Dean&Francis

ISSN 2959-6157

A Comprehensive of CPU and GPU Performance and Applications in Autonomous Vehicles

Haoxuan Wu[1, *]

[1]Department of Unity Concord International School, Chiang Mai, 50200, Thailand
*Corresponding author: wuhaoxuanpeter2005@ldy.edu.rs

Abstract:

The rapid advancement of autonomous vehicle technology over the past decade has significantly increased the complexity of intelligent transportation systems. This need is clearly reflected in the requirements for high-performance computing in autonomous vehicles — a requirement that artificial intelligence, machine learning and big data analytic have all come to meet through their integration. We need the CPU and GPU power to handle processing in real-time data, sensor fusion and decision-making. We investigate how the CPU and GPU perform in an autonomous driving scenario, concluding that these tasks are complementary to one another describing where each can be best used within a heterogeneous computing architecture. The research delves into how these computing chips can be best used under the demand for real-time processing, reliability and efficiency. The study is to better understand the techniques for improving CPU-GPU collaboration, and applies these new findings on performance-intensive tasks like image recognition or track construction. The technical part involves conducting driving scenarios in the real world on three different types of CPU and several GPU, with key metrics such as processing latency, power consumption and accuracy. The actual experimental results show the advantages of GPU parallel computing and deep learning, while CPU still has a good advantage in multi-tasking ability and logical calculation. The results are important for the implementation of future autonomous driving systems, highlighting heterogeneous computing architectures as a means to optimize and support safe, effective vehicle operations.

Keywords: NVIDIA Orin X; Image processing; autonomous driving GPU; Heterogeneous computing; CPU and GPU collaboration

1. Introduction

In recent years, the autonomous vehicle technology has developed rapidly and they are playing a leading role in future transportation systems. These expensive systems are made more complicated and therefore error-prone with the rise in importance of artificial intelligence, machine learning, big data analytic. That complexity, however, creates a significant need for computing capability within vehicles. And since Autonomous vehicles require processing large amounts of data in real-time: Sensor Inputs, Environmental Awareness, and Path Planning accordingly; they need to make fast decision that guarantees the safety efficiency of their drive. This is where high performance computing comes into play, with the CPU (Central Processing Unit) and GPU (Graphics Processing Unit). The CPU is good at control logic and multi-tasking, while the GPU can process a large amount of data in parallel which benefits with image recognition tasks or deep learning. We can then come up with a cooperative model where CPU and GPU should be in autonomous driving systems that function thanks to sharing heavy computing loads.

The performance and use of CPU and GPU in autonomous vehicles, as well examining how they are used together within these systems is explored respectively by this paper. The study will explore just how to best employ CPU and GPU in autonomous driving, as they offer unique capabilities which might be paired for real-time processing, reliability and efficiency. Beyond these, the paper will investigate how heterogeneous computing architectures mixing CPU and GPU allow to reach maximum performance and processing those complex and computations of autonomous driving applications.

The paper is structured as follows. The computing requirements of autonomous vehicles will be discussed along with the main applications and advancements of CPU, GPU in this domain as part of the literature review. This second of 3 sections will describe the inter-processor communication between CPU and GPU in heterogeneous computing architectures, as well as ways to maximize their efficiency. Application: ORB-SLAM and Image Processing The application section will be for how CPU/GPU perform a real-world task, i.e. image processing/fountain catechizing as well path planning. In the end, the method of experiment design and results analysis going to compare different CPU, GPU performances and a summary will be made based on these findings. Then research directions are introduced.

1. Literature Review

2.1 Computing Needs of Autonomous Vehicles

Since any autonomous vehicle must have a sufficiently powerful computational in order to complete the steps from sensing an environment, and making decisions as how one should drive. Central to one of these needs is processing data in real time for which vehicles need millisecond timescale sensor data. Some of them are Lizards, radars, cameras and Ultrasonic sensors. Sensor fusion is useful because it enables the vehicle system to better understand what is happening in its environment. It is also necessary that a powerful enough processor accompanying the ADAS as it must calculate variables like, Road conditions or possibly other obstacles on road and according to traffic rules choose best path of any vehicle. Significant for its image recognition, essential to recognizing road signs, detecting pedestrians and staying in lanes, this task is well suited for the parallel processing power available from GPU. This leads to the necessity of high-performance computing solutions in areas like autonomous vehicles which involve real-time processing, sensor fusion, path planning and image recognition.

2.1.1 CPU in Autonomous Driving

A subset of this, such as CPU are integral to autonomous driving in providing the real-time and multitasking capabilities that enable safe operation and performance of a vehicle. However, new milestones like NVIDIA Orin X CPU have charted a rapid rise in processing power that allows more effective management of real-time data and multitasking tasks. Currently there is only minimal research available on the use of CPU in an embedded real-time reckless driving detection system for vehicle motion analysis which can achieve substantial efficiency and accuracy, one example being Tae-Wook Heo's study that was published this year 2020 [1].

Another study on CNN based autonomous driving optimization in 2020 by Yitong Huang concludes Xcentric which is aimed at enhancing performance with techniques like CPU-GPU collaboration and parallel execution modes [2].

Moreover, According to the research done by Maximilian fink 2019 mentioned that how CPU could also improve detection accuracy with multi-scale action of objectdetection and classification in autonomous vehicles [3].

Seyyed Hamed Naghavi et al, "A Deep Learning Approach for Real-Time Object Detection using CPU," June 2018. In a nutshell, Thomadakis' 2024 research presents areal-time framework for CPU-GPU cooperation that delays autonomous techniques in the execution process and hence enables better performance improvement together with scalability. And some studies highlight that CPU are of paramount significance in controlling the sophisticated functions involved with autonomous driving, their greatest strengths lying within high reliability and processing capabilities but they have potential drawbacks based on power efficiency and thermal constraints.

2.1.2 GPU in Autonomous Driving

For example, GPU are indispensable in self-driving cars for image processing tasks during the deep learning inference or video rendering (real-time 3D). For instance, a 2020 study from Tae-Wook Heo proposes an embedded deep learning application to detect real-time reckless driving using GPUs and produces excellent resource efficiency and precise vehicle motion analysis. The 2022 real-time 3D LIDAR processing study shows different inexpensive GPU can be used for similarly-priced commercial autonomous vehicle applications In the other vein, Yitong Huang (2020) goes on to investigate multi-tenant parallel CNN inference in autonomous driving and demonstrates that exploiting both CPU–GPU collaboration at feature extraction time and execution-independent par- allelism modes improves performance significantly.

In his prior year 2019 research, Maximilian Fink contended that detection of autonomous vehicles is a multi-object problem and GPU were crucial in increasing accuracy and processing speed. Seyyed Hamed Naghavi, 2018 This study uses GPU to boost the speed and accuracy of autonomous driving systems for real-time object detection as well as classification. This confirms the well-documented benefit provided by GPU in managing massive parallel computational workloads, especially with large datasets while also introducing known problems related to power dissipation and thermal management [4].

2.1.3 Heterogeneous Computing in Autonomous Vehicle Architectures

Heterogeneous computing architectures combine CPU and GPU to form the backbone of autonomous vehicles, delivering maximum performance for tackling intensive computational tasks. A case-in-point example can be found in the work of Polykarpos Thomadakis (2024) who introduced a runtime framework that allows for CPU-GPU collaboration and delivers substantial performance gains as well as scalability [5].

Also the 2021 work from Fei Yin that introduces a Cpu-GPU hybrid architecture to optimize clustering algorithms, achieving significant speedup on computing tasks [6].

Juan Fang suggests the optimal utilization of resources in a heterogeneous system and Through this study, shorter execution times will be achieved for tasks used by autonomous systems [7].

The structure of dual mobile GPU and FPGA heterogeneous collaboration computing to be used in different autonomous driving situations. All of these studies highlight the critical role of heterogeneous computing architectures for performance optimization and safe autonomous driving.

3. Methodology

3.1 Image Processing and Recognition

Autonomous vehicles have always had to fulfil demanding image processing and recognition requirements — the real-time analysis of large datasets from onboard cameras is no easy feat. GPU (especially those with architectures like NVIDIA Ampere) are built to speed up deep learning operations, which boosts the accuracy and quickness of image recognition. This is important for jobs including object detection, lane monitoring and traffic sign identification. The ability to process massive volumes of image data at speed is crucial in applications such as autonomous systems where the faster decisions are made, and more importantly correctly edition ed. Moreover, the algorithm used for path planning in-road environments is not useful at off-road. by Ether y Ramirez-Robles et al [8].

Proposed a real-time path planning algorithm designed to navigate in complex off-road environments using autonomous vehicles. Similarly, Runqi Qiu et al [9].

Proposed a terrain-informed real-time path optimization system to enhance the capabilities of off-road autonomous vehicles for navigating through challenging and dynamic environments. This work highlights the ability of GPU to efficiently perform both image recognition and more specialized path planning tasks, which are necessary for overall robustness across diverse driving conditions.

3.2 3D Mapping and Path Planning

The GPU is responsible for real-time rendering of complex 3D environmental data, while the CPU manages path optimization calculations in navigating through these environments (figure 1). For example, in the case of autonomous vehicles: a GPU generates real-time 3D maps to perceive obstacles and road conditions, whereas CPU computes optimal path planning algorithm to navigate seamlessly through dynamic environments The joint work of the enhanced computational processing makes possible for this system to execute path planning and optimization within a small period with respect to other kinds, dealing unexpected situations. The real-time path optimization is essential for a perform ant autonomous driving, and requires collaboration at the upper level of CPU with GPU.

4. Experimental Design

To evaluate the performance of CPU and GPU in autonomous vehicles, the experiment should simulate real driving scenarios and test CPU and GPU performance under different tasks. The experimental scenarios include urban traffic, complex road conditions, and obstacle detection. The tools include benchmark software and sensor simulation tools for performance evaluation. Performance indicators include processing latency, power consumption, and accuracy. To ensure the accuracy and fairness of the experiments, different CPU and GPU will process the same datasets, focusing on their efficiency and precision in 3D map construction and path planning [10].

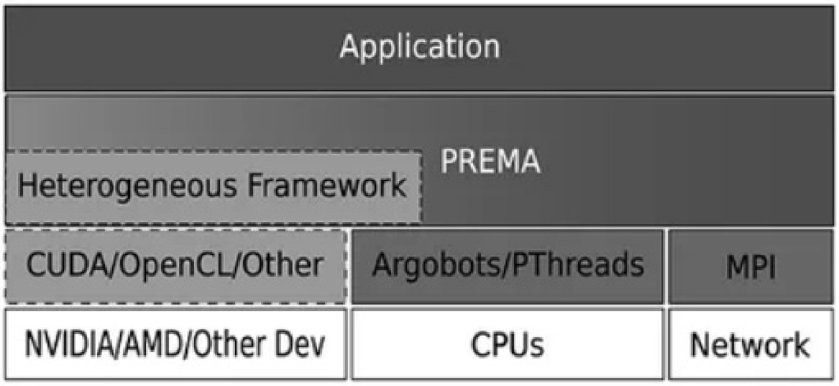

Fig. 1 A high-level representation of the heterogeneity-aware PREMA [1].

As shown in figure .1(CPU-only; see Thomadakis et al., 2022 and Thomadakis and Chrisochoides, 2023), and the heterogeneous tasking framework (in the current work). On top of that stands the application, which leverages these capabilities through a simple but powerful interface.

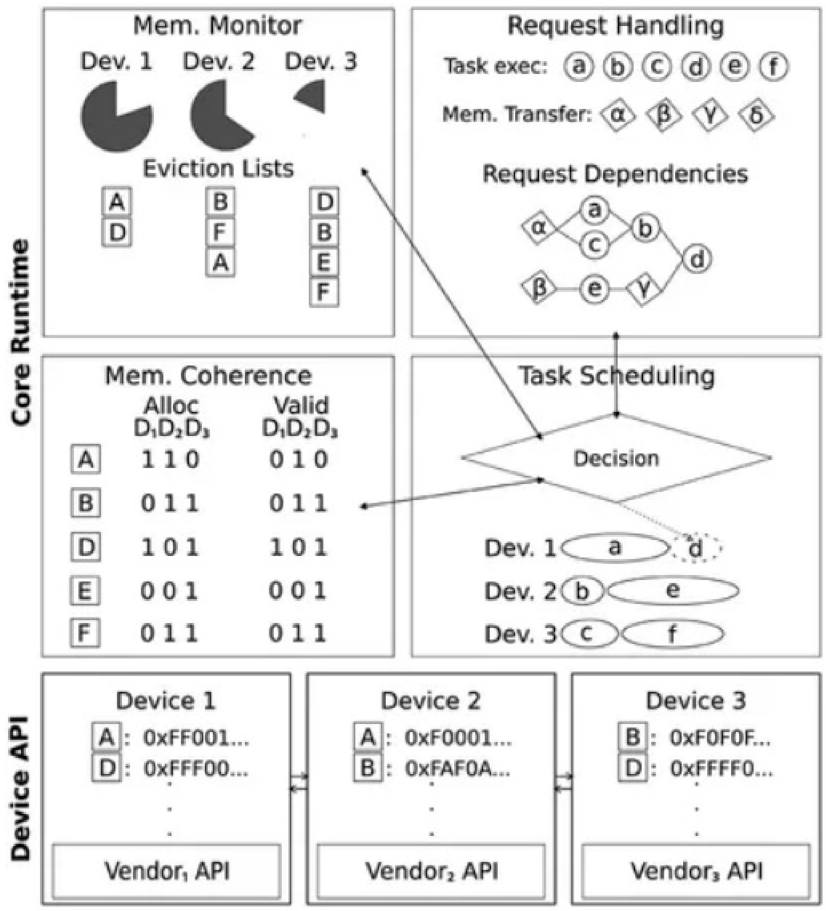

Fig.2 A high-level representation of the heterogeneous tasking framework software stack and operations [2].

As shown in figure .2 The operations performed by the Core Runtime include (a) Memory monitoring to keep track of the available device memory and deal-locate unused objects when running low on resources, (b) Memory transfer and task execution request handling that dispatches such requests when it is safe, (c) Memory coherence among different copies of the same hetero_object in multiple devices, and (d) Task scheduling to optimize for a reduction in memory transfers and optimize overall execution time. The Device API exposes an abstract "device class" that encapsulates the implementation of different vendor interfaces uniformly. The device API maps high-level abstractions like hetero_objects and hetero_tasks to actual device-specific constructs

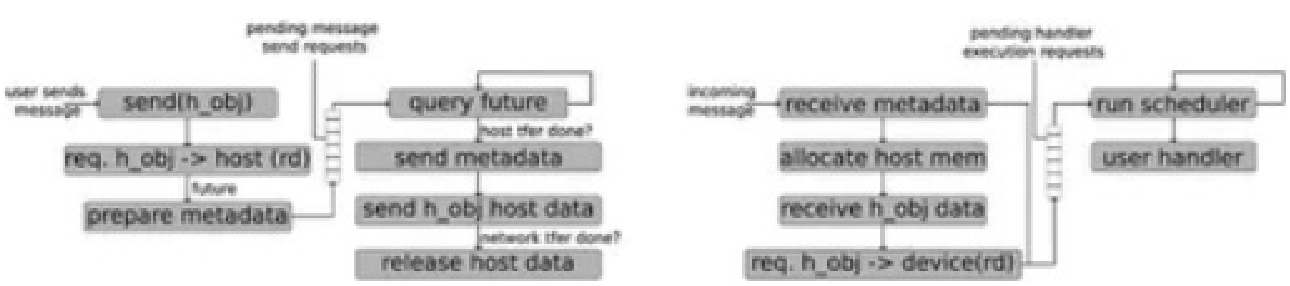

Fig.3 Heterogeneous data send (left) and receive (right) in PREMA with no GPU-aware interconnect [3].

As shown in figure .3 Depending on the capabilities of the underlying communication library and the target device hardware, the actual implementation of the memory transfers differs to leverage heterogeneity-aware communication substrates.

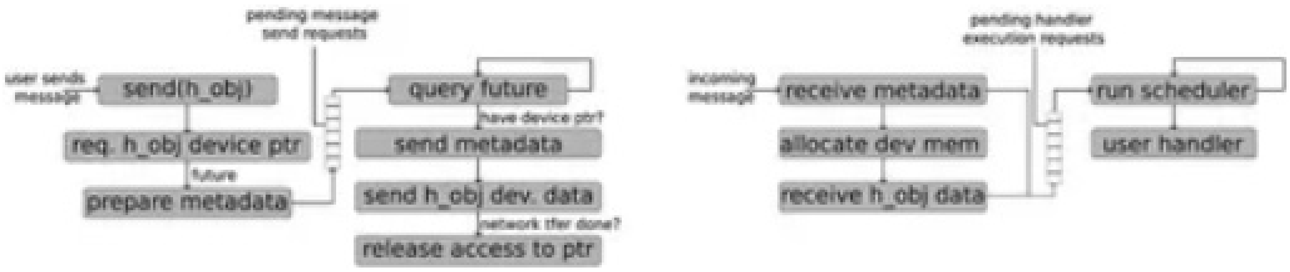

Fig.4. Heterogeneous data send (left) and receive (right) in PREMA with GPU-aware interconnect [4].

As shown in figure .4 The host-staging step can be skipped if the communication library/hardware and compute devices support direct transfers between distributed devices (currently only tested for CUDA-OpenMPI). In this case, the hardware can perform the data movement between the GPU devices directly through the network interface card (NIC), avoiding the intermediate transfers to the main memory.

The experimental results show that GPU have a significant advantage in image processing and 3D map construction, with faster processing speed and higher accuracy. CPU on the other hand, excel in path planning and decision control, effectively optimizing vehicle travel paths. When comparing different CPU and GPU, NVIDIA GPU, optimized for deep learning, demonstrate strong advantages, while high-performance CPU like NVIDIA Orin X are particularly outstanding in multitasking and logic computation. Performance differences are primarily influenced by chip architecture design, memory bandwidth, and hardware acceleration modules.

5. Conclusion

IThis paper reviews the performance and applications of CPU and GPU in autonomous vehicles, analyzing their different strengths in real-time data processing, image recognition, 3D map construction, and path planning. The experimental results indicate that GPU excel in parallel tasks and deep learning inference, while CPU are more efficient in logic control and multitasking. Both play crucial roles in autonomous driving systems, with heterogeneous computing architectures maximizing performance. Additionally, the results reveal key differences in power consumption, processing speed, and accuracy across different chips, providing insights for future autonomous driving chip design.

Future research should focus on further optimizing heterogeneous computing architectures, exploring the application of next-generation CPU and GPU in autonomous driving, such as neural network processors and quantum computing. Additionally, as autonomous driving technology matures, system security and low-power design will become critical challenges for the next
generation of in-vehicle computing chips.

References

[1] Tae-Wook Heo,Woojin Nam,Jeongyeup Paek,JeongGil Ko Autonomous Reckless Driving Detection Using Deep Learning on Embedded GPUs. Institute of Electrical and Electronics Engineers Inc. 2020,12(01): 464-472.

[2] Yitong Huang, Yu Zhang, Boyuan Feng, Xing Guo, Yanyong Zhang, Yufei Ding. A Close Look at Multi-tenant Parallel CNN Inference for Autonomous Driving. Network and Parallel Computing: 17th IFIP WG 10.3 International Conference NPC 2020, Revised Selected Papers, 2020, pp. 92-104. Springer. DOI: 10.1007/978-3-030-79478-1_8.

[3] Maximilian Fink, Stefan Feiner, Marc Erich Latoschik. Deep Learning-Based Multi-scale Multi-object Detection and Classification for Autonomous Driving. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1450-1459. IEEE.

[4] Seyyed Hamed Naghavi, Mohammad Reza Keyvanpour. Real-Time Object Detection and Classification for Autonomous Driving. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(3): 764-778.

[5] Polykarpos Thomadakis, Nikos Chrisochoides. Runtime Support for CPU-GPU High-Performance Computing on Distributed Memory Platforms. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(6): 1310-1321.

[6] Fei Yin, Xiao Liu, Jing Zhang. Cluster Optimization Algorithm Based on CPU and GPU Hybrid Architecture. Journal of Parallel and Distributed Computing, 2021, 154: 98-109.

[7] Juan Fang, Xiaohui Zhang. Resource Scheduling Strategy for Performance Optimization Based on Heterogeneous CPU-GPU Platform. Journal of Supercomputing, 2022, 78(2): 2387-2402

[8] Ethery Ramirez-Robles, Juan Perez-Limon. Real-time Path Planning for Autonomous Vehicle Off-road Driving. Sensors, 2024, 24(1): 122-134.

[9] Runqi Qiu, Xiaoping Chen. Real-Time Terrain-Aware

Path Optimization for Off-Road Autonomous Vehicles. IEEE Robotics and Automation Letters, 2024, 9(2): 675-682.
[10] Polykarpos Thomadakis, Nikos Chrisochoides. Runtime Support for CPU-GPU High-Performance Computing on Distributed Memory Platforms. Frontiers in High Performance Computing, 2024, 2: Article 1417040. DOI: 10.3389/fhpcp.2024.1417040. Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.