Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 2, 2024

Applications and Challenges of AI Chips in Autonomous Driving Technology

Tianhao Huang

Faculty of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China
hwangtinhau@163.com

Abstract: The rapid development of autonomous driving technology puts great demands on high-performance computing, and AI chips play a crucial role in it due to their efficient parallel computing capability. The aim of this review is to explore the application of AI chips in different layers of the autonomous driving system, including the perception layer, the decision-making layer, and the execution layer, and to analyse their key technical issues in terms of power consumption management, computational efficiency, scalability, reliability and security. In addition, this paper discusses the impact of the manufacturing process on the performance of AI chips, as well as its future development trends and challenges.

Keywords: AI chips; Autonomous driving; Perception layer; Decision layer; Execution layer; Power management; Computational efficiency; Manufacturing process.

1. Introduction

Autonomous driving technology is considered to be a revolutionary innovation for future transport systems, with the advantages of enhancing traffic safety, improving transport efficiency and reducing energy consumption. With the development of technology, the application of AI chips in autonomous driving systems has become more and more widespread, and has become the core of realising advanced autonomous driving functions. This review will systematically analyse the current status of the application of AI chips in autonomous driving technology, design challenges and its future development direction.

2. Overview of autonomous driving technologies

2.1. Definition and evolution of autonomous driving

Definition: Autonomous driving technology refers to the realisation of autonomous driving operation of vehicles without human intervention through the vehicle's own sensors, controllers, actuators and other devices. Its purpose is to realise autonomous driving of vehicles through technical means such as environment perception, planning decision-making, automatic control, etc., in order to improve driving safety, efficiency and comfort. Autonomous driving technologies are usually categorized into different levels of automation, ranging from partially automated driving to fully automated driving. Using the classification method of the Society of Automotive Engineers (SAE), automated driving technologies are described by dividing them into 6 levels (L0-L5, shown in Table 1).

Development history: The development of automatic driving technology can be traced back to decades ago, but in recent years has experienced rapid development, from the initial assisted driving system to fully automated driving, automatic driving technology has made great breakthroughs. It is roughly divided into three stages: assisted driving system, advanced driver assistance system and automatic driving. [2]

Assisted driving systems: These mainly include blind spot monitoring, lane keeping assist, adaptive cruise control, automatic emergency braking and other functions. The driver remains the primary controller and needs to monitor and operate the vehicle at all times. Basic camera, radar and ultrasonic sensors are used. The aim is to improve driving safety and reduce some of the burden on the driver. Corresponds to SAE levels L1 and L2 above. reduces the number of traffic accidents, especially collisions at low speeds and stopping, gradually allows users to accept and trust autonomous driving technology, and lays the groundwork for future advanced stages.

Advanced Driver Assistance: Based on the Assisted Driving System, more automated functions are added, such as Traffic Jam Assist, Motorway Driving Assist and Automatic Parking System. The driver still needs to be alert, but in some cases the vehicle is able to perform more driving tasks autonomously. The addition of laser radar (LiDAR), more advanced camera and radar systems, as well as greater computing power and software algorithms, further reduces the operational burden on the driver and increases driving comfort and safety. L3, which corresponds to the SAE level, reflects a significant increase in sensor technology and data processing capabilities, enabling vehicles to sense and process more environmental information. Laws, regulations and standards for advanced driver assistance systems are beginning to emerge, providing legal protection for the promotion and popularisation of the technology. Significantly improving the user experience, more driving tasks can be completed by the system under specific conditions.

Fully Automatic Driving: The vehicle is able to complete the driving task from start to finish completely autonomously without the intervention of a human driver. The driver can become a complete passenger and no longer needs to be involved in the driving process. Comprehensively adopting a variety of sensors, including high-precision LIDAR, millimetre-wave radar, high-definition cameras and high-precision maps, combined with powerful AI algorithms and computational capabilities, the vehicle can achieve true driverlessness, improve traffic efficiency, reduce traffic accidents and provide a convenient way of travelling. L4 and

L5, which correspond to SAE levels, are the embodiment of major breakthroughs in algorithms and computing capabilities, especially the maturity of AI and machine learning technologies, and have far-reaching impacts on traffic management, urban planning, and social and legal systems, including employment, legal responsibilities, and ethical issues.

Table 1 Specific information

Hierarchy	Specific information
L0	L0, i.e. a vehicle driven entirely by a human driver
L1	Also known as assisted driving, with the addition of early warning and alert ADAS functions, including lane departure warning, forward collision warning, blind spot detection warning lights, etc., which are mainly early warning alerts and have no active intervention function.
L2	It is known as semi-autonomous or partially autonomous driving, with features such as adaptive cruise, emergency automatic braking and lane keeping assistance.
L3	With integrated intervention functions, including automatic acceleration, automatic braking, automatic steering, etc., vehicle systems at L3 level are able to rely on their own sensors to sense the driving environment, but monitoring tasks still need to be led by the human driver, and human driver intervention is still required in emergency situations.
L4	The term "high degree autonomous driving" means that in limited areas or limited environments (e.g. fixed parks, closed and semi-closed highways, etc.), the vehicle can fully sense the environment and intervene autonomously in emergencies without the need for any intervention by the human driver. In Level L4, the vehicle can be free of steering wheel, accelerator and brake pedals, but its application can only be limited to special scenarios and environments.
L5	That is, fully automatic driving, L5 does not require a driver, nor does it require anyone to intervene with the steering wheel and throttle, braking, etc., and is not limited to driving in specific scenarios, but can adapt to any scenario and environment of automatic driving.

2.2. Components of an automated driving system

Sensory layer: includes sensors such as cameras, radar and lidar.

Decision level: path planning and behavioural decisions.

Executive layer: the control system of the vehicle.

The car can sense the external driving environment through the on-board sensors in the driving process, and analyse and make decisions on the collected information transmitted to the on-board terminal in order to select the best driving scheme, and finally execute and realize automatic driving, thus achieving the purpose of liberating the driver's brain and limbs. [3]

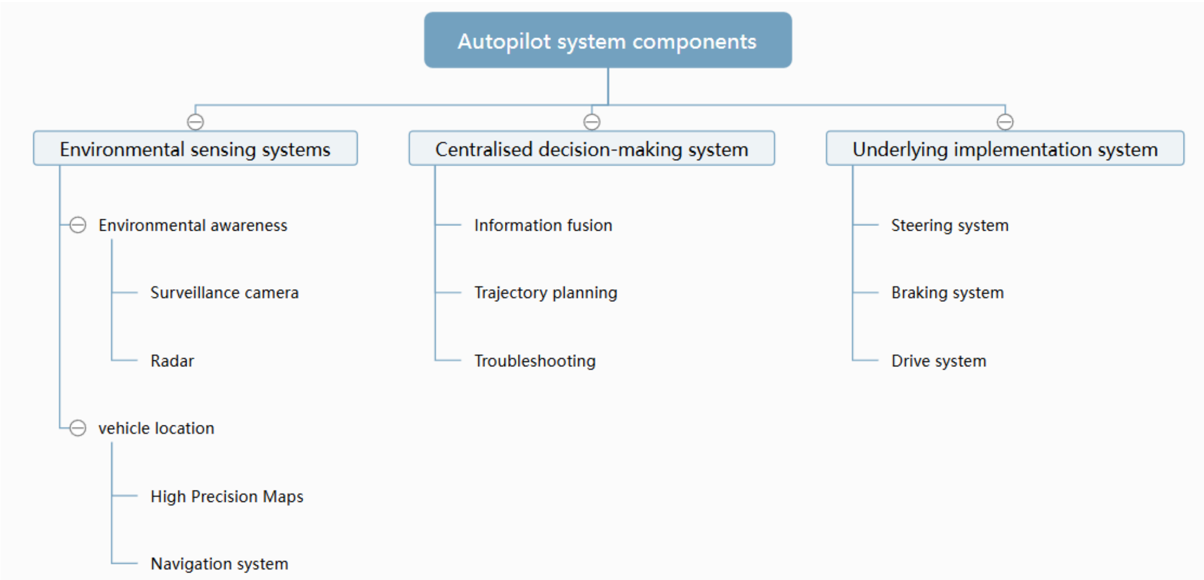

Figure 1. Autopilot system components[4]

3. Basic concepts ofAI chips

3.1. Definition and classification ofAI chips

DEFINITION: An AI chip (Artificial Intelligence Chip) is a class of hardware accelerators specifically designed to perform Artificial Intelligence (AI)computational tasks. These chips are designed to increase the processing speed, energy efficiency, and performance of AI models. They are commonly used for machine learning, deep learning, and neural network computing.

Classification: GPU, FPGA, ASIC, TPU, etc.

Table 2. Mainstream AI chips [5]

Typology	GPU	FPGA	ASIC	brain-like chip
Degree of customisation	General purpose	Semi-customisation	Fully customised	Human brain simulation
Vantage	Highly versatile	Flexibility Low power consumption	Dedicated Lower power consumption Low cost of mass production	Cognitive ability Efficient communication Very low power consumption
Drawbacks	High power consumption	High programming threshold High unit price for mass production Peak power is not as good as GPU	High cost of premise inputs Long development cycle High technology risk	
Applicable algorithms	Massively parallel computing	Massively parallel computing	Deep learning algorithm	Universal Intelligent Computing
Areas of application	3D image processing	Dedicated areas with frequent algorithm updates or small market sizes	Specialised areas of high market demand	Brain-like application areas
Applicable Scenarios	Cloud	Cloud	Cloud, Intelligent Terminal	Cloud, Intelligent Terminal
Representative Manufacturers and Products	NVIDIA:Tesla	Xilinx: V7-690T	Google: TPU Cambrian:DIANNAO	IBM: TrueNorth Qualcomm: Zeroth
Maturity level	Proven application for training in the cloud	Faster development in the data centre business	Fast evolutionary stage	In the exploratory phase
Room for development	GPUs and FPGAs are monopolised by Nvidia, Xilinx and other large manufacturers outside the enclosure, and the technical and ecological barriers are very high. Domestic AI chip enterprises can only carry out secondary development or optimisation based on their basic architecture.		The emergence of international giants has yet to be seen, and there is a good opportunity for "China Chip" to overtake the world.	The global start-up time is almost the same, and there are still a lot of basic theories to be researched.

3.2. Difference between AI chips and traditional CPU/GPUs

The main difference between AI chips and traditional CPU/GPUs is their design purpose and direction of architectural optimisation. While traditional CPUs (Central Processing Units) are designed for general-purpose computational tasks with high flexibility and low parallel processing power, and are suitable for complex control logic and diverse tasks, GPUs (Graphics Processors) are optimised for parallel processing of a large number of simple computational tasks, and were initially designed for graphic rendering, but are now widely used for data-parallel computational tasks such as image processing and deep learning. In order to perform complex parallel computations and fast graphics rendering, GPUs have far more cores than CPUs, but each core has a relatively small cache and a simpler unit of digital logic, making them better suited to computationally-intensive tasks. Intel's GPUs are primarily used as integrated graphics cards on Intel's motherboards and CPUs, whereas Nvidia and AMD have an advantage in discrete graphics cards. Nvidia and AMD have an advantage in discrete graphics. The training process of deep neural networks is extremely computationally intensive, and the data and operations can be highly parallelised. GPUs have the ability to parallelise massive amounts of data and are equipped with a large amount of computational resources for floating-point vector operations, which coincides with the needs of deep learning, and were therefore first introduced to run deep learning algorithms, and have become one of the mainstays of the high-performance computing field. However, since GPUs cannot support complex programme logic control, high-performance CPUs are still required to form a complete computing system.[6]

In contrast, AI chips, such as neural network processor NPUs, are specifically designed to accelerate AI and deep learning tasks.AI chips typically employ massively parallel processing units, low-precision computation, and specialised data-flow architectures that are capable of performing neural network computations with a much higher energy-efficiency ratio. They also integrate specialized hardware acceleration modules (e.g., tensor processing unit TPUs) that optimise matrix multiplication and convolution operations to significantly improve AI computational performance and energy efficiency. As a result, AI chips offer higher efficiency and performance than traditional CPUs and GPUs when handling deep learning and neural network tasks.

4. Application ofAI chips in autonomous driving

4.1. AI chip applications in the perceptual layer

Image recognition: camera data processing, object recognition and classification.

Obstacle detection: radar and lidar data processing.

Taking the application of ADAS (Advanced Driver Assistance System) as an example, ADAS needs to deal with a large number of real-time data collected by sensors such as LIDAR and cameras, and process the data in a very short period of time and provide timely feedback. The superiority of ADAS is reflected in the algorithms for optimisation of the control model and comprehensive information processing, which include neural network control and deep learning algorithms, etc. The rapid development of AI chips can meet the requirements for computing speed in ADAS such as image analysis and environment awareness. The rapid development of AI chips can meet the requirements for computing speed of image analysis and environment perception in ADAS. The computational delay of traditional chips can't meet the application scenario of unmanned driving, and only AI chips can deal with real-time traffic information and all kinds of sensor feedback information at any time. Now the major AI chip manufacturers have released or are studying the special ADAS processor chip, for example, the United States Qualcomm in the International Consumer Electronics Show launched the first Snapdragon 820A platform based on automotive-grade ADAS chip products. This product can achieve the recognition of road moving obstacles, but there is still a certain distance from commercial use. In addition, Intel, Renesas, Infineon and other companies have launched their own AI chip products based on intelligent driving.[7]

4.2. AI chip applications for decision-making

Route planning: Route selection based on environment and objectives.

Behavioural prediction: predicting the behaviour of other road users.

The application of AI chips at the decision-making level greatly improves the intelligence and reliability of automatic driving systems. Through efficient data processing and deep learning algorithms, AI chips can achieve real-time path planning, accurate behavioural prediction and rapid decision-making, providing strong technical support for self-driving vehicles, thus promoting the rapid development and popularization of self-driving technology.

4.3. AI chip applications at the execution layer

Vehicle control: real-time control of vehicle steering, acceleration and braking. Compared with the traditional vehicle control methods, intelligent control methods are mainly reflected in the application of the control object model and the use of comprehensive information learning, including neural network control and deep learning methods, etc. Thanks to the rapid development of AI chips, these algorithms have been gradually applied in vehicle control. Through the assistance of AI chip to vehicle driving, it is believed that accidents can be effectively avoided.[8]

5. Key technical issues in AI chip design

5.1. Power management

Chip energy efficiency ratio. Improving the computational performance of a chip per unit of power consumption is a core issue. The energy efficiency ratio directly affects the range of electric vehicles and the thermal management of the chip.

Low power design techniques. How to minimise unnecessary power consumption without compromising system responsiveness and accuracy.

5.2. Computational eficiency

Computing power. The huge amount of data required for deep learning computation causes the memory bandwidth to become a bottleneck for the whole system, which is known as the "memory wall" problem.

Data processing speed and latency. Related to the first problem, the large number of memory accesses and the large number of operations in the MAC array cause an increase in the overall power consumption of the AI chip.[8]

5.3. Scalability and Integration

Scalable architecture design. Architectural innovation is an unavoidable issue for AI chips. An important question needs to be answered: will there be a standalone AI processor like a general-purpose CPU? If so, what is its architecture? If not, the current AI chips, whose main goal is to satisfy specific applications, must only exist in the form of IP cores, which will eventually be integrated by a variety of SoCs (system-on-a-chip). [6]

Highly integrated chip design, integrating multiple functional modules (e.g. CPU, GPU, NPU, DSP, memory controller, etc.) on the same chip to improve system integration and overall performance.

5.4. Reliability and Safety

Fault tolerance. Ensures that the chip remains stable and mission-critical in the event of a hardware failure.

Data security and privacy protection. Prevent data leakage and malicious attacks on the chip, guaranteeing the overall security of the system.

6. Impact ofmanufacturing process on AI chip performance

6.1. Selection ofprocess nodes

Embedded FPGAs are useless if the technology does not fulfil the customer's needs for foundries and process nodes and their respective variants with the various sizes they want and the options they need. [9] Choosing the right process node is a critical aspect of AI chip design and manufacturing. Multiple factors such as performance, power consumption, cost, heat dissipation, reliability and technology development need to be taken into account in order to achieve optimal chip performance and cost- effectiveness in autonomous driving applications. [10]

6.2. Process Technology Development

The development of process technology is of great significance to the performance improvement of AI chips and the advancement of autonomous driving technology. By adopting advanced lithography, new materials and new structures, chipmakers are able to continuously shrink process nodes, improve transistor density and performance, and provide more efficient and reliable computing power for autonomous driving systems.

6.3. Process-Performance Correlation

Advances in manufacturing processes have had a significant impact on chip performance. Smaller process nodes can significantly increase computing power, reduce power consumption, and improve speed and integration, while also posing some reliability challenges. By continuously optimising the manufacturing process, the application of AI chips in autonomous driving technology will be more extensive and efficient, providing a solid technological foundation for the realisation of safe and intelligent autonomous driving.

7. Future trends and challenges

7.1. Limitations of current AI chips in autonomous driving

Artificial intelligence chips are currently in their infancy, and their development still has a long way to go. At present, all countries are still exploring the development of AI chips and looking for the direction of its development. In the field of AI chips, there is not yet a universal chip that is applicable to many aspects.[8]Although AI chips have significant advantages in enhancing autonomous driving technology, they still have limitations in terms of power consumption, reliability, data processing, versatility, manufacturing costs, and algorithm support.

7.2. Technology Innovation and Research Direction

In order to overcome these limitations, continuous innovation and improvement in hardware design, manufacturing process, heat dissipation management, security and algorithm optimisation will be required in the future. The technological innovations and research directions of the AI chip in the field of automated driving cover a wide range of areas such as high-efficiency processor design, low-power technology, advanced manufacturing process, new material application, intelligent algorithm optimisation, high-integrity design, security and reliability enhancement, and programmability and flexibility enhancement. Through continuous innovation and breakthroughs in these areas, AI chips will further promote the development of automated driving technology and realise a smarter, safer and more efficient automated driving system.

7.3. Future demand for AI chips for autonomous driving

The future of autonomous driving puts higher demands on AI chips, including computing power, power consumption, reliability, security, real-time processing capability, integration, flexibility and advanced manufacturing processes. Chip is the basis of supporting AI technology, for the moment, how to solve the current architectural problems, innovation of a new set of AI chip programme is a pressing task, but also a new challenge and opportunity, want to make AI really change this society, change people's lives, but also down-to-earth, to solve the current problems, the development of AI chip road is long and distant.[8]

References

[1] Duan W. (2024). A Brief Introduction to Automated Vehicle Driving Technology. Chinese Automatic Identification Technology (02), 66-68.

[2] Liu, W. & Lu, Cunhao. (2023). Research on environment sensing sensors for self-driving cars. Automotive Practical Technology (10), 197-203. doi:10. 16638/j.cnki.1671-7988.2023.010.040.

[3] Zong, Su-Chan. (2022). Analysis of the development trend of intelligent driving in new energy vehicles. Automotive and New Power (05), 21-24. doi:10. 16776/j.cnki.1000-3797.2022.05.005.

[4] Wang, Rainwater Y. & Zhang, P.. (2021). Analysis of key technologies for automotive autonomous driving. Automotive Practical Technology (23), 20-22+29. doi:10. 16638/j.cnki.1671- 7988.2021.023.006.

[5] Wang, X. (2020). Concept and application analysis of artificial intelligence chips. China New Communications (20), 112-113.

[6] Liu, Hengqi. (2019). Development and application of AI chips. Electronics and Software Engineering (22), 91-92.

[7] Wang, J. Q., Huang, H., & Zhou, Q. G.et al. (2019). An overview of autonomous driving development and key technologies. Electronic Technology Applications (06), 28-36. doi:10. 16157/j.issn.0258-7998.199062.

[8] Li, Li-Ting. (2019). Research Report on Artificial Intelligence Chip Technology Progress and Industrial Development. Xiamen Science and Technology (01), 1-9.

[9] Yin, Shouyi, Guo, Heng & Wei, Shaojun. (2018). Current status and trend of artificial intelligence chip development. Science and Technology Herald (17), 45-51.

[10] Jeff Dorsch.(2018). Field programmable gate array FPGA chips and their applications. Integrated Circuit Applications (01), 77-79. doi:10. 19339/j.issn.1674-2583.2018.01.020.