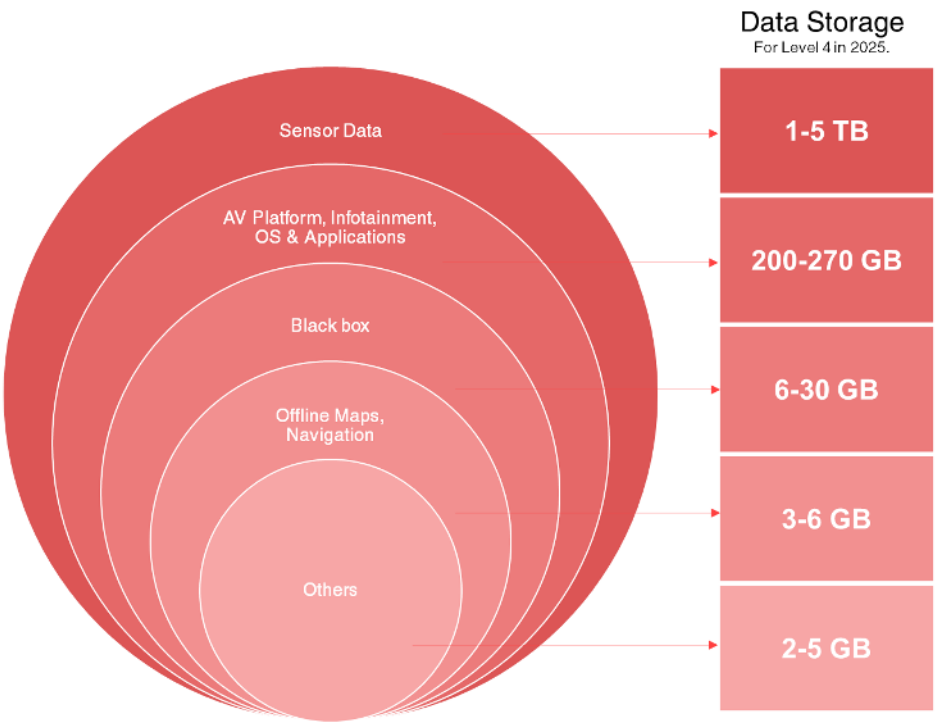
- The autonomous vehicle platform required to run various autonomous/ADAS features will drive additional storage needs for an advanced operating system, media, gaming, productivity, voice AI applications, and other offline content.
- A 'black box' is expected to become a legal and safety requirement by regulators. Data generated from connected and safety services, OTA updates, and similar services will require some local storage space.

HD Maps

- Unlike current 2D maps used for navigation on cars, high-precision (HD) maps vary greatly in refresh rates, sampling methods, and cost. Due to the high frequency of dynamic data updates, HD maps typically use real-time online updates via wireless transmission.
- The use of HD maps in autonomous vehicles can complement sensor deficiencies, increase positioning accuracy, and reduce sensor costs. Traditional 2D Map refers typically to the lowest level of static map layers. The HD map includes a static layer, semi-static layer, semi-dynamic layer, and a dynamic layer. The static map layer can be updated once a month or as needed. However, dynamic maps are updated every hour, minute, and second.

Exhibit 6 shows the different categories and corresponding range of data storage requirements in a future autonomous vehicle – from robo-taxi to enterprise fleet to passenger vehicle.

Exhibit 6: Data Storage Requirements in Future Autonomous Vehicles



Data Storage Strategies

A combination of cloud and edge computing strategies will eventually prevail in autonomous vehicles, necessitating the development of AI computing systems and high-speed in-car communication enabled by high-performance storage. Autonomous systems' requirements for high-speed data transfer, retrieval, and processing will demand a hybrid of cloud and local storage.

Today, the amount of data generated by vehicles is relatively modest because the number of sensors is limited at Level 2 autonomy. Until OEMs are confident of the machine learning models to enable Level 4-5, most will bank on storing more data rather than less, and find ways to upload data to the OEM's cloud to strengthen machine learning models. As OEMs move to Level 4-5, with more robust edge intelligence, the evolution of sophisticated data networks, and advanced data compression techniques, there will be a point where the amount of data stored locally will level-off or possibly decline.

"Until OEMs are confident of the machine learning models to enable Level 4-5, most will bank on storing more data rather than less, and find ways to upload data to the OEM's cloud to strengthen machine learning models"

Autonomous vehicles cannot rely entirely on cellular networks for communication with the cloud. But they can't completely do away with cellular connectivity either. While edge computing is critical in driving, various signals can help the autonomous system make decisions faster. These signals may come from communication between other vehicles (V2V) and with the infrastructure (V2I). Companies such as Mobileye and HERE will also provide rapid real-time 3D geographic, traffic, and road condition information to alert autonomous systems of hazards. Therefore, high-speed real-time, low-latency transmission is critical for success of Level 4-5 ADAS systems. Coming 5G networks can reduce latency to less than ten milliseconds, provided information is stored close to the network edge. However, 5G will take several years to roll out in all markets with broad coverage. Until then, systems will be based on 4G LTE networks and will have to rely on locally generated, stored and processed data.

The automotive industry will face four challenges when it comes to handling data generated by autonomous vehicles:

- Size of data sets
- High-speed, real-time data processing
- Security and privacy
- Cost of network bandwidth for transmitting data to the cloud

Network infrastructure and cost of bandwidth will play a significant role in establishing a business model that addresses the most important question – who will pay for the cost of communication? Two underlying factors will define the data offloading channel and financial model – **time criticality** and **who is deriving the value from the data**. While time criticality will define what network to use (cellular, Wi-Fi, ethernet) for data transfer, who derives value from the data will define who will pay for the cost of transmission.

While OEMs will likely extract the most value from the data being generated, the cost associated with data transmission will become a deciding factor in how and what is transmitted. For example, the cost to an OEM to transfer all the data gathered by a single Level 4-5 autonomous vehicle, for a day, over a cellular network, could run to thousands of dollars. This is unlikely to be a viable solution across an entire fleet of cars.

There are challenges at each stage of the data journey which warrant prudent data collection, storage, and usage strategies. An autonomous vehicle will be subject to environment, design, and cost limitations, whereas the network and cloud will have limitations in bandwidth, connectivity, latency, and security. Autonomous vehicles will, therefore, need to have advanced storage and intelligent offload systems. The periodic offload can be done while cars are at charging stations over an ethernet connection, over Wi-Fi while parked at home, or during routine maintenance trips at service centers. The data itself may be pared down to simpler metadata rather than raw unprocessed data streams depending on the need for the type of data to train the models or enrich the database.

To overcome these issues, OEMs and suppliers need to collaborate, share, and learn to achieve scalability and innovate. The local data storage strategy required for the four different types of vehicles included in this study – passenger vehicles, commercial vehicles, robo-taxis, and OEM test vehicles – will also vary significantly. For example, OEM test vehicles are expected to have a relatively large proportion of data stored locally, compared to passenger vehicles. Test vehicles and commercial vehicles will also tend to regularly visit a base where physical storage systems can be manually removed and replaced.

Exhibit 7: Data Storage Requirements by Vehicle Type by Autonomy Level

Vehicle Type / Autonomy Level / Application Storage Requirement (GBs)	2020E	2025E	2030E
Passenger Car 			
Level 3	608.6	993.6	1,431.9
Level 4	-	1,176.9	1,037.1
Level 5	-	-	942.2
Commercial Vehicle 			
Level 3	-	2,056.6	3,153.8
Level 4	-	3,275.6	2,509.1
Level 5	-	-	2,258.8
OEM Test Vehicle 			
Level 4	3,646.0	5,497.5	6,977.7
Level 5	-	7,044.6	10,081.1
Robo Taxi 			
Level 4	-	6,981.1	8,148.3
Level 5	-	-	10,862.6

THE OPPORTUNITY FOR STORAGE PLAYERS IN AUTONOMOUS VEHICLES

Today, the most common storage solution for in-vehicle infotainment (IVI) systems are e.MMCs[3] . However, they cannot support autonomous driving capabilities that need sophisticated graphical user interfaces and faster speed to access models from memory. In addition, there must be adequate storage space in the vehicle's infotainment system to store multimedia and high-resolution maps.

In the evolution of fully autonomous driving, signals from sensors are processed at high speed by the drive computer, requiring high-bandwidth and processing speed. Besides, drive computers will likely include redundant systems running in parallel and comparing decision outcomes in real-time, potentially multiplying memory requirements.

Automated driving systems require proper storage to manage the large amounts of data at high speeds. Therefore, as the level of autonomy increases from Level 1 to Level 4 and 5, vehicles will need an exponential increase in storage capacity, bandwidth, and speed. To cater to such needs, storage technology is correspondingly increasing from SLC NAND[4] to e.MMC/UFS[5] to embedded SSD[6].

"As the level of autonomy increases from Level 1 to Level 4 and 5, vehicles will need an exponential increase in storage capacity, bandwidth, and speed"

SLC NAND

- Applications: Event log in Event Data Recorder (EDR), code storage in embedded system, data storage in dashboard
- The capacity is small and needs to be managed by the system
- Widely used from Level 1 to Level 5 with a maximum capacity of 4GB

e.MMC/UFS

- Applications: Infotainment systems, navigation system and code storage in ADAS
- MLC eMMC/UFS has been so far widely used in the ADAS systems since it provides a good compromise between performance and cost, data security, durability, price, and capacity
- Further, a move to 3D NAND (TLC) eMMC/UFS has further helped from a cost perspective
- It will be used in Level 2 to Level 4

UFS/embedded SSD

- Applications: Storage of high-resolution maps, autonomous vehicle computer, AI database, black box data recorder
- Price of SSD is higher than other storage systems but can provide better speed, larger capacities, and higher bandwidth
- It will be used in Level 3 to Level 5

Automotive Specific Storage Solutions

The requirements of automotive storage are different from that of consumer electronics. Automotive applications require high-reliability and greater resilience to data corruption than consumer applications. While consumer electronics are mostly common parts and off-the-shelf solutions, automotive products are specifically designed

[3] eMMC - Embedded Multimedia Card
[4] SLC NAND – Single-Level Cell NAND
[5] UFS - Universal Flash Storage
[6] SSD - Solid-State Drive

for the automotive environment. Since memory chips are soldered and not socketed, recall and repairs that swap the entire board will be expensive operations for OEMs.

Safety, reliability, and quality are critical for automotive memory storage units. Automotive memory must, at a minimum, conform to the following three standards – ISO26262, AEC-Q100, and IATF16949.

- ISO26262 leverages a common standard to measure the functional safety of a system in service. It defines regulations and requirements in the product development process, from conceptual development through to component retirement.

- AEC-Q100 is used to evaluate the reliability and stability of products. Automotive memory must meet at least Grade 2 and 3 tests to verify product stability at high temperatures.
    - Grade 2: -40°C to +105°C ambient operating temperature range
    - Grade 3: -40°C to +85°C ambient operating temperature range

- IATF16949 certification establishes conformity to the zero-defect quality management standard of the supply chain. At every stage of the memory supply chain, including memory manufacturers, controller suppliers, assembly, and testing vendors, all must be certified for IATF16949 to make sure the parts are traceable back to the original wafer batches and wafer lot.

The three major areas which comprise a memory component include:

- NAND memory type (SLC, MLC, TLC) and geometry node or layers
- Controller and firmware
- Packaging and testing

A key part of an automotive-grade memory is its controller. Poorly designed controller architecture and/or firmware algorithms can wipe out the advantages of the most reliable NAND memory.

Although most leading NAND manufacturers have passed the three quality standards, those manufacturers with in-house proprietary designed, high-quality controllers, have significant competitive advantages. While most third-party memory vendors can provide good quality automotive devices, they are challenged to assure long-term stable supply and consistent quality. Thus, there are advantages for vendors that have their own production line and in-house research and development.

KEY TAKEAWAYS

Data storage solutions will be critical enablers in the evolution of autonomous vehicles, the next big leap for the automotive industry.

Numerous sensors, cameras, radars, and LiDARs, the backbone of an autonomous vehicle, will generate massive amounts of data, most of which will have to be stored and processed locally.

The volume of data generated will vary by the application type of the vehicle. For example, commercial vehicles and robo-taxis will capture more data and require more storage compared to passenger vehicles, due to longer runtimes per day and being the eyes and ears on the road for the OEMs and fleet owners. Counterpoint Research expects the requirement for in-vehicle storage to range from 1.2TB in Level 3-Level 4 passenger vehicles to 7TB in Level 4 robo-taxis by 2025. Despite such variations, we believe that OEMs will need to plan for at least 2TB or more of onboard storage for Level 4 and Level 5 autonomous vehicles.

Simultaneously, automakers need a plan to create a data off-loading infrastructure to facilitate the easy transmission of stored data to the cloud to build robust machine learning models and new data-driven services.

However, a large data storage capacity is just one aspect of data storage solutions for autonomous vehicles. Level 3 and above autonomous vehicles will use a combination of cloud and edge computing strategies necessary to optimize the sophisticated requirements of AI computing systems, high-speed responses, and data transmission costs. Storage players are evolving to meet such requirements. We believe that in-vehicle storage systems will move from SLC/MLC NAND to UFS/embedded SSD for Level 3 to Level 5 autonomous vehicles.

While data storage players will become key partners of the automotive players in the coming years, they also need to prepare for stable and reliable long-term supplies. Unlike consumer electronics, supplying to automakers will require meeting relevant quality standards to prove reliability at high temperatures and other harsh environmental conditions, as well as ensuring the functional safety of the system.

AUTHORS, COPYRIGHT, USER AGREEMENT AND OTHER GENERAL INFORMATION

Aman Madhok
Senior Analyst, Smart Automotive & Mobility
E: aman@counterpointresearch.com
M: +91 95603 84548

Brady Wang
Associate Director, Components & Technologies
E: brady.wang@counterpointresearch.com
M: +886 9365 91212

Vinay Piparsania
Research Director, Smart Automotive & Mobility
E: vinay@counterpointresearch.com
M: +91 99710 05882

COUNTERPOINT TECHNOLOGY MARKET RESEARCH
Hong Kong | USA | South Korea | India | UK | Argentina | China

20F Central Tower, 28 Queen's Road Central, Hong Kong
info@counterpointresearch.com

opinions, conclusions or information contained in the report. All trademarks displayed in this report are owned by