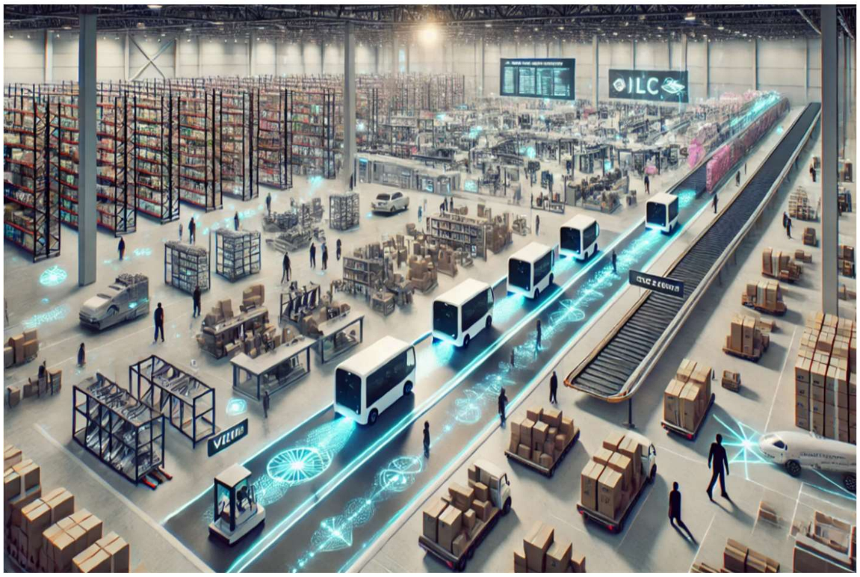


# Guidance of autonomous vehicles through Visible Light Communication

Antonio Joaquim Martins de Carvalho
(Licenciado)

Dissertação para obtenção do grau de Mestre em Engenharia de Eletrónica e Telecomunicações no Perfil de Telecomunicações.

Orientadores:
  Doutora Maria Manuela de Almeida Carvalho Vieira
  Doutora Paula Maria Garcia Louro
  Doutor Mário Pereira Véstias

Júri
  Presidente : Doutor Rui Antonio Policarpo Duarte
  Vogais:
    Doutora Paula Maria Garcia Louro
    Doutor Paulo Jorge Passo Sério Lourenço

Outubro de 2025

# Guidance of autonomous vehicles through Visible Light Communication

Antonio Joaquim Martins de Carvalho
(Licenciado)

Dissertação para obtenção do grau de Mestre em Engenharia de Eletrónica e Telecomunicações no Perfil de Telecomunicações.

Orientadores:
Doutora Maria Manuela de Almeida Carvalho Vieira , ISEL
Doutora Paula Maria Garcia Louro, ISEL
Doutor Mário Pereira Véstias, ISEL

Júri
Presidente : Doutor Rui Antonio Policarpo Duarte, ISEL
Vogais:
Doutora Paula Maria Garcia Louro, ISEL
Doutor Paulo Jorge Passo Sério Lourenço, ISEL

Outubro de 202

# Acknowledgements

Gostaria de expressar a minha mais profunda gratidão à minha família, em especial à minha esposa, pelo apoio incondicional, paciência e incentivo constante ao longo de todo este percurso. Sem a vossa compreensão e motivação, teria sido muito mais difícil alcançar este marco.

Agradeço também aos meus orientadores e professores, pela orientação científica, disponibilidade e partilha de conhecimento, que foram fundamentais para o desenvolvimento deste trabalho.

Não menos importante, quero reconhecer todos os colegas, amigos e demais pessoas que, de forma direta ou indireta, contribuíram para que este projeto fosse possível. Cada gesto de ajuda, conselho ou palavra de encorajamento teve um papel importante na concretização desta etapa.

A todos, o meu sincero obrigado.

# Statement of integrity

I declare that this dissertation report is the result of my personal and independent research. Its content is original, and all sources listed in the bibliographic references were consulted and are duly mentioned in the text. I further declare that all scientific and technical references relevant to the development of the work are duly cited and included in the bibliographic references.

The author

Antonio Joaquim Martins de Carvalho

Lisbon,10 October, 2025

# Resumo

Os Veículos Guiados Automatizados (AGVs) e os Robôs Móveis Autónomos (AMRs) estão a transformar a automação industrial, trazendo melhorias na eficiência, segurança e precisão operacional. Apesar destes avanços, continuam a existir limitações relevantes em ambientes interiores, onde as tecnologias de navegação convencionais se revelam pouco fiáveis ou ineficazes. A Comunicação por Luz Visível (VLC) surge como uma solução promissora, explorando o espectro da luz visível para possibilitar uma localização precisa e uma comunicação fiável. Ao aproveitar a infraestrutura LED já existente, a VLC oferece posicionamento com precisão ao nível da célula e permite a troca de dados em tempo real, constituindo uma alternativa economicamente viável e energeticamente eficiente aos sistemas tradicionais.

Esta dissertação analisa a integração da VLC em sistemas AGV e AMR, com particular enfoque em ambientes estruturados, como armazéns e centros logísticos. São igualmente abordados os principais desafios da VLC, incluindo a dependência da linha de vista e a suscetibilidade a interferências da luz ambiente, bem como estratégias para os mitigar. No quadro proposto, são ainda analisadas abordagens de coordenação centralizada, evidenciando o contributo da VLC para a navegação, atribuição de tarefas e gestão de tráfego em ambientes controlados.

Foi realizado um estudo de simulação com base numa grelha VLC e num coordenador central, aplicado a um ambiente de armazém estruturado em grelha. Os resultados demonstram que a VLC assegura uma localização estável ao nível da célula e suporta a operação coordenada dos veículos. Nos cenários avaliados, a abordagem heurística revelou o melhor desempenho em termos de produtividade, com menos colisões e ausência de bloqueios, enquanto um controlador central não adaptativo mostrou menor eficiência em situações de carga elevada; já o controlo preliminar baseado em aprendizagem não superou a abordagem de referência nos horizontes testados. Em síntese, este trabalho confirma a viabilidade da coordenação centralizada suportada por VLC e identifica tanto limitações práticas (linha de vista, luz ambiente, carga de comunicação) como oportunidades claras de evolução futura.

Palavras Chave

Veículos Guiados Automatizados (AGVs); Robôs Móveis Autónomos (AMRs); Comunicação por Luz Visível (VLC); Localização em ambientes interiores; Coordenação centralizada; Logística inteligente.

# Abstract

Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) are transforming industrial automation, delivering improvements in efficiency, safety, and operational accuracy. Despite these advancements, AGVs and AMRs still face limitations in indoor environments, where conventional navigation technologies may be unreliable or restricted. Visible Light Communication (VLC) emerges as a promising solution, exploiting the visible light spectrum to enable precise localization and reliable communication. By taking advantage of existing LED infrastructures, VLC provides accurate positioning and continuous real-time data exchange, offering a cost-effective and energy-efficient alternative to traditional systems.

This dissertation examines the integration of VLC into AGV and AMR systems, with a focus on applications in structured environments such as factories, warehouses, and logistics hubs. The study also considers the main challenges associated with VLC, including line-of-sight dependence and susceptibility to ambient light interference, and discusses strategies to mitigate these issues. Within this framework, centralized coordination approaches are analyzed, highlighting how VLC can enhance navigation, task allocation, and traffic management in controlled environments.

A simulation study based on the proposed VLC grid and a centralized coordinator was conducted in a grid-based warehouse environment. Results show that VLC delivers stable cell-level localization and supports coordinated operation. Across the evaluated scenarios, the rule-based baseline achieved the highest throughput with few collisions and no deadlocks, whereas a non-learning centralized controller was inefficient under load; preliminary learning-based control did not surpass the baseline within the tested horizons. Overall, the study confirms the feasibility of VLC-enabled centralized coordination and identifies practical constraints (line-of-sight, ambient light, channel load) together with clear avenues for improvement.

Keywords

Automated Guided Vehicles (AGVs); Autonomous Mobile Robots (AMRs); Visible Light Communication (VLC); Indoor localization; Centralized coordination; Intelligent logistics.

# LIST OF ACRONYMS

| | |
|---|---|
| ACC | Adaptive cruise control |
| ADAS | Advanced driver assistance system |
| ADC | Analog-to-Digital Converter |
| ADS | Automated driving system |
| AGV´s | Automated Guided Vehicles |
| AI | Artificial Intelligence |
| AMRs | Autonomous Mobile Robots |
| AV | Autonomous Vehicles |
| BSM | Blind Spot Monitoring |
| CAV | Connected Autonomous Vehicles |
| CNN | Convolutional Neural Network |
| CRC | Cyclic Redundancy Check |
| DAC | Digital-to-Analog Converter |
| DDT | Dynamic driving task |
| DQL | Deep Q-Learning |
| DRL | Deep Reinforcement Learning |
| EM | Electromagnetic |
| GPS | Global Positioning Syste |
| IMU's | Inertial Measurement Units |
| ITS | Intelligent Traffic Systems |
| LED | Light Emitting Diode |
| LDW | lane departure warning |
| LIDAR | Light Detection and Ranging |
| LKA | Lane keeping assistance |
| LOS | Line of Sight |
| MARL | Multi-Agent Reinforcement Learning |
| MPC | Model Predictive Control |
| MUI | Multi-User Interference |
| ODD | Operational Design Domain |
| OEDR | Object and event detection and response |
| OOK | On-Off Keying |
| PPO | Proximal Policy Optimization |
| PWM | Pulse Width Modulation |
| RL | Reinforcement Learning |
| RF | Radio Frequency |
| RNN | Recurrent Neural Network |
| RTK | Real-Time Kinematic |
| SAE | Society of Automotive Engineers |
| SLAM | Simultaneous Localization and Mapping |
| TDMA | Time Division Multiple Access |
| V2I | Vehicle-to-Infrastructure |
| V2P | Vehicle-to-Pedestrian |
| V2N | Vehicle-to-Network |
| V2V | Vehicle-to-Vehicle |
| V2X | Vehicle-to-Everything |
| VLC | Visible Light Communication |

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1. Introduction

This first chapter serves as the introduction to the dissertation. It presents the motivation for the work, the proposed objectives, the organization of the document, and the resulting outputs.

## 1.1 Motivation

The rapid evolution of technology, particularly in artificial intelligence (AI), neural networks and deep learning algorithms, has significantly impacted autonomous vehicle (AV) development. While early advancements were primarily aimed at personal transportation, a more immediate and transformative application has emerged in the industrial sector. Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) have become central to smart manufacturing, warehouse logistics, and factory automation, reducing human error and optimizing efficiency. Manually operated industrial vehicles, such as forklifts and conveyor systems, often lead to inefficiencies, increased costs, and safety hazards. AGVs, equipped with AI-driven decision-making, sensor fusion, and 5G-enabled communication, address these challenges by ensuring precise navigation, obstacle avoidance, and seamless coordination in structured factory settings.
A key motivation for this research lies in the integration of Visible Light Communication (VLC) into AGV navigation and fleet management. Unlike personal autonomous vehicles that rely on GPS, AGVs often operate in GPS-limited environments. VLC, leveraging LED infrastructure, provides high-precision indoor positioning and real-time communication, making it an ideal alternative for industrial automation. By enabling Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication, VLC enhances AGV efficiency and safety in logistics hubs and controlled settings. Additionally, Reinforcement Learning (RL) and Deep Q-Learning (DQL) offer promising solutions for AGV fleet management, allowing AGVs to adapt to warehouse conditions, optimize traffic flow, and minimize downtime.
This research is driven by the demand for autonomous industrial solutions that improve supply chain efficiency. By focusing on AGVs and VLC-driven navigation systems, this study bridges the gap between traditional material handling and smart automation. The findings will contribute to AI-enhanced, high-precision AGV navigation, shaping the future of autonomous logistics and industrial automation.

The activities developed in the scope of this Master's Dissertation are part of the work plan of the research project IPL/IDI&CA2024/INUTRAM_ISEL.

## 1.2 Objectives

This work aims to explore the application of autonomous vehicles for collecting, transporting, and delivering goods in indoor environments. It addresses the limitations that GPS faces indoors, where precise location data is critical. These limitations are addressed through Visible Light Communication (VLC) technology, which ensures accurate positioning by leveraging communication between vehicles and between vehicles and infrastructure.

The study further proposes an intelligent AGV flow management system for controlled aisles, utilizing VLC to enhance safe navigation along warehouse corridors. By adopting cooperative communication including Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Infrastructure-to-Vehicle (I2V) communication this approach enables access to previously unavailable information. This variety of data can optimize AGV flow and improve safety.

To further improve the performance of autonomous guidance, this research incorporates simulation software, allowing for realistic testing and optimization of vehicle control strategies, VLC-based localization, and intelligent traffic management. By simulating various scenarios, the impact of VLC communication on autonomous navigation can be analyzed, refining system performance before real-world deployment.

The combination of precise indoor positioning and intelligent traffic management systems creates a framework for optimizing the movement of autonomous vehicles in structured environments. This research emphasizes using VLC as a cornerstone technology, highlighting its potential to revolutionize indoor navigation and traffic control.

## 1.3 Document outline

The present thesis is organized into five chapters, beginning with this introductory chapter, which presents the motivation, objectives, and overall structure of the work. The second chapter provides a detailed state-of-the-art review, covering key concepts related to autonomous vehicles, their communication technologies, and the role of Visible Light Communication (VLC) in vehicular guidance and traffic control systems. This chapter also explores recent advancements in reinforcement learning techniques and their applications in optimizing intelligent transportation systems.

The third chapter focuses on the methodology employed in this research. It describes the design and development of a VLC-based intelligent traffic control system, in an industrial indoor environment, detailing the implementation of vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (I2V) communication protocols. This chapter also discusses system architecture, including signal modulation, message encoding, and the integration of reinforcement learning techniques to enhance decision-making processes in dynamic traffic environments.

In the fourth chapter, the proposed system is evaluated through simulation in the SUMO, GAZEBO or CARLA yet to be discuss software is used to create realistic traffic scenarios, allowing an evaluation of the system’s performance, assessing VLC-based communication's impact on traffic flow, coordination, and congestion, while comparing dynamic and intelligent control approaches. Neural network models RL are tested and analyzed for optimal real-time decision-making.

The fifth chapter presents the conclusions drawn from the study, summarizing the main findings and contributions of the research. The effectiveness of VLC as a communication medium for autonomous vehicle guidance and traffic control is discussed, along with the benefits of reinforcement learning in optimizing traffic flow and reducing congestion and identifies the limitations of the current implementation and proposes potential improvements to enhance system performance.

# 2. State of the Art

Although autonomous vehicles (AVs) in their generic form, such as those imagined transporting people and goods from point A to point B on public roads, are not the primary focus of this document, it is important to outline the current state of automation technologies within industrial contexts. The primary emphasis of this work is on Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs), which play a pivotal role in structured environments like warehouses, factories, and logistics centers. AGVs and AMRs have seen significant advancements in navigation, obstacle avoidance, and decision-making capabilities, driven by AI and machine learning technologies.

Throughout the next sections, we will address the technical definitions, characteristics, and frameworks governing AGVs and AMRs from both a technological and operational perspective. We will also highlight how these autonomous systems fit into the highest levels of automation, comparable in functionality to Level 5 as defined by the Society of Automotive Engineers (SAE) in its standard J3016. This comparison serves to illustrate the capabilities of AGVs and AMRs in achieving fully autonomous operations within structured indoor environments.

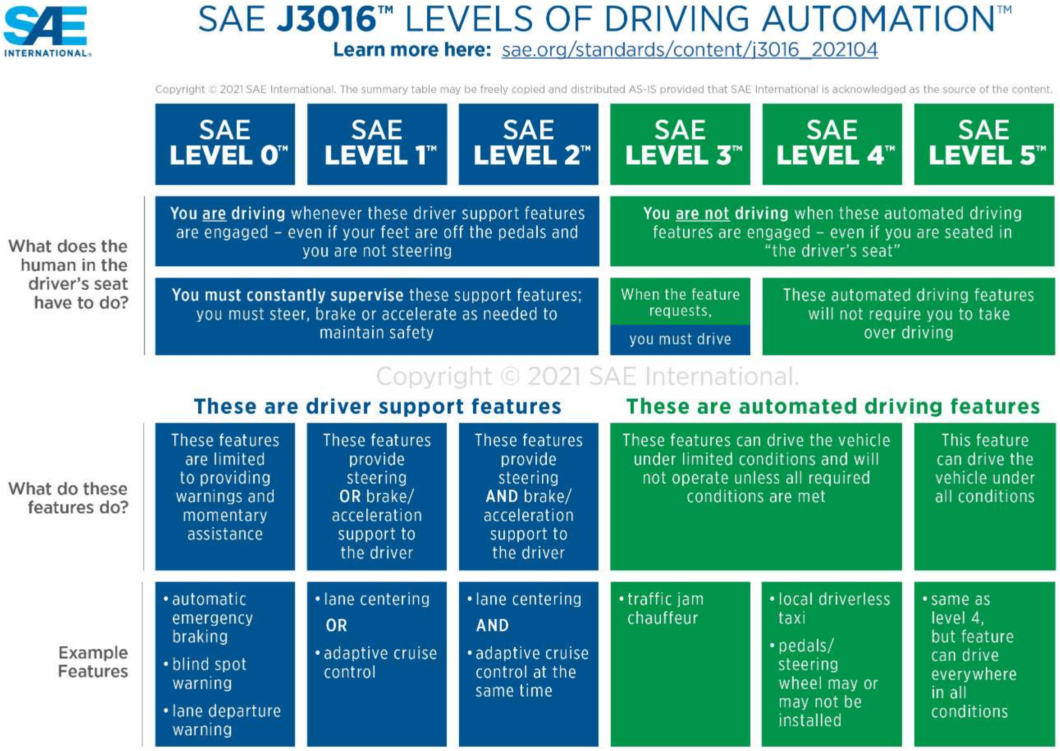
## 2.1. Autonomous Vehicles (AV)

An autonomous vehicle is a system capable of sensing its environment and operating without human involvement. In the context of personal transportation, an autonomous car does not require a human passenger to take control at any time, nor is one required to be present in the vehicle. It can navigate a wide range of environments, performing tasks equivalent to those of an experienced human driver [1]. In industrial settings, AGVs AMRs perform similar functions within structured environments such as warehouses, factories, and distribution centers. AGVs rely on predefined navigation systems, such as magnetic strips or LiDAR guidance, while AMRs use advanced perception technologies, including Simultaneous Localization and Mapping (SLAM) and sensor fusion, to autonomously adapt to dynamic environments. These autonomous systems are integral to modern logistics and smart manufacturing, optimizing operational efficiency while reducing human intervention.

### 2.1.1 AV classification

Understanding the classification of autonomous vehicles allows us to appreciate the evolution of these transport systems, understanding the necessary steps and technological advancements that have led to the current state. As defined by SAE International [2] and adopted in the United States by NHTSA (USDoT) [3] and in Europe by ERTRAC [4] [5], autonomous vehicles are classified into six levels of automation, ranging from Level 0 (no automation) to Level 5 (full autonomy). Level 0 involves no automation, with the driver responsible for all aspects of dynamic driving tasks (DDT), while Level 5 represents full automation, where vehicles operate in any environment without human intervention. While SAE levels focus on personal vehicles, industrial autonomous systems like Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) are classified based on navigation and operational criteria governed by standards such as ASTM F45, ISO 3691-4, ANSI/ITSDF B56.5, IEC 63398, and EN 1525. Unlike SAE Level 5 AVs, which navigate unstructured environments, AGVs and AMRs operate in structured settings like warehouses, using predefined paths or sensor-based navigation. Despite this difference, both Level 5 AVs and fully autonomous AGVs and AMRs share a focus on advanced navigation, obstacle avoidance, and decision-making capabilities. The ASTM F45 standard, for example, defines performance metrics for AGVs and AMRs, emphasizing navigation accuracy, docking performance, and real-time fleet coordination. The ISO 3691-4 standard outlines safety requirements for AGVs in environments shared with humans, ensuring safe interaction and minimized risk.

In figure 1 it is summarized diagram of the levels of driving automation published by SAE International focus on personal vehicles.



SAE INTERNATIONAL

SAE **J3016**™ LEVELS OF DRIVING AUTOMATION™

**Learn more here:** sae.org/standards/content/j3016_202104

| | **SAE LEVEL 0™** | **SAE LEVEL 1™** | **SAE LEVEL 2™** | **SAE LEVEL 3™** | **SAE LEVEL 4™** | **SAE LEVEL 5™** |
|---|---|---|---|---|---|---|
| What does the human in the driver's seat have to do? | **You are driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering | | | **You are not driving** when these automated driving features are engaged – even if you are seated in "the driver's seat" | | |
| | **You must constantly supervise** these support features; you must steer, brake or accelerate as needed to maintain safety | | | When the feature requests,<br>you must drive | These automated driving features will not require you to take over driving | |
| | **These are driver support features** | | | **These are automated driving features** | | |
| What do these features do? | These features are limited to providing warnings and momentary assistance | These features provide steering **OR** brake/ acceleration support to the driver | These features provide steering **AND** brake/ acceleration support to the driver | These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met | | This feature can drive the vehicle under all conditions |
| Example Features | • automatic emergency braking<br>• blind spot warning<br>• lane departure warning | • lane centering<br>**OR**<br>• adaptive cruise control | • lane centering<br>**AND**<br>• adaptive cruise control at the same time | • traffic jam chauffeur | • local driverless taxi<br>• pedals/ steering wheel may or may not be installed | • same as level 4, but feature can drive everywhere in all conditions |

*Figure 1 - SAE Levels of automation. [2]*

Level 0 involves no automation, with the driver responsible for all aspects of dynamic driving tasks (DDT). This level is limited to issuing warnings, assisting the driver to potential obstacles such as blind spot monitoring (BSM) and lane departure warning (LDW), they do not have any intervention in the vehicle driving task.

Level 1 introduced assisted driving systems such as adaptive cruise control (ACC) and lane keeping assistance (LKA), which can intervene in vehicle operation but require human supervision.

At Level 2, partial automation is achieved, where systems like ACC and LKA operate simultaneously to manage speed and steering, though human oversight remains essential.

Level 3 marks a transition to conditional automation. With this level we enter automation with vehicles able to operate autonomously in a specific set of parameters also referred to as Operational Design Domain (ODD). This concept is explored by HongSeok Cho in[6], operating out of these parameters requires fallback to human drivers to take control [7]. There are currently vehicles on the market that implement this level of automation such as Tesla, BMW (Series 7), Mercedes (EQS or S-Class).

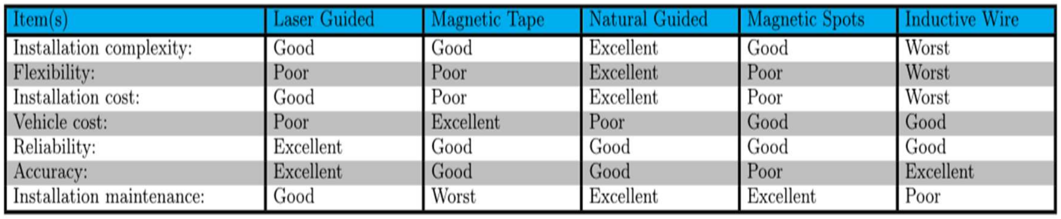
At Level 4, high automation is achieved, where vehicles can perform the entire DDT without human intervention within their ODD [2]. Companies such as Google Waymo [8] have implemented this level of automation, focusing on autonomous vehicles as a service rather than a commodity.

Level 5 represents full automation, where vehicles are capable of operating in any environment without human input. This requires advanced object and event detection and response (OEDR) capabilities, as well as robust decision-making systems capable of handling unpredictable situations and adverse conditions such as severe weather. Level 5 AVs also rely on real-time infrastructure communication through Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) networks to optimize navigation and ensure safety. Developing this infrastructure at scale involves collaboration among technology providers, city planners, and governments

While SAE levels focus on personal vehicles, industrial autonomous systems like Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) are classified based on navigation and operational criteria governed by standards such as ASTM F45, ISO 3691-4, ANSI/ITSDF B56.5, IEC 63398, and EN 1525. Unlike SAE Level 5 AVs, which navigate unstructured environments, AGVs and AMRs operate in structured settings like warehouses, using predefined paths or sensor-based navigation. Despite this difference, both Level 5 AVs and fully autonomous AGVs and AMRs share a focus on advanced navigation, obstacle avoidance, and decision-making capabilities. The ASTM F45 standard, for example, defines performance metrics for AGVs and AMRs, emphasizing navigation accuracy, docking performance, and real-time fleet coordination. The ISO 3691-4 standard outlines safety requirements for AGVs in environments shared with humans, ensuring safe interaction and minimized risk

While these classifications primarily focus on road vehicles, industrial autonomous systems such as AGVs and AMRs follow different criteria based on their navigation methods and operational autonomy. AGVs traditionally operate on predefined paths, using technologies such as magnetic strips, QR codes, or LiDAR reflectors to navigate within structured warehouse environments. Some systems also rely on wire guidance embedded in the floor or optical sensors to detect painted or reflective lines. More advanced approaches incorporate natural feature navigation and SLAM to increase flexibility, although with higher cost and complexity. These guidance methods determine how AGVs move along aisles and intersections, usually under centralized fleet management that assigns routes, controls right-of-way, and ensures safety in shared spaces. This makes AGVs particularly reliable in repetitive and predictable tasks but less adaptable to sudden layout changes. AGVs are classified based on navigation systems such as fixed-path (magnetic strips, QR codes) or more flexible approaches like LiDAR and SLAM. In [9] there is a comprehensive study of the different forms of AGV orientation, which summarizes in a simple and clear way the various advantages and disadvantages of each orientation mode (see Table 1).

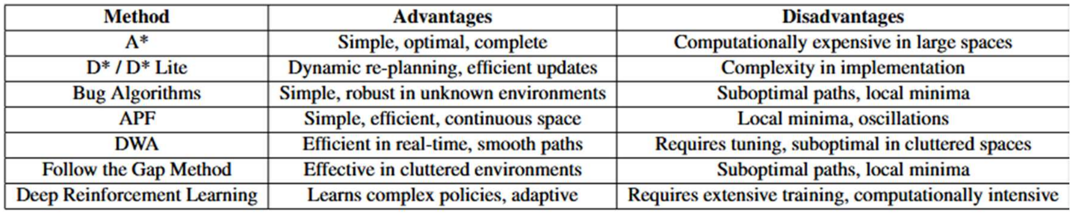
*Table 1 - AGV local navigation technique comparison [9]*



| Item(s) | Laser Guided | Magnetic Tape | Natural Guided | Magnetic Spots | Inductive Wire |
|---|---|---|---|---|---|
| Installation complexity: | Good | Good | Excellent | Good | Worst |
| Flexibility: | Poor | Poor | Excellent | Poor | Worst |
| Installation cost: | Good | Poor | Excellent | Poor | Worst |
| Vehicle cost: | Poor | Excellent | Poor | Good | Good |
| Reliability: | Excellent | Good | Good | Good | Good |
| Accuracy: | Excellent | Good | Good | Poor | Excellent |
| Installation maintenance: | Good | Worst | Excellent | Excellent | Poor |

In contrast, AMRs differ from AGVs in their ability to navigate dynamically without relying on fixed routes, using Simultaneous Localization and Mapping (SLAM) [9], deep learning, and multi-sensor fusion to adapt their paths in real time. This allows AMRs to adapt to real-time warehouse layouts and changing environmental conditions. In the master thesis [10] AGVs are described as operating on predefined routes such as magnetic strips or QR codes, whereas AMRs are analyzed in terms of trajectory prediction, path planning, and collision avoidance. Table 2 extracted from this work provides a clear comparison of different path planning methods.

*Table 2 - Comparison of Path Planning Methods [10]*



| Method | Advantages | Disadvantages |
|---|---|---|
| A* | Simple, optimal, complete | Computationally expensive in large spaces |
| D* / D* Lite | Dynamic re-planning, efficient updates | Complexity in implementation |
| Bug Algorithms | Simple, robust in unknown environments | Suboptimal paths, local minima |
| APF | Simple, efficient, continuous space | Local minima, oscillations |
| DWA | Efficient in real-time, smooth paths | Requires tuning, suboptimal in cluttered spaces |
| Follow the Gap Method | Effective in cluttered environments | Suboptimal paths, local minima |
| Deep Reinforcement Learning | Learns complex policies, adaptive | Requires extensive training, computationally intensive |

### 2.1.2 AV advantages and benefits

Progressive technological development has led to significant advancements in autonomous systems. This technology enables driving automation systems to assist or even replace human drivers, improving driving precision and helping reduce energy consumption, pollution, congestion, and road accidents [11] the advance toward automation began [12] with the implementation of active and passive safety systems in modern vehicles. For the AGV´s and AMRs the advance started in 1953 – 1956 with “Guide-o-Matic”[13]. These systems transforming industrial automation and logistics operations technologies offer numerous benefits in terms of efficiency, safety, and operational cost savings. Unlike autonomous vehicles (AVs) designed for road environments, AGVs and AMRs are optimized for structured indoor settings, such as warehouses, factories, and logistics centers. This focus enables these systems to achieve a high level of precision and reliability in environments where GPS signals may be unavailable or unreliable. AGVs and AMRs excel in navigation and obstacle avoidance, utilizing technologies such as LiDAR, cameras, radar, and AI-driven algorithms like Dijkstra algorithm[13]. These systems provide real-time situational awareness, enabling safe and accurate movement through dynamic environments. AGVs demonstrate reliability and efficiency in repetitive tasks due to their structured navigation, while AMRs achieve greater flexibility by leveraging techniques such as Simultaneous Localization and Mapping (SLAM) to adapt to changing condition [9].

The safety benefits of AGVs and AMRs are notable. By eliminating the need for human drivers in industrial environments, these systems reduce the risk of workplace accidents, which are often caused by human error. Enhanced safety mechanisms, including collision avoidance systems, emergency stop functions, and predictive AI models, ensure reliable operation even in high-traffic areas. This level of safety parallels the benefits offered by Level 5 AVs, which aim to eliminate road accidents through fully autonomous operations.

Financially, the use of AGVs and AMRs results in substantial cost savings. Automated systems operate around the clock without the need for breaks, leading to increased productivity and reduced labor costs. Additionally, optimized navigation and traffic flow reduce maintenance costs for industrial facilities. These financial benefits are comparable to those projected for Level 5 AVs, which promise to reduce costs related to accidents, insurance, and fuel consumption [9]. While Level 5 AVs are designed to operate in any environment without human input, AGVs and AMRs excel in structured environments where they can leverage fixed infrastructure and predictable workflows. However, both technologies share common challenges related to real-time obstacle detection, decision-making in dynamic environments, and seamless integration with other automated systems. Advanced AI techniques, such as (RL) and (DQL), are increasingly being applied to AGVs and AMRs to enhance these capabilities, enabling them to learn from experience and optimize their performance over time.

In logistics, AGVs and AMRs play a crucial role in automating material handling tasks, enabling round-the-clock operations, and improving supply chain reliability. Similar to the impact of Level 5 AVs on transportation networks, AGVs and AMRs contribute to the development of intelligent logistics systems by gathering data on operational efficiency, traffic patterns, and potential bottlenecks. This data-driven approach allows businesses to make informed decisions, enhance resource allocation, and optimize overall performance. As automation technologies continue to evolve, both AGVs and Level 5 AVs are paving the way for more efficient, safer, and sustainable operations in their respective domains. AGVs and AMRs are critical to the success of Industry 4.0 initiatives, facilitating smart manufacturing, intelligent warehouse management, and seamless logistics processes. By addressing current limitations and integrating advanced AI and communication technologies, these autonomous systems are poised to further revolutionize industrial automation and create lasting benefits across multiple sectors.

### 2.1.3 AV Sensors Technology

In the previous section, we discussed various systems that enable vehicle automation. This automation relies on a sophisticated suite of sensors that "sense" the environment and relay crucial information in real time. AGVs and AMRs rely on a sophisticated array of sensors to perceive their environment and ensure safe, precise, and efficient operations in industrial settings. Unlike autonomous vehicles (AVs) designed for road navigation, AGVs and AMRs operate in structured indoor environments such as warehouses and factories. The choice and integration of sensors are critical for enabling real-time navigation, obstacle avoidance, and decision-making. AGVs and AMRs commonly use a combination of LiDAR, radar, ultrasonic sensors, GPS, and safety scanners to achieve a high level of environmental awareness and situational understanding. The following image illustrates a typical commercial AGV solution equipped with key sensors, including a 3D LiDAR, cameras, and safety sensors.

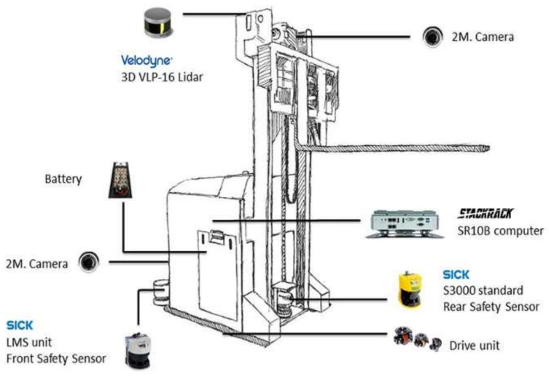


*Figure 2 - AGV´s Sensors. [45]*

Cameras play a crucial role in AGVs and AMRs due to their ability to capture high-resolution images of the surrounding environment. These images are processed using computer vision and machine learning algorithms [14] to detect obstacles, identify objects, and interpret visual cues such as signs and markers. Cameras are also essential for applications that require detailed visual information, such as quality inspection tasks.

LiDAR (Light Detection and Ranging) is a vital sensor for AGVs and AMRs, providing high-resolution 3D maps of the environment. LiDAR systems emit laser pulses and measure the time taken for the light to return after reflecting off objects. This data is used to generate detailed point clouds, enabling the vehicle to determine the size, shape, and distance of obstacles. LiDAR offers excellent accuracy and is especially effective in low-light conditions.

Radar (Radio Detection and Ranging) is another important sensor used in AGVs and AMRs. Radar systems emit electromagnetic waves and detect their reflections to determine the position, speed, and movement of objects. Radar is highly reliable in various weather conditions and excels in detecting objects in low-light environments. However, radar may struggle with false positives or difficulty distinguishing stationary objects from the background. The integration of radar with other sensors, such as cameras and LiDAR, through sensor fusion improves overall performance.

Ultrasonic sensors are commonly used for short-range detection tasks, such as collision avoidance during docking or navigating tight spaces. These sensors emit high-frequency sound waves and measure their reflections to detect nearby objects. Ultrasonic sensors are cost-effective and reliable for proximity detection and can be used as Safety scanners but have limited range and accuracy.

GPS provides global positioning data for navigation, allowing AGVs and AMRs to determine their location relative to the larger environment. However, in many indoor environments, GPS signals are weak or unavailable, leading to the exploration of alternative technologies like Visible Light Communication (VLC)】 [15]. VLC leverages the existing LED infrastructure to provide high-precision positioning and real-time communication between AGVs, AMRs, and a central control system. This technology is particularly beneficial in environments where GPS signals are unreliable, such as large warehouses or manufacturing facilities. Sensor fusion[16] is a critical aspect of AGV and AMR navigation. By combining data from multiple sensors, sensor fusion creates a comprehensive, real-time view of the environment. For example, when radar detects an object, cameras and LiDAR can be used to verify its position and type, improving the accuracy of the vehicle's response. This approach overcomes the limitations of individual sensors, enhancing overall system reliability and safety

## 2.2 Visible Light Communication (VLC)

Visible Light Communication (VLC) is a new paradigm that could revolutionize the future of wireless communication [17]. Operating within the visible light spectrum, specifically between 400 and 800 THz, VLC employs Light Emitting Diodes (LEDs) as both sources of illumination and data transmitters. Figure 3 illustrates the electromagnetic spectrum, highlighting the visible light range used in VLC systems. Unlike conventional radio frequency (RF) systems, which suffer from spectrum congestion, VLC takes advantage of the abundant and unlicensed light spectrum. This allows for higher data rates and more efficient communication.

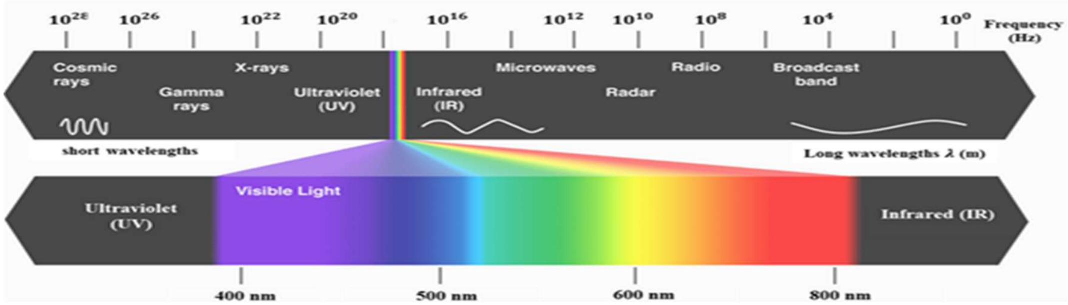


*Figure 3 - VLC Spectrum. [46]*

The core functionality of VLC relies on the ability of LEDs to modulate their light intensity at high frequencies, imperceptible to the human eye. On the receiving side, photodetectors or camera sensors decode the transmitted signals. This enables seamless data exchange without affecting lighting performance. VLC offers several advantages over RF communication. It leverages existing LED lighting infrastructure with minimal modifications. It is energy-efficient and provides enhanced security due to its reliance on line-of-sight transmission. These features make VLC particularly suitable for indoor environments such as warehouses, airports, and logistics centers. In such contexts, the technology allows a very fine-grained localization, sufficient to support navigation and coordination tasks with high reliability. This work focuses on VLC's role in autonomous vehicle systems. By integrating VLC, vehicles can enhance communication, positioning, and operational efficiency, especially in indoor environments where GPS is unreliable.

### 2.2.1 VLC Applications

VLC has a wide range of applications across multiple domains. It is used in smart cities, the Internet of Things (IoT), and vehicle communications [18]. It also plays a key role in autonomous vehicle systems. Figure 4 illustrates several VLC applications. These include road traffic safety, medical applications, industrial automation, and high-speed internet. VLC is also valuable for underwater communication, indoor localization, and RF-restricted areas such as aircraft and oil platforms.

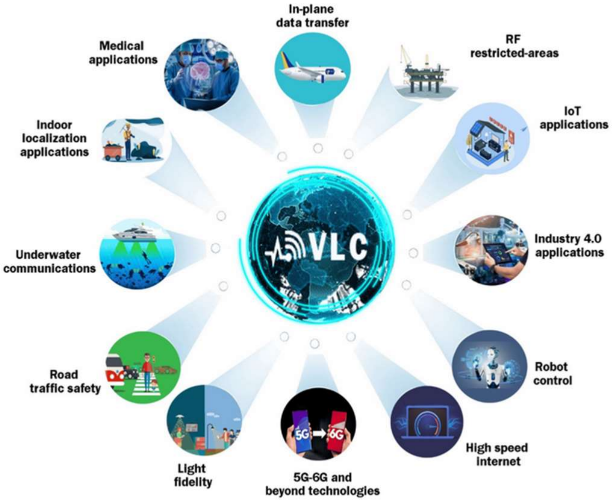


*Figure 4 - VLC Applications[47]*

This work focuses on VLC's role in autonomous vehicle communication, particularly in controlled indoor environments like warehouses and airports or other industrial settings. These locations provide ideal conditions for VLC deployment, offering low interference and easy infrastructure integration.

In industrial contexts, VLC enables several critical functions. Automated Guided Vehicles (AGVs) rely on VLC grids implemented through LED lighting to achieve precise positioning and reliable communication. This allows accurate navigation along aisles, coordination at intersections, and safe interaction with other vehicles and infrastructure. By ensuring centimeter-level localization and reliable coordination, VLC contributes to safer AGV operation in shared spaces, improving fleet management and optimizing material handling workflows.

While VLC has also been explored in road transport — for example, using headlights and taillights as transmitters to support Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I) communication — its greatest potential lies indoors, where conventional RF-based systems face reliability issues. In warehouses, VLC-based communication not only ensures safe AGV guidance but also supports workflow optimization by providing continuous and interference-free data exchange [19], [20].

By utilizing the high bandwidth and low congestion of visible light, VLC offers a scalable and reliable communication solution tailored to modern intelligent transportation systems

### 2.2.2 VLC Transmitter and Receiver Equipment

The ability of LED technology to modulate light intensity at high frequencies, without affecting its primary lighting function forms the foundation of VLC's functionality. This modulation takes place above the flicker fusion threshold [21] [22], ensuring that illumination remains imperceptibly unchanged to the human eye.

On the receiver side, photodetectors or camera sensors are deployed to decode these modulated light signals. The integration of VLC systems is particularly appealing because it leverages existing LED lighting infrastructure to transmit data, making it both cost-effective and practical for large-scale deployment.

LEDs serve as the core components of VLC transmitters, offering high-speed data transmission capabilities while maintaining their illumination functions. The inherent energy efficiency of LEDs further strengthens their role as transmitters in VLC systems. Shown in figure 5 is a simple VLC transmitter and receiver system obtained from [23], where the digital signal is converted and modulated through a DAC, amplified and transmitted via LEDs. On the receiver side, a photodiode converts the incident light back into an electrical signal, which is then amplified, filtered, and digitized by an ADC, enabling the recovery of the original data stream.

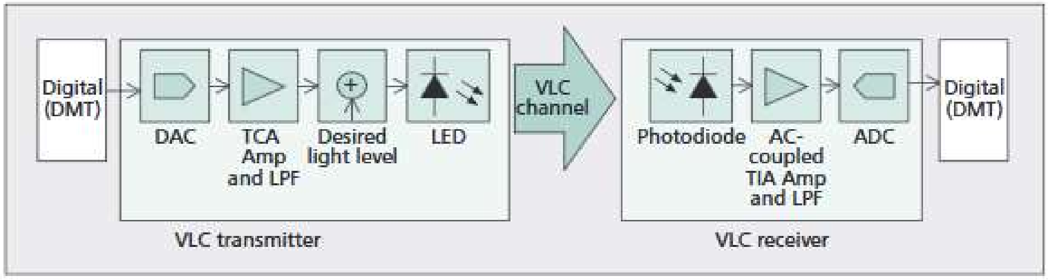


*Figure 5 - VLC transmitter and receiver system (Obtained from [23])*

The approach in [24] key point for the start of this work, where the VLC system consists of tetra-chromatic LEDs installed in the ceiling of an industrial environment. Each LED operates across different wavelengths (red, green, blue) to define specific navigation cells. These cells provide spatial information that enables AGVs and AMRs to accurately determine their position within indoor environments. Each navigation cell transmits optical signals, which are detected by a photodiode-based VLC receiver mounted on the AGV or AMR. The receiver is composed of a heterostructure with selective spectral sensitivity to filter and amplify light signals. This enables it to decode modulated signals effectively, allowing for real-time navigation and obstacle detection. The VLC receiver measures the optical power delivered by each LED in the navigation cell, enabling vehicles to determine their position with remarkable exactness. This setup supports real-time navigation and communication by integrating spatial information with control algorithms for AGV or AMR path optimization and movement coordination.

The VLC system achieves this positioning using line-of-sight (LOS) communication, where Lambertian beam distributions model [25] each LED's coverage area. Modulation techniques like On-Off Keying (OOK) and Manchester coding encode data into variations in light intensity. These methods provide both high-speed communication and robust signal decoding, ensuring accurate data exchange and minimal decoding errors. By using this infrastructure-based VLC setup, AGVs and AMRs can navigate complex layouts and detect potential obstacles or hazards in real-time, improving safety and operational efficiency.

The key advantages of this VLC-based approach include immunity to electromagnetic interference, compatibility with existing LED infrastructure, and enhanced data security due to the LOS communication. This makes VLC a reliable and scalable technology for real-time navigation and communication in industrial automation environments such as warehouses, logistics hubs, and manufacturing facilities. For vehicle-to-Infrastructure (A2I) communication, LED panels and ceiling lights equipped with VLC transmitters, as illustrated in figure 6, relay critical navigation data to AGVs and AMRs over short distances, enabling precise path planning, real-time navigation, and obstacle avoidance.

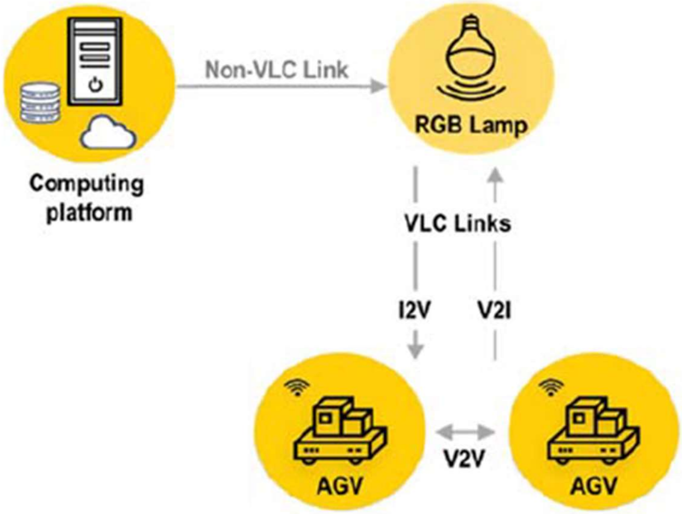


*Figure 6 - Critical navigation data. (Obtained in [19])*

In infrastructure-to-vehicle (I2V) communication, fixed LED markers exemplified in figure 7 can transmit safety-related information and real-time route updates. Enhancing the accuracy and efficiency of indoor fleet operations.

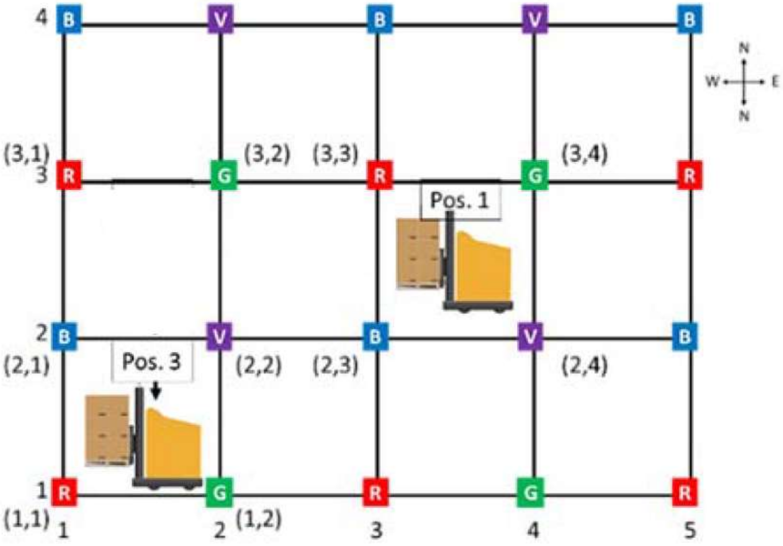


*Figure 7 - Infrastructure-to-vehicle (I2V) communication. (Obtained in [19] )*

Vehicle-to-vehicle (AGV-to-AGV) communication leverages VLC technology by using light signals emitted from AGV-mounted LEDs, allowing units to share positional data and coordinate movements in confined spaces. This setup supports the exchange of safety-critical information, such as collision avoidance alerts and traffic coordination signals, ensuring smoother interactions between multiple AGVs and AMRs. VLC's reliance on LOS communication enhances security, as unauthorized interception of signals is significantly more challenging compared to radio-frequency-based communication methods.

VLC's utility in AGV and AMR navigation is further bolstered by its ability to achieve precise positioning in indoor environments. Such high-resolution localization can be achieved with minimal hardware modifications and at a relatively low cost, making VLC an attractive option for real-time positioning and navigation in warehouses, logistics hubs, and manufacturing facilities.

Furthermore, advanced modulation techniques like On-Off Keying (OOK) and Pulse Width Modulation (PWM) [26]enable the efficient encoding of data into variations in light intensity. For more advanced applications, multiplexing methods are employed to combine multiple data streams, maximizing the utilization of available bandwidth. This encoded information is then transmitted to the receiver, where photodetectors or cameras decode the signals for further processing. The synergy between VLC transmitters and receivers forms the backbone of this communication technology, making it a promising alternative to traditional wireless methods for AGVs and AMRs operating in structured indoor environments.

### 2.2.3 VLC Challenges

While Visible Light Communication offers numerous advantages, it is not without its challenges. One of the most significant limitations of VLC is its reliance on line-of-sight (LOS) communication [26].[27] Any physical obstruction between the transmitter and receiver can disrupt the signal, posing challenges in dynamic environments where objects or vehicles may block the light path. Additionally, VLC systems are susceptible to interference from ambient light sources, such as sunlight and fluorescent lamps, which can degrade signal quality and reliability. Another challenge lies in the limited coverage area of VLC systems. As the intensity of visible light diminishes with distance, the effective range of VLC communication is restricted compared to RF systems. This limitation necessitates the deployment of additional infrastructure to ensure consistent coverage in larger spaces. However, in controlled indoor environments, such as those considered in this work, these challenges are less pronounced. The static and predictable nature of such settings minimizes the risk of obstructions and interference, allowing VLC to function effectively. Furthermore, the high precision required for certain VLC applications, such as centimeter-level positioning, demands meticulous calibration and alignment of the system’s components. Despite these challenges, the benefits of VLC, particularly its high bandwidth, energy efficiency, and enhanced security, outweigh its limitations. By addressing these issues through careful system design and optimization, VLC can unlock its full potential as a transformative communication technology. Within the scope of this work, the focus remains on leveraging VLC’s strengths in indoor environments to facilitate efficient and reliable communication for autonomous vehicle systems.

## 2.3 Traffic Control System

Navigation for AGVs in modern industrial and logistics environments demands a shift from rigid, fixed-path guidance to adaptive, real-time control systems much like the evolution seen in urban traffic control. Traditional AGV systems, akin to fixed-time traffic signals in cities, often lack the flexibility to dynamically respond to fluctuating conditions, resulting in inefficient routing, delays, and increased energy consumption when encountering congestion or unexpected obstacles. Recent research has demonstrated that leveraging technologies such as VLC can overcome these limitations. For instance, [15] propose an indoor localization and navigation framework where trichromatic white LEDs serve as transmitters and specialized photodiodes capture positional data. This infrastructure-to-vehicle (I2V) communication enables continuous monitoring of AGV locations and dynamically adjusts guidance signals to reflect real-time operational conditions mirroring the adaptive data inputs now revolutionizing urban traffic control. Moreover, integrating advanced AI-driven methods such as Reinforcement Learning (RL) and Deep Q-Learning (DQL) further refines AGV decision-making. Reviews of dynamic obstacle avoidance techniques [9] illustrate how these methods empower AGVs to rapidly assess and navigate around both static and moving obstacles, thereby optimizing their paths and reducing operational delays. In parallel, the emerging integration of 5G-based communication networks offers a robust backbone for coordinating AGV fleets. As detailed in recent reviews[13], 5G networks support low-latency, high-reliability connections that allow a central management agent to aggregate sensor data and issue real-time directives such as speed adjustments, lane reassignments, and synchronized movement to ensure smooth, collision-free operations even during peak activity periods. Modern AGV navigation systems are transitioning towards a model that incorporates dynamic data inputs and AI-powered control, analogous to the evolution of urban traffic management. By integrating VLC for precise indoor positioning, adopting advanced obstacle avoidance strategies, and leveraging 5G networks for real-time fleet coordination, these systems pave the way for safer, more efficient, and responsive material handling and smart manufacturing operations.

### 2.3.1 Reinforcement Learning and Deep Q-Learning

Reinforcement Learning (RL) [28] Deep Q-Learning (DQL) [29]have emerged as pivotal techniques in optimizing industrial traffic control and fleet management, particularly for Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs). These methods provide adaptive and intelligent decision-making strategies by leveraging real-time data, enhancing navigation efficiency, and ensuring optimized coordination in structured environments such as warehouses, factories, and logistics hubs[9]. RL functions on the principle of trial-and-error learning, where an agent interacts with its environment, observes the current state, performs actions, and receives feedback in the form of rewards [30]. In industrial traffic systems, states represent real-time operational conditions such as vehicle congestion, queue lengths, and task assignments. Actions encompass navigation decisions, dynamic rerouting, and path planning, while rewards are designed to reflect improvements in efficiency, reduced idle times, and enhanced obstacle avoidance[9]. This iterative process allows AGVs and AMRs to refine their strategies dynamically, ultimately optimizing their operations for increased productivity and safety [13]. Deep Q-Learning (DQL) extends traditional RL by integrating deep neural networks (DNNs) to approximate Q-values, enabling it to manage high-dimensional and complex state-action spaces effectively. The use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) facilitate real-time pattern recognition and prediction, making AGV navigation more efficient [31]. In industrial applications, DQL processes multi-modal sensor data such as LiDAR, vision cameras, and infrared proximity sensors to enhance navigation precision and ensure collision avoidance [9]. One of the key advantages of DQL in industrial automation lies in its ability to optimize fleet coordination. Multi-agent reinforcement learning (MARL)[32], [33] systems facilitate cooperative learning among multiple AGVs and AMRs by implementing shared reward mechanisms. These approaches ensure seamless coordination, preventing traffic bottlenecks and improving overall operational efficiency. Furthermore, by utilizing edge computing and cloud-integrated RL systems, AGVs and AMRs can process vast amounts of real-time operational data, allowing for decentralized yet coordinated decision-making[34]. Traffic control systems for AGVs and AMRs have traditionally relied on pre-programmed routing and fixed scheduling, which lack adaptability in dynamic environments. The integration of RL-based traffic management systems allows for adaptive control, where traffic signal phases and fleet coordination strategies adjust dynamically to current workload distributions and task priorities. Reinforcement learning enhances the decision-making process by considering historical performance metrics and optimizing real-time path planning for autonomous fleets [9] [34]. Visible Light Communication (VLC) has also emerged as a critical technology in enhancing AGV and AMR operations within industrial settings. By integrating VLC with RL-based fleet coordination, AGVs can achieve precise positioning and real-time data transmission, improving navigation accuracy in structured environments such as warehouses and factories. VLC-based navigation enables infrastructure-to-vehicle (I2V) and vehicle-to-infrastructure (V2I) communication, further enhancing decision-making accuracy in fleet operations.[24][31]. Additionally, advanced reinforcement learning frameworks leverage model predictive control (MPC)[35], [36] and imitation learning techniques to refine AGV behavior. These hybrid approaches enable AGVs to anticipate dynamic changes in operational conditions, reducing unexpected delays and optimizing warehouse traffic flow. Studies indicate that RL-optimized AGV traffic control systems can achieve significant reductions in congestion and improve task execution times, leading to higher throughput in smart manufacturing environments. By integrating deep learning with RL, AGVs and AMRs can continuously adapt to evolving logistics demands, minimizing idle times and improving task efficiency. The combination of deep reinforcement learning (DRL) and decentralized traffic management enhances the scalability of AGV systems, allowing them to operate autonomously across large-scale industrial environments. The ongoing advancements in AI-driven navigation and VLC-enhanced communication are setting new benchmarks for intelligent fleet coordination, ensuring a more efficient and responsive future for industrial automation [31] [34].

In conclusion, RL and DQL provide a transformative impact on AGV and AMR fleet management by enabling adaptive learning, real-time optimization, and cooperative navigation strategies. The integration of these AI-driven techniques with VLC and edge computing further enhances the efficiency and scalability of autonomous systems in industrial settings, paving the way for a fully automated and intelligent logistics infrastructure.

### 2.3.2 Central Agent AV Control

The central agent in an Intelligent Traffic Control System plays a crucial role in managing Automated Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) within industrial and logistics environments. This system is essential for optimizing traffic flow, ensuring efficient navigation, and preventing congestion in automated warehouses, smart manufacturing floors, and distribution centers. By integrating real-time data from Visible Light Communication (VLC) infrastructure and onboard AGV/AMR sensors, the central agent enables coordinated movement, collision avoidance, and route optimization [26].

The central agent functions as the primary decision-maker in AGV/AMR navigation, processing real-time data from VLC signals, vehicle-to-infrastructure (V2I) links, and infrastructure-to-vehicle (I2V) communications. It continuously monitors AGV positions, speeds, and load statuses, ensuring that routes are dynamically adjusted based on warehouse traffic density, operational priorities, and environmental conditions. Similar to how urban traffic control systems adapt to vehicle flow, this approach minimizes bottlenecks and optimizes material transport efficiency [25] [26]. VLC serves as a key enabler for precise indoor localization and bidirectional communication between AGVs/AMRs and infrastructure. Studies have demonstrated that VLC-equipped LED systems can transmit highly accurate positional and operational data, significantly improving AGV navigation accuracy and reducing delays. For instance, in a warehouse environment, tetrachromatic LEDs mounted on the ceiling provide high-resolution localization for AGVs, while photodetectors on AGVs decode incoming signals to determine their position relative to the infrastructure. This allows AGVs to receive navigation instructions, detect lane assignments, and adjust their speed dynamically in response to real-time traffic conditions. The decision-making process of the central agent is powered by (RL) and (DQL), enabling it to learn optimal traffic control strategies for AGV fleets. This AI-driven system continuously refines AGV/AMR routing strategies by analyzing historical and real-time operational data, leading to predictive route optimization where AGVs are dynamically rerouted based on congestion and operational demand. Collision avoidance mechanisms are implemented as the agent preemptively adjusts AGV paths when potential conflicts are detected, and adaptive speed control ensures that AGVs adjust their velocity based on warehouse congestion levels and priority loads. This approach mirrors multi-intersection coordination in urban traffic systems, where RL-based algorithms optimize signal timings across multiple intersections. In the context of smart warehouses, the central agent synchronizes multiple AGVs to avoid traffic jams at critical junctions, such as loading docks, storage aisles, and picking stations. In some studies, this functionality is also described under the concept of an Intersection Manager (IM), which specifically refers to the allocation of right-of-way at crossing points in warehouses or assembly lines. In this dissertation, the IM is considered part of the broader central agent, which coordinates traffic across all junctions and ensures smooth operation without collision risks. To manage large-scale AGV fleets in smart factories, the central agent integrates 5G-based communication and edge computing architectures. The benefits of 5G in AGV management include ultra-low latency communication for real-time AGV control, high-reliability data exchange ensuring AGVs receive up-to-date navigation instructions, and cloud-assisted AGV coordination where edge nodes process local traffic data while a centralized AI system manages global optimization. By leveraging 5G connectivity and VLC-based communication, AGVs operate with precise coordination, reducing traffic congestion within industrial facilities and optimizing warehouse throughput. Modern AGV and AMR navigation systems require intelligent, adaptive traffic management to handle increasing demands in smart manufacturing and automated logistics. The central agent, equipped with VLC-based communication, reinforcement learning algorithms, and 5G networking, ensures dynamic, real-time control over AGV fleets. This AI-driven traffic control system significantly improves efficiency, safety, and operational throughput in industrial environments, paving the way for next-generation automated warehouses and smart factories,

# 3 Methodology

The development of this project is anchored in the simulation and control of Automated Guided Vehicles (AGVs) within a custom-designed warehouse environment. The primary objective is to implement a centralized decision-making framework in which all navigation logic resides in a single control agent, and AGVs operate purely as sensory and actuation units. This architecture allows the vehicles to focus exclusively on executing the instructions received from the central agent, without local decision-making autonomy. AGVs continuously report their positions, detected obstacles, and operational status to the central agent, which, in turn, determines the optimal sequence of actions for each unit in real time.

The simulator developed in this work represents a significant evolution from the original baseline environment available in the Task-Assignment Robotic Warehouse (TA-RWARE) open-source project. While the original simulator was primarily designed to test heuristic task allocation strategies in a grid-based layout, its architecture was heavily reliant on local agent logic. This limited its ability to simulate environments governed by centralized intelligence, as well as its capacity to integrate more advanced communication protocols. Furthermore, the baseline lacked features such as high-fidelity localization models, advanced collision management, and communication-based task coordination.

In contrast, the current simulator incorporates substantial modifications to support VLC-based positioning and communication as the backbone of AGV localization and command delivery. The environment now includes a visible light communication grid that not only enables precise position estimation but also facilitates robust bidirectional communication between vehicles and the central controller. This VLC-based system replaces GPS, which is impractical in indoor environments, with a scalable and infrastructure-dependent solution that is both cost-effective and technically aligned with real-world industrial deployments.

The methodology followed in this study focuses on the design, development, and evaluation of this upgraded simulator to assess centralized AGV coordination strategies under realistic indoor communication and localization constraints. The simulation environment serves as a controlled testbed in which various decision-making approaches can be implemented, measured, and compared. This chapter is divided into three sections: Section 3.1 describes the simulator in detail, highlighting its architecture and operational flow; Section 3.2 provides an in-depth analysis of the VLC-based communication system; and Section 3.3 discusses the integration of centralized control with reinforcement learning for optimizing AGV behavior.

## 3.1 Simulator Description and Operation

The simulator developed for this work is a grid-based warehouse environment specifically designed to model the movement, coordination, and task execution of Automated Guided Vehicles (AGVs) under a centralized control paradigm. Its purpose is to provide a realistic yet computationally manageable representation of an industrial warehouse in which AGVs collect items from storage shelves and deliver them to designated drop-off points. While the foundation of the simulator is based on the open-source Task-Assignment Robotic Warehouse (TA-RWARE) [37]project, its architecture has been significantly extended and restructured to address the specific requirements of this research, including the integration of VLC-based localization, reinforcement learning, and advanced traffic management logic.

At its core, the environment represents the warehouse as a two-dimensional grid where each cell can correspond to a free pathway, a storage shelf, a drop-off zone, a wall, or a VLC-enabled localization cell. This discrete representation simplifies the implementation of pathfinding, collision detection, and task allocation while preserving the ability to encode complex interaction rules between agents and the environment. Each AGV is modeled as an agent with a defined position, orientation, and internal state variables that capture its readiness to perform a task, current route, and operational status. The grid structure ensures deterministic navigation and facilitates the simulation of both small-scale and large-scale layouts by simply adjusting the grid dimensions and cell types. The simulator operates in discrete time steps. At each step, the central agent receives the full state of the environment, including the positions and statuses of all AGVs, the condition of shelves and drop-off zones, VLC-based localization data, and any reported obstacles or congestion points. Using this information, the central agent calculates the optimal set of macro-actions for each AGV. A macro-action represents a high-level command such as “move to a specified cell,” “pick up an item,” or “deliver to a drop-off zone.” These macro-actions are decomposed into micro-actions (e.g., move forward, turn left, turn right) that are executed within the grid, with the environment checking for collisions and updating the state accordingly.

The framework follows a centralized decision-making paradigm. All navigation logic is concentrated in the central agent, and AGVs act purely as sensors and actuators. They continuously report their positions, operational status, and any detected obstacles to the central agent, which in turn issues navigation and task commands. This ensures globally consistent decisions aligned with the optimization strategy.

Figure 8 illustrates a snapshot of the simulated warehouse environment operating under centralized control. The grid layout distinguishes the different structural components: white cells as navigable pathways, cyan cells as accessible sides of storage shelves, purple cells as shelf bodies, and black cells as drop-off zones. AGVs are represented as orange icons with arrows indicating their orientation, while semi-circular designs denote stationary states. This representation enables rapid visual assessment of both the static layout and the dynamic operational state of the AGV fleet.

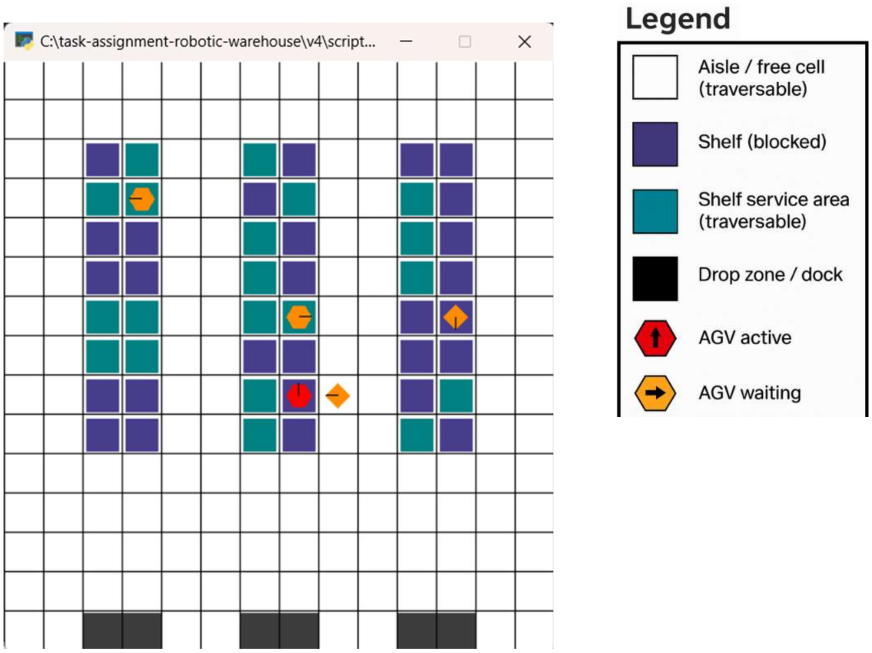


*Figure 8 - Snapshot of the simulated warehouse environment*

A key feature of the simulator is its integration with reinforcement learning (RL) for decision optimization. In particular, the Proximal Policy Optimization (PPO) algorithm is implemented. PPO is a state-of-the-art on-policy RL method that constrains policy updates to avoid destabilizing changes, achieving a balance between learning stability and continuous performance improvement. This makes it suitable for dynamic environments such as warehouses, where traffic patterns and task distributions change frequently. By interacting with the environment over many episodes, PPO enables the central agent to learn adaptive strategies for task allocation and route planning, improving throughput and reducing congestion over time. The operational logic of PPO follows an Actor–Critic structure, where two neural networks work in tandem. The Actor generates candidate actions (e.g., assigning a task to a specific AGV or selecting a route), while the Critic evaluates the quality of these actions. Over time, the policy is updated in small, controlled steps, ensuring that new strategies are gradually integrated without destabilizing previously learned behaviors.

Figure 9 illustrates the PPO architecture as applied in the simulator: the central agent receives state information, the Actor proposes an action, which is executed; the Critic evaluates it, and the policy is updated in a closed feedback loop.

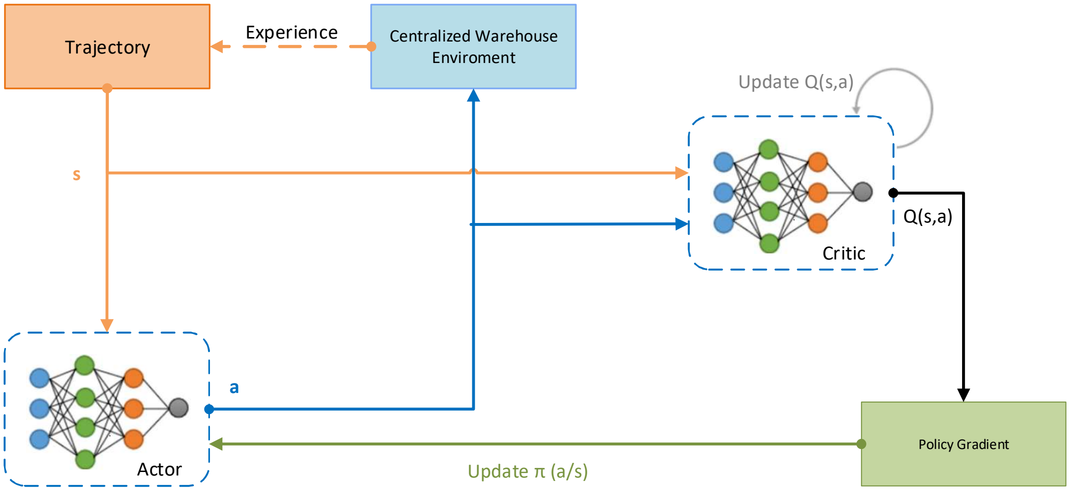


*Figure 9 - PPO architecture as applied in the simulator.*

Beyond reinforcement learning, the simulator includes an enhanced collision detection and avoidance system that extends the basic TA-RWARE logic. In the original version, collision handling was limited to preventing two AGVs from occupying the same cell in the same time step. The updated system continuously monitors planned paths several steps ahead, identifying potential conflicts before they occur. When a possible collision is detected, the central agent proactively adjusts routes or temporarily pauses one vehicle to avoid blockage. This predictive mechanism is crucial in centralized control, where fleet efficiency depends on coordinated movement.

Task generation is handled through a dynamic request queue, where each task specifies a pick location and a delivery location. Tasks are generated based on configurable parameters such as warehouse throughput, number of shelves, and complexity of operations. The central agent assigns these tasks according to its policy, which in PPO-driven runs is influenced by the learned strategy. The simulator tracks every task from assignment to completion, recording time-to-delivery, number of collisions, and travelled distance, providing a rich dataset for performance evaluation.

Scalability was a key design consideration. Grid size, number of AGVs, pickers, and request queue length can be modified to simulate a variety of scales and operational loads. The modular architecture also enables integration of new features such as alternative localization systems or decision-making algorithms without disrupting the core loop.

From a software perspective, the simulator is structured around a main environment class that manages initialization, agent placement, task allocation, pathfinding, collision checks, and state updates. Pathfinding is implemented either through the pyastar2d[38] python library. The rendering subsystem produces either live visualization or frame sequences for analysis, enabling observation of AGV movements, task execution, and congestion points. With VLC localization enabled, the overlay grid is displayed, assisting in verification of localization-based navigation.

The operational loop of the simulator can be summarized in Figure 10. This is a visual example of the implemented simulation environment, showing the arrangement of shelves, passageways, and AGVs in different operational states under centralized control. The VLC positioning grid is overlaid on the circulation lanes, allowing each AGV to determine its position with high precision.

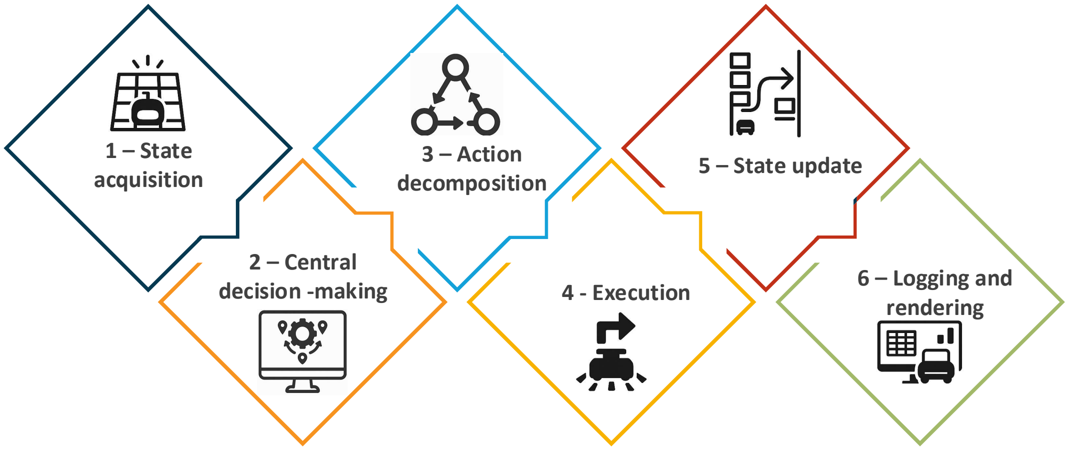


*Figure 10 - Operational loop of the simulator*

1. **State acquisition** – AGVs report their current positions (via VLC), statuses, and any detected obstacles.
2. **Central decision-making** – The central agent processes the state and determines macro-actions for each AGV.
3. **Action decomposition** – Macro-actions are converted into sequences of micro-actions.
4. **Execution** – AGVs perform micro-actions within the grid, with collision checks and dynamic route adjustments.
5. **State update** – The environment updates positions, task progress, and metrics.

6. **Logging and rendering** – Metrics are saved for later analysis, and if enabled, the current state is rendered visually.

During each simulation episode, the system logs a variety of performance metrics to quantitatively and qualitatively evaluate efficiency. These include:

- Number of completed deliveries, reflecting throughput and productivity.
- Number of collisions and stuck events, measuring safety and traffic management effectiveness.
- Total and per-agent distance travelled, assessing route efficiency and potential energy usage.
- Idle time per AGV, indicating underutilization or task allocation bottlenecks.
- Total reward assigned by the decision policy, serving as an overall optimization indicator.

By systematically recording these metrics, the simulator enables direct comparison of control policies, validation of design choices such as VLC localization, and identification of inefficiencies. Through its combination of discrete modeling, centralized control, reinforcement learning integration, and predictive collision avoidance, the platform offers a controlled yet realistic testbed for the study and optimization of AGV coordination in modern warehouse environments.

## 3.2 VLC Communication Architecture

The VLC-based positioning system implemented in the simulator provides a precise, infrastructure-dependent method of Automated Guided Vehicle (AGV) localization, specifically tailored for indoor environments where GPS is ineffective and odometry may suffer from drift. The approach overlays warehouse pathways with a visible light communication (VLC) grid, enabling AGVs to determine their position with cell-level accuracy based on light identifiers. Each pathway cell is associated with a unique VLC identifier, defined by a combination of color and spatial position within the grid, ensuring unambiguous localization and eliminating reliance on error-prone odometry.

The VLC system functions as both a localization mechanism and a bidirectional communication model between AGVs and the central coordination agent. Conceptually, LED-based transmitters are positioned in the warehouse ceiling grid, projecting encoded optical identifiers toward the floor. These identifiers represent location codes and could, in a real system, embed operational commands from the central agent. In turn, AGVs would provide uplink data such as status updates and sensor readings through high-intensity LEDs mounted on the vehicles.

The physical transmission aspects of VLC are not simulated in this work only the logical positioning and communication flow are modeled. The simulator abstracts the optical layer and focuses on deterministic positioning logic provided by the VLC grid, along with structured data exchange between AGVs and the central agent. Each simulation step associates the AGV's current (x, y) position with a VLC identifier from the grid, simulating the reading of a light signal. This value is included in the AGV's state and passed to the central agent through the *environment's step() outputs*, representing a perfect logical lookup rather than hardware decoding. In parallel, communication between agents is modeled as a synchronous data exchange: AGV states (position, load status, task status, etc.) are packaged into observation dictionaries and sent to the central agent via the gym-like API *(env.step()),* representing downlink messages.

The central agent computes macro-actions, decomposes them into micro-actions, and returns them to the AGVs, simulating uplink commands. No modulation, encoding, or channel errors are simulated the emphasis is on control logic and message structure.

From an operational perspective, the simulation represents VLC-based localization in a functional, logic-driven sequence as illustrated in Figure 11.

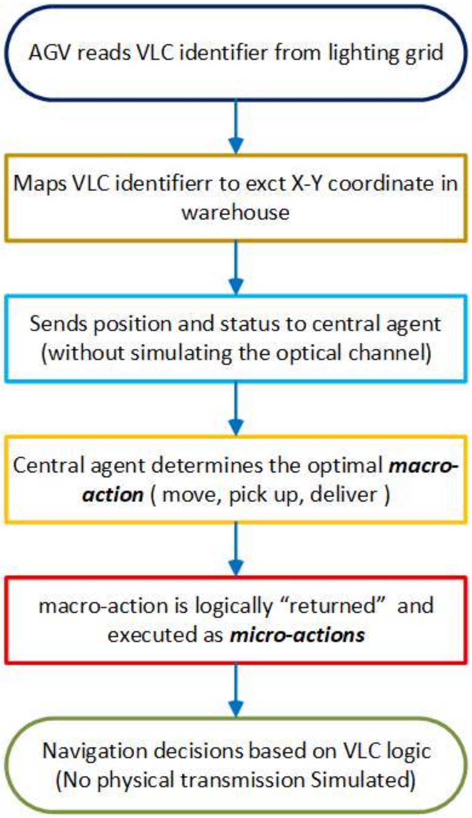
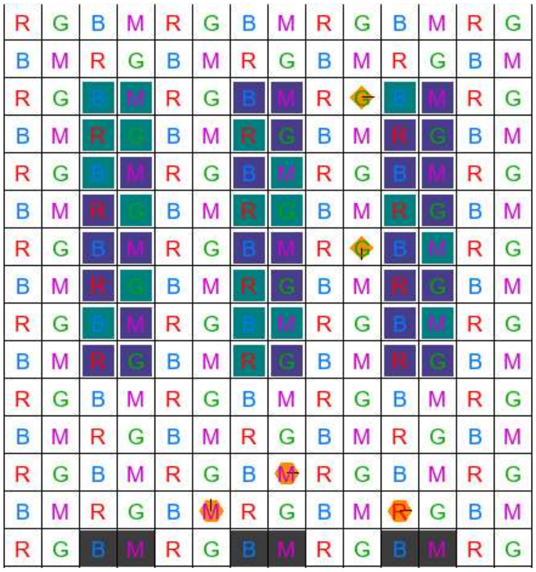


Figure 11 - VLC-based localization logic

The communication protocol follows a hardware-inspired architecture with both physical and logical definitions. The physical layer specification though not physically implemented conceptually defines LED transmitters, photodiode receivers, modulation schemes, and encoding methods. The logical layer implements structured message flow between AGVs and the central agent, including mechanisms for error detection, correction, and acknowledgement. Two frame formats are defined: a simple frame for essential position reporting and command reception, and an extended frame for richer operational data. The simple frame minimizes latency and bandwidth usage, while the extended frame provides more detailed situational context.

key innovation is the VLC localization grid itself, which overlays the warehouse pathways with color-coded light signals. Each color or color combination corresponds to a predefined spatial code, enabling scalable and deterministic localization independent of odometry or GPS. By mapping each VLC code to grid coordinates, the system ensures precise position updates synchronized with simulation steps

Figure 12 depicts the VLC grid overlay on warehouse pathways showing Red/Green/Blue/Magenta(R/G/B/M) identifiers and their spatial distribution



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |

*Figure 12 - VLC grid overlay on warehouse pathways showing R/G/B/M identifiers and their spatial distribution*

This figure illustrates the grid pattern superimposed on the warehouse floor, highlighting how each cell corresponds to a color-coded VLC identifier. The spatial distribution of Red, Green, Blue, and Magenta signals ensures that AGVs detect unique codes at every position, enabling accurate localization and synchronized control by the central agent. From a functional perspective, the VLC communication architecture in the simulation abstracts AGV–central agent interactions into periodic message exchanges represented as digital frames. While the physical channel is not simulated, the information content, structure, and timing of these frames follow realistic VLC design principles.

Although the simulated environment represents an indoor warehouse, the use of 16-bit fields for the Task ID and coordinate values (Position X and Position Y) ensures precision and scalability. Using only 8 bits would limit the range per axis to 255 units, insufficient for larger layouts or finer resolutions (e.g., 0.1 m per cell). Furthermore, task identifiers would be exhausted quickly in multi-AGV scenarios with high task turnover. By opting for 16 bits, the simulator supports larger areas, greater granularity, fractional coordinates in fixed-point format (e.g., Q8.8, meaning 8 bits for the integer part and 8 bits for the fractional part), and avoids overflow issues in long operations. The additional transmission time—only a few hundred microseconds at 1 Mbps in a real link—is negligible compared to the operational benefits.

In Table 3, the simplified frame structure is presented. With only 80 bits (10 bytes), such a frame could be transmitted in approximately 60 μs at 1 Mbps in a real VLC system. This compact design ensures low latency, making it ideal for frequent position updates or basic command exchanges. Its limitation lies in restricted field resolution (16-bit coordinates), including essential fields such as synchronization, AGV ID, position coordinates, command code, and checksum for error detection preventing detailed sensor data or complex status reporting.

Table 3 - Simple VLC Frame Structure

| Simple VLC Frame Structure | | |
|---|---|---|
| **Field** | **Size (bits)** | **Description** |
| Sync | 8 | Indicates the start of the frame (synchronization) |
| AGV ID | 8 | Unique identifier for the AGV (0–255) |
| Task ID | 16 | Current assigned task identifier. |
| Position X | 8 | Current X-coordinate in the warehouse grid. |
| Position Y | 8 | Current Y-coordinate in the warehouse grid. |
| Command Code | 8 | Encoded instruction from the central agent (e.g., MOVE, STOP, PICK). |
| Timestamp | 16 | Milliseconds since last reset. |
| Checksum | 8 | Error detection via simple XOR or CRC-8. |

In Table 4, the extended frame format is shown. This 120-bit (15-byte) frame would take approximately 180 µs to transmit at the same rate in a real VLC channel, providing high-resolution coordinates, detailed sensor readings, robust status flags, and enhanced error detection. While more bandwidth-intensive, it offers substantial benefits for complex coordination scenarios.

Table 4 - Complex VLC Frame Structure

| Complex VLC Frame Structure | | |
|---|---|---|
| **Field** | **Size (bits)** | **Description** |
| Sync | 8 | Indicates the beginning of the frame (synchronization). |
| AGV ID | 8 | Unique identifier for the AGV (0–255) |
| Task ID | 16 | Current X-coordinate in the warehouse grid. |
| Position X | 16 | Current Y-coordinate in the warehouse grid. |
| Position Y | 16 | AGV speed in cm/s. |
| Command Code | 8 | Encoded instruction from the central agent (e.g., MOVE, STOP, PICK). |
| Heading | 8 | Orientation in degrees (0–255 scale). |
| Status Flags | 8 | Bitmask for AGV status (e.g., carrying load, idle, stuck). |
| Sensor Data Block | 32 | Encoded readings from onboard sensors (temperature, proximity, battery level). |
| Timestamp | 16 | Milliseconds since last reset. |
| Checksum | 8 | Error detection via simple XOR or CRC-8. |

From a telecommunication standpoint, the choice between simple and complex frames reflects a fundamental trade-off between control loop speed and data richness. Simple frames minimize latency and bandwidth requirements, making them ideal for rapid position updates and essential command exchanges, while complex frames provide more detailed information at the cost of slightly longer transmission times. In real VLC systems, environmental factors such as ambient light noise and occlusions could introduce retransmissions, thereby increasing both latency and channel utilization. Although such effects are not physically simulated in the present work, the model provides a conceptual framework for evaluating how different frame formats may influence decision-making performance in practical deployments.

The effective performance of the VLC communication link depends directly on symbol rate, modulation scheme, and ambient light conditions. At a nominal data rate of 1 Mbps with OOK modulation, the simple frame transmits in approximately 60 microseconds, whereas the complex frame requires about 180 microseconds. Under ideal conditions, these transmission times are negligible when compared to the scale of AGV movement. In practice, the simulator does not model the physical link in detail but abstracts the communication process into synchronous data exchanges at every simulation step. Downlink communication from the central agent to the AGVs is represented by the actions issued within the step() function, which are applied directly during the update cycle. Uplink communication is modeled by the AGVs returning their current state through structured observations generated by methods such as _get_obs(). These observation dictionaries include information such as position, load status, and task state, effectively capturing what would be contained in a real VLC frame.

The distinction between simple and complex frames is implemented as a configuration of observation content: simple frames include only essential variables, while complex frames provide richer details at the cost of larger payloads. While concepts such as Time Division Multiple Access (TDMA) scheduling and error detection using checksums or cyclic redundancy codes (CRC) are referenced to place the model in a realistic context, they are not physically implemented in the simulator. Instead, communication is abstracted into synchronous data exchanges at every simulation step.

Downlink communication from the central agent to the AGVs is modeled as an instantaneous broadcast of control information, with the message format configurable between simple and complex frames. Uplink communication is represented by each AGV returning its state in structured observations generated through functions such as _get_obs(), effectively emulating the idea of reserved time slots without modeling a real TDMA protocol. To explore robustness under different conditions, the simulator also allows the artificial introduction of impairments such as transmission delays, frame losses, or simplified error flags, creating a controlled framework for assessing how communication assumptions and message complexity influence AGV coordination and overall warehouse performance.

## 3.3 Integration of VLC Localization and Centralized Control

The integration of Visible Light Communication (VLC)-based localization with a centralized control architecture represents a pivotal element in the proposed operational framework for Automated Guided Vehicles (AGVs) in structured warehouse environments. While the previous sections outlined the simulator's operational logic (Section 3.1) and detailed the VLC communication architecture (Section 3.2), this section consolidates these perspectives by describing how deterministic optical positioning is fused with global decision-making mechanisms to achieve robust, scalable, and fault-tolerant warehouse operations. This integration is not merely a technical combination of two distinct subsystems, a positioning layer and a control layer but rather the design of a tightly coupled architecture where localization accuracy, communication efficiency, and decision-making responsiveness form a closed operational loop.

In such a configuration, the VLC grid serves as the primary spatial reference framework, providing precise and infrastructure-dependent localization. Each pathway cell in the warehouse is assigned a unique VLC identifier, enabling the AGV to determine its position with cell-level accuracy at every operational step. The central agent maintains a continuously updated global map of AGV positions, trajectories, and operational states, enabling it to issue optimized navigation commands, task assignments, and conflict resolution directives in real time. This deterministic loop allows for the deployment of advanced control strategies whether heuristic-based, rule-based, or reinforcement learning (RL)-driven with the assurance that decisions are always grounded in accurate and up-to-date positional data.

The integration process begins at the localization layer, where the VLC grid provides continuous position updates. At each system tick, the AGV's onboard VLC receiver captures the identifier of the light source directly above its current grid cell. These raw identifiers are mapped to absolute Cartesian coordinates in the warehouse reference system. This mapping not only confirms the AGV's present location but also enables the system to verify route adherence, detect deviations from planned trajectories, and identify emerging conflict points with other AGVs or static obstacles. Once positional data is acquired, it is immediately relayed to the control layer, where it is merged with additional operational parameters reported by each AGV such as load status, battery level, task progress, and any fault conditions. This integration of spatial and operational data is supported by the uplink VLC channel, enabling a continuous, bidirectional exchange of information between the AGVs and the central agent.

The result of this fusion is a comprehensive state representation that serves as the foundation for centralized decision-making. To ensure efficiency and consistency, the control process follows a deterministic update cycle:

1. **Positional data update** – The latest coordinates from the VLC localization grid are recorded.
2. **Operational state ingestion** – Status information from each AGV is collected and verified.
3. **Decision computation and command broadcast** – The central agent determines the optimal macro-action for each AGV and transmits the corresponding commands.

By synchronizing communication and control in this way, the architecture eliminates asynchronous update issues. Every decision is therefore made using the most current and complete dataset, which significantly improves the reliability and responsiveness of the system.

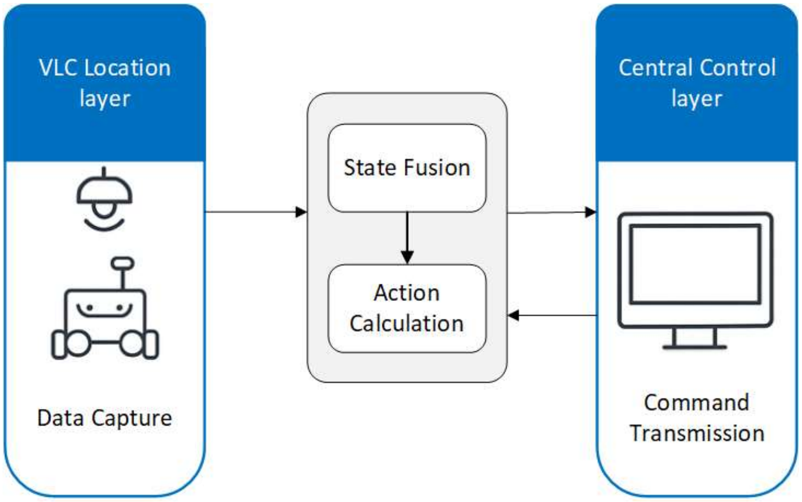


Figure 13 - Data flow

Figure 13 illustrates the sequential exchange of information between the VLC localization layer and the central control layer. The diagram highlights the main processing stages from data capture in the localization layer, through state fusion and action calculation, to the final transmission of commands to each AGV making explicit how the control loop operates in a synchronized and deterministic manner.

Centralized control architectures fundamentally differ from decentralized ones in that all decision-making occurs at a global level, rather than being distributed among individual agents. In a VLC-enabled centralized control system, this distinction becomes particularly significant, as high-accuracy, infrastructure-based localization provides the central agent with a consistent and comprehensive view of the operational environment. This global perspective enables capabilities that are difficult to achieve in decentralized systems, such as predictive collision avoidance by forecasting AGV trajectories several steps ahead, dynamic task reassignment in response to delays, faults, or localized congestion, and fleet-wide workload balancing to minimize idle time and maximize throughput. By contrast, decentralized systems rely heavily on local sensing and peer-to-peer coordination, which can result in fragmented situational awareness, slower conflict resolution, and suboptimal resource allocation, particularly in dense or high-traffic warehouse environments. Some studies point out that the multi-robot coordination literature consistently reports centralized approaches generally improve efficiency and reliability in highly coordinated tasks, while decentralized solutions may excel in scalability and fault tolerance[39] Similarly, studies have shown that multi-robot systems with centralized control often achieve higher task completion rates and more uniform coverage when compared to decentralized counterparts[40] .

Leveraging VLC-based positioning within the centralized control architecture enables the deployment of advanced predictive control strategies. By maintaining a historical record of AGV positions in the central database, the system can identify medium-term inefficiencies such as recurring bottlenecks, extended idle times, or under-utilized routing paths. A common scenario involves congestion near drop-off zones and their approach corridors, where AGVs frequently converge to complete deliveries. These insights support targeted macro-level adjustments, including the redistribution of pick-up and drop-off points to better balance traffic, the refinement of path-planning heuristics to reduce congestion, and the reprioritization of tasks to smooth operational flow during peak periods.

When integrated into reinforcement learning (RL)-based control policies, this predictive capacity enhances both performance and learning efficiency. The availability of high-fidelity historical datasets allows the learning agent to associate environmental states not only with spatial coordinates but also with operational contexts, such as congestion probability or task density in specific zones. This improved situational awareness leads to more stable policies and faster convergence, particularly in complex environments where inconsistent localization could otherwise introduce disruptive noise. As shown by [32], hierarchical multi-agent reinforcement learning models benefit significantly from structured global information in warehouse coordination tasks. Likewise,[41] demonstrate that attention-based RL architectures leveraging shared historical data can outperform baseline approaches in multi-AGV scheduling scenarios with high spatial overlap.

Figure 14 illustrates how patterns in AGV movement especially the concentration of activity near drop-off zones can be transformed into actionable insights for the central agent, enabling proactive route adjustments, balanced task distribution, and the prevention of traffic bottlenecks before they impact operational performance.

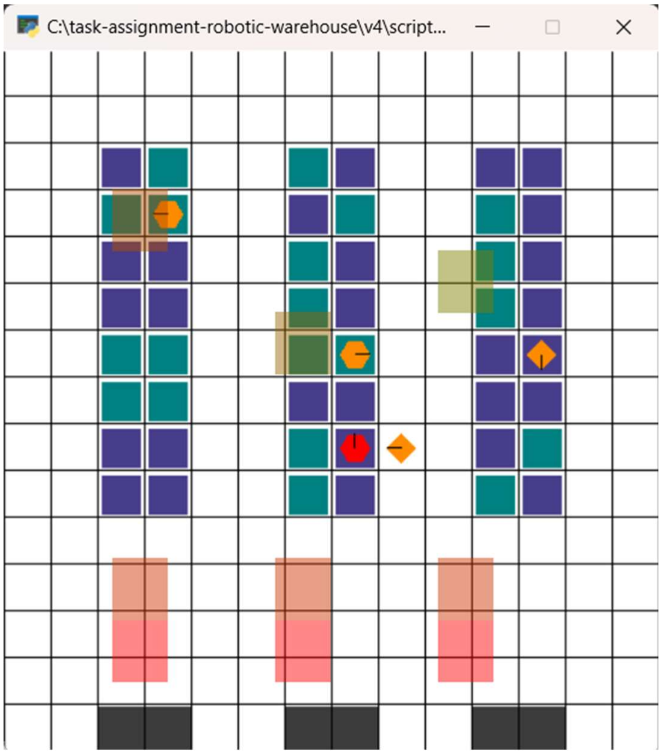


*Figure 14 - Conceptual heat map of predicted congestion*

As the system scales to accommodate larger fleets and more dynamic environments, ensuring not only fault tolerance but also efficient communication becomes critical. The scalability of the proposed architecture stems from two foundational components: the unique positional encoding provided by each VLC grid cell and the optimization of communication channels. As fleet size increases, so does the frequency of status updates transmitted over the uplink channel, which can lead to communication congestion. To mitigate this risk, the system implements coordinated transmission strategies allocating time slots for AGV status updates, prioritizing time-critical messages such as collision alerts or fault reports and employing data compression for non-essential information.

These strategies are aligned with findings in multi-robot system research and communications engineering. For example, [42] present an auction-based task allocation method integrated with motion planning under human supervision, demonstrating how continuous task inflow and dynamic reallocation can be effectively managed while maintaining coordination efficiency in mixed human–robot. Similarly, a centralized coordination framework using dynamic priority logic has been shown to effectively resolve motion conflicts among AGVs while containing communication overhead. Implementing time-based scheduling of routing and scheduling actions, as proposed by [43], also supports effective coordination without overwhelming the communication network. For typical AGV operations in VLC-enabled industrial environments, status updates include position, velocity, task state, battery level, and error flags. Recent industrial measurement campaigns confirm that such updates are transmitted as compact UDP packets with payloads on the order of hundreds of bytes at periodic rates between 5 Hz and 50 Hz [44], The corresponding communication bandwidth can be estimated using the standard formulation:

$$B = \frac{P \cdot 8 \cdot f \cdot N}{10^3} \tag{3.1}$$

B – Bandwidht (kbps)
P – Payload size in bytes per update
f – Update rate per second (Hz)
N – Number of AGV´s
8 – Convert Bytes to bits
$10^3$ – Cobert bits to kbits

Using this relation, the communication requirements for various fleet sizes and update rates were computed and are summarized in **Table 5 – Bandwidth Usage**. The values are derived by dividing the assumed frame sizes (≈60 bits for the simple format and ≈180 bits for the complex format) by the nominal data rate of 1 Mbps, and then scaling according to the number of AGVs and their reporting frequency. The results clearly demonstrate the linear relationship between fleet size, update frequency, and bandwidth demand, underlining the importance of scheduling and prioritization strategies to ensure real-time performance in large-scale VLC-based deployments.

*Table 5 - Bandwidth Usage*

| Fleet Size | Update Rate 1 Hz (kbps) | Update Rate 5 Hz (kbps) | Update Rate 10 Hz (kbps) |
|---|---|---|---|
| Small (5 AGVs) | 12.0 | 60.0 | 120.0 |
| Medium (20 AGVs) | 48.0 | 240.0 | 480.0 |
| Large (50 AGVs) | 120.0 | 600.0 | 1200.0 |

While VLC remains the primary localization method in the proposed framework, integrating it with complementary technologies such as RF communication, LiDAR, or Ultra-Wideband (UWB) positioning can enhance system robustness. These hybrid configurations allow each modality to contribute its strengths for instance, VLC provides high-precision indoor positioning, while RF handles longer-range or higher-bandwidth communication needs. This multimodal approach improves the resilience and flexibility of the system, particularly in warehouses with partial VLC coverage or where frequent reconfigurations occur.

One of the most significant advantages of an infrastructure-based VLC system is its potential for standardization. Once a warehouse is equipped with a VLC grid and its corresponding localization map, the same setup can be replicated across multiple facilities with minimal adjustments. This not only reduces deployment and calibration costs but also enables trained reinforcement learning models to be transferred between environments while maintaining consistent behavior and performance. Despite its benefits, the integrated VLC-control architecture introduces several practical constraints. Significant changes to shelving layouts or aisle configurations require updates to both the lighting infrastructure and the VLC-based localization map, making layout rigidity a factor in system adaptability. In terms of communication, while VLC provides sufficient capacity for control and status messages, it may fall short when transmitting high-resolution sensor data or complex telemetry streams, necessitating auxiliary RF links. Furthermore, because VLC relies on line-of-sight communication, temporary occlusions such as those caused by tall loads or equipment can degrade positioning accuracy. These situations require well-defined fallback strategies, such as the use of dead reckoning or estimated trajectory recovery by the central agent.

The combined use of VLC-based localization and centralized control forms a compelling architecture for indoor AGV coordination. Deterministic positioning accuracy, paired with global optimization capabilities, supports advanced navigation logic, real-time task assignment, and predictive congestion management. The system's design also supports scalability, multi-agent fault tolerance, and integration with reinforcement learning frameworks for adaptive behavior. By building on the simulation environment presented in Section 3.1 and the communication mechanisms detailed in Section 3.2, this architecture enables extensive testing under varied operational conditions from idealized zero-latency scenarios to degraded signal environments. Such flexibility ensures that the proposed control strategies remain both theoretically grounded and practically viable in real-world deployments.

# 4 Simulation

The simulation stage serves as the experimental backbone of this research, providing a controlled and repeatable environment in which the proposed VLC-based centralized coordination framework can be evaluated. Unlike physical trials, the simulator allows for extensive testing under diverse scenarios without incurring the cost, complexity, or logistical constraints of deploying an actual AGV fleet. It also enables the injection of specific variables such as communication delays, positioning errors, or sudden task surges to observe system behavior under controlled stress conditions. This section details how the simulation was designed, configured, and executed to reflect both the architectural principles outlined in Chapter 3 and the operational dynamics of an industrial warehouse. It covers the environment setup and agent modeling, the specific role of the VLC grid in positioning and coordination, and the evaluation metrics used to assess performance. The implementation is based on a customized simulator named DeepWarehouseSelf, designed to compare heuristic coordination strategies with reinforcement learning-based decision-making in a centralized control context.

While the simulation remains a simplified representation of real-world conditions, it offers valuable insights into system behavior and serves as a tool to explore the feasibility of the proposed coordination approach.

## 4.1 Simulation Environment Setup

The simulation configuration was designed to mirror realistic warehouse operations while enabling fine-grained control over experimental variables. The virtual environment is structured as a two-dimensional grid representing the facility layout, where each cell corresponds to a discrete spatial coordinate. Shelves are arranged in a regular configuration of two rows and seven columns (14 total), creating aisle pathways that AGVs must traverse to complete delivery tasks. Pick-up and drop-off zones are strategically placed to ensure that AGVs navigate through multiple VLC cells, making localization accuracy a determining factor in performance.

To ensure consistency with the architectural and communication principles established earlier, the simulation integrates a series of enhancements aimed at accurately representing the operational characteristics described in Chapter 3. In particular, the VLC-based localization model is emulated through the deployment of virtual transmitters, each broadcasting encoded identifiers that follow the VLC frame structure defined in

Section 3.2. These identifiers are detected by the AGVs, which decode their current position and transmit this information to the central coordination agent.

The size of the warehouse, number of agents, and simulation timing parameters were varied depending on the test scenario. AGV fleet size ranged from 3 to 10, while up to 4 pickers were introduced in mixed-operation simulations. Each simulation episode is divided into discrete time steps, each representing 100 milliseconds of virtual time. The total duration of an episode varied from 2000 to 5000 steps depending on whether the simulation was being used for training, validation, or stress testing. To clarify these variations, the main simulation parameters are summarized in Table 6

Table 6 - Main simulation parameters

| Parameter | Value [min – max] | Description |
|---|---|---|
| Warehouse Size | [Width] × [Height] | Grid dimensions representing the full simulation environment (in number of cells) |
| Number of Shelves | [Rows] × [Colunms] | Shelves arranged in two parallel rows with seven shelf columns each (Original 2x7 , 14 Shelfs) |
| Number of AGVs | 3 – 10 | Adjusted to simulate low, medium, and high-density fleet operations |
| Number of Pickers | 0 – 4 | Optional human-like agents; included in complex scenarios with mixed coordination |
| Time Step Resolution | 100 ms | Each simulation step corresponds to 100 milliseconds of virtual time |
| Steps per Episode | 2000 – 5000 | Total simulation steps per episode; longer runs used for training stability and policy testing |
| VLC Frame Modes | 2 (Simplified, Extended) | Simplified mode includes only position/task |
| | | Extended adds heading, status, and sensor data |
| VLC Coverage | Full, Partial, Dropout zones | Full: complete lighting grid; Partial: intermittent coverage; Dropout: zones with no signal |
| Scenario Types | 4 types | Low-traffic, High-density, Mixed-load, and VLC-degraded simulations used to assess system limits |

AGVs are modeled as purely reactive entities with no embedded intelligence. Navigation, task assignment, and decision-making are entirely handled by a centralized control agent, which collects positional and status updates from the AGVs and issues corresponding movement instructions. Communication between the agents and the central controller follows structured frame formats defined in Chapter 3. Two distinct communication modes were tested:

Simplified Frame Mode, in which AGVs transmit only their location and task status, resulting in low-bandwidth communication but reduced decision granularity.

Extended Frame Mode, which adds heading direction, AGV status, and sensor readings to the transmitted data, offering richer situational awareness at the cost of greater communication overhead.

Each simulation step follows a consistent sequence: AGVs detect their position via the VLC grid, transmit an update frame to the central agent, receive movement instructions, and execute the commanded action. Possible interactions include task pick-up, delivery, or navigation within shared pathways. Throughout the simulation, the system logs a set of key performance metrics for each episode. These metrics are exported in both JSON and CSV formats to support later analysis, comparison, and visualization.

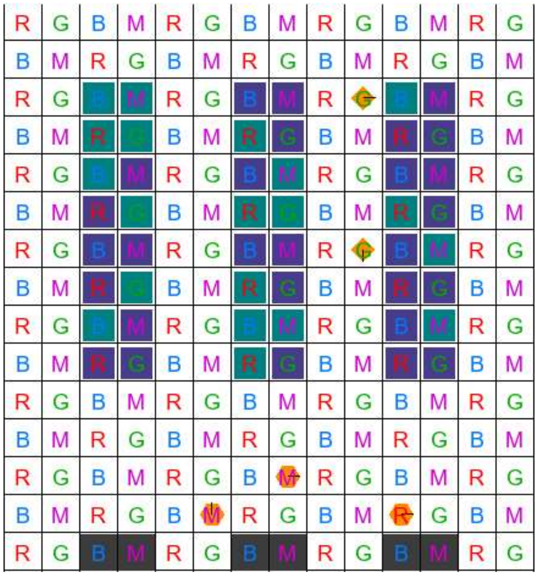
To help interpret simulation results and assess system performance under different operating conditions, Table 7 summarizes the core evaluation metrics used in this study. For each metric, the table presents its expected behavior, the outcome associated with efficient coordination, and the signs that may indicate suboptimal control, learning failure, or system bottlenecks.

Table 7 - Core evaluation metrics

| **Metric** | **Expected Trend / Range** | **Desirable Outcome** | **Undesirable Outcome** |
|---|---|---|---|
| Deliveries Completed | Should increase over time | High delivery count per AGV | Low or stagnated delivery rate |
| Collisions | Should be close to zero | No collisions | Frequent collision events |
| Idle Time per AGV | Should remain minimal | Minimal idle duration | Extended inactivity or delays |
| Distance Traveled | Context-dependent | Efficient, goal-directed movement | Redundant or excessive travel |
| Stuck Events (Deadlocks) | Should be rare or non-existent | No deadlocks | Persistent or repeated deadlocks |
| Global Reward Score | Should increase with learning | High and stable cumulative score | Negative or erratic reward patterns |

The VLC grid implementation in the simulator operates on a cell-based coverage model. Each VLC transmitter has an effective coverage area defined in the warehouse grid, and its ID is embedded in the transmitted frame. The simplified frame mode provides only positional and task-related data, while the extended mode includes sensor feedback and status metadata, enabling more adaptive and situationally aware control strategies. This dual-mode capability allows for controlled experiments comparing minimalist and enriched communication models under otherwise identical conditions.

Figure 15 illustrates a representative warehouse layout with the VLC grid rendered explicitly using overlay techniques. The VLC cell identifiers are superimposed at the center of each active transmitter zone, demonstrating the coverage granularity available for localization.



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |
| B | M | R | G | B | M | R | G | B | M | R | G | B | M |
| R | G | B | M | R | G | B | M | R | G | B | M | R | G |

Figure 15 - Illustration of warehouse layout with the VLC grid rendered

Simulation scenarios are designed to test the system under varying conditions:

Low-traffic environments for baseline efficiency.

High-density AGV traffic for testing congestion handling and collision avoidance.

Mixed-load tasking to evaluate prioritization and task reassignment logic.

Localization degradation conditions to simulate real-world lighting failures or obstructions.

A key strength of the simulation framework lies in its modular extensibility. Operational parameters such as warehouse dimensions, number of agents, observation modes, and episode length can be reconfigured externally typically via command-line arguments without altering the core simulation logic. This enables rapid scenario iteration and facilitates the introduction of advanced control constraints, including dynamic path restrictions and task prioritization schemes.

For example, a test scenario with 7×2 shelves, 10 AGVs, and 4 pickers can be executed using:

```
“python scripts/run_centralagent.py --shelf_columns 7 --shelf_rows 2 --column_height 10 --num_agvs 10 --num_pickers 4 --steps 5000 --observation_type global”
```

The main simulation parameters used throughout this work are summarized in Table 8, which outlines the configurable aspects of the virtual warehouse, AGV fleet, and communication setup. Each parameter includes its tested range and a brief description of its role within the simulation environment. These values define the operational conditions under which all experimental results were obtained.

Table 8 - Configurable aspects of the virtual warehouse

| Parameter | How to Change | Notes |
|---|---|---|
| Number of AGVs / Pickers | Command line args: --num_agvs, --num_pickers | Used to instantiate the environment |
| Warehouse layout | Pass layout preset or modify layout file | Many layouts (Scenario1, etc.) supported |
| Observation type (frame content) | --observation_type global vs local_extended | Affects how rich the communication is |
| Max steps / episode length | --steps 5000 | Used for controlling simulation horizon |
| Reward type | --reward_type individual | Configures reward propagation |
| Seed | --seed 123 | Ensures reproducibility |

Complementarily, Table 9 lists the core performance metrics recorded during each simulation episode. These indicators were used to assess system efficiency, coordination quality, and learning progression across both heuristic and reinforcement learning strategies. Each metric includes an interpretation of its ideal and worst-case outcomes, providing context for result analysis in subsequent sections.

Table 9 - Recorded performance metrics

| Parameter | Where to Change | Notes |
|---|---|---|
| Update frequency | Hardcoded in agent loop (e.g., 1 update per step) | You'd need to modify how often AGVs send frames |
| Payload content | encode_state(...) in encode_state.py | To add/remove fields (e.g., battery, heading) |
| Frame type toggle | Conditional logic in get_observation() | Tied to observation_type enum |
| Message size cap / compression | Needs to be implemented (not currently in place) | Could simulate limited-bandwidth channels |

For the visual component, the simulation includes a rendering module capable of displaying the warehouse layout, AGV positions, task locations, and agent trajectories. The VLC grid overlay, when relevant, is generated through dedicated rendering scripts to support spatial interpretation and analysis. This visual feedback is crucial for validating agent behavior, ensuring proper path allocation, and detecting anomalies such as inefficient routing or poor task distribution.

Figure 16 presents a snapshot captured during a simulation episode in which AGVs are performing active pick-up and delivery operations under centralized control. The figure illustrates the system’s spatial coordination and highlights the centralized agent’s ability to manage multiple agents simultaneously within a constrained layout.

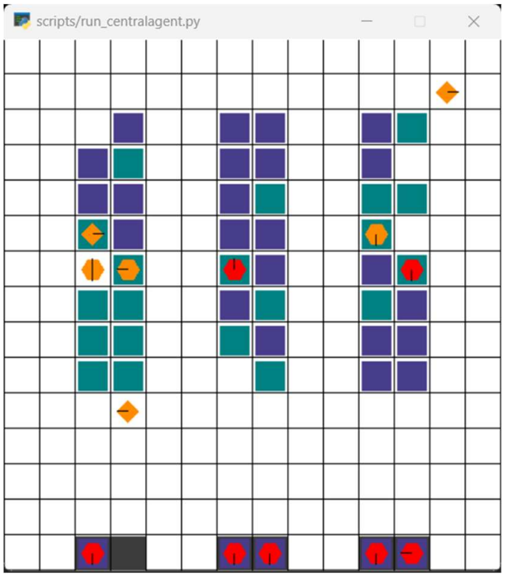
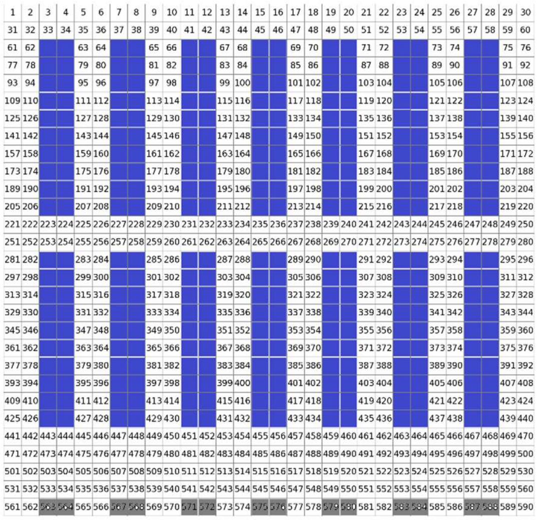


*Figure 16 - Simulation episode snapshot captured*

Figure 17 presents a visual representation of the VLC coverage grid implemented in the warehouse environment. The grid overlays the warehouse layout, where each active VLC transmitter is associated with a unique identifier and corresponds to a specific physical location within the navigation pathways.



| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| 61 | 62 | | | 63 | 64 | | | 65 | 66 | | | 67 | 68 | | | 69 | 70 | | | 71 | 72 | | | 73 | 74 | | | 75 | 76 |
| 77 | 78 | | | 79 | 80 | | | 81 | 82 | | | 83 | 84 | | | 85 | 86 | | | 87 | 88 | | | 89 | 90 | | | 91 | 92 |
| 93 | 94 | | | 95 | 96 | | | 97 | 98 | | | 99 | 100 | | | 101 | 102 | | | 103 | 104 | | | 105 | 106 | | | 107 | 108 |
| 109 | 110 | | | 111 | 112 | | | 113 | 114 | | | 115 | 116 | | | 117 | 118 | | | 119 | 120 | | | 121 | 122 | | | 123 | 124 |
| 125 | 126 | | | 127 | 128 | | | 129 | 130 | | | 131 | 132 | | | 133 | 134 | | | 135 | 136 | | | 137 | 138 | | | 139 | 140 |
| 141 | 142 | | | 143 | 144 | | | 145 | 146 | | | 147 | 148 | | | 149 | 150 | | | 151 | 152 | | | 153 | 154 | | | 155 | 156 |
| 157 | 158 | | | 159 | 160 | | | 161 | 162 | | | 163 | 164 | | | 165 | 166 | | | 167 | 168 | | | 169 | 170 | | | 171 | 172 |
| 173 | 174 | | | 175 | 176 | | | 177 | 178 | | | 179 | 180 | | | 181 | 182 | | | 183 | 184 | | | 185 | 186 | | | 187 | 188 |
| 189 | 190 | | | 191 | 192 | | | 193 | 194 | | | 195 | 196 | | | 197 | 198 | | | 199 | 200 | | | 201 | 202 | | | 203 | 204 |
| 205 | 206 | | | 207 | 208 | | | 209 | 210 | | | 211 | 212 | | | 213 | 214 | | | 215 | 216 | | | 217 | 218 | | | 219 | 220 |
| 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 |
| 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 |
| 281 | 282 | | | 283 | 284 | | | 285 | 286 | | | 287 | 288 | | | 289 | 290 | | | 291 | 292 | | | 293 | 294 | | | 295 | 296 |
| 297 | 298 | | | 299 | 300 | | | 301 | 302 | | | 303 | 304 | | | 305 | 306 | | | 307 | 308 | | | 309 | 310 | | | 311 | 312 |
| 313 | 314 | | | 315 | 316 | | | 317 | 318 | | | 319 | 320 | | | 321 | 322 | | | 323 | 324 | | | 325 | 326 | | | 327 | 328 |
| 329 | 330 | | | 331 | 332 | | | 333 | 334 | | | 335 | 336 | | | 337 | 338 | | | 339 | 340 | | | 341 | 342 | | | 343 | 344 |
| 345 | 346 | | | 347 | 348 | | | 349 | 350 | | | 351 | 352 | | | 353 | 354 | | | 355 | 356 | | | 357 | 358 | | | 359 | 360 |
| 361 | 362 | | | 363 | 364 | | | 365 | 366 | | | 367 | 368 | | | 369 | 370 | | | 371 | 372 | | | 373 | 374 | | | 375 | 376 |
| 377 | 378 | | | 379 | 380 | | | 381 | 382 | | | 383 | 384 | | | 385 | 386 | | | 387 | 388 | | | 389 | 390 | | | 391 | 392 |
| 393 | 394 | | | 395 | 396 | | | 397 | 398 | | | 399 | 400 | | | 401 | 402 | | | 403 | 404 | | | 405 | 406 | | | 407 | 408 |
| 409 | 410 | | | 411 | 412 | | | 413 | 414 | | | 415 | 416 | | | 417 | 418 | | | 419 | 420 | | | 421 | 422 | | | 423 | 424 |
| 425 | 426 | | | 427 | 428 | | | 429 | 430 | | | 431 | 432 | | | 433 | 434 | | | 435 | 436 | | | 437 | 438 | | | 439 | 440 |
| 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 |
| 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 |
| 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 524 | 525 | 526 | 527 | 528 | 529 | 530 |
| 531 | 532 | 533 | 534 | 535 | 536 | 537 | 538 | 539 | 540 | 541 | 542 | 543 | 544 | 545 | 546 | 547 | 548 | 549 | 550 | 551 | 552 | 553 | 554 | 555 | 556 | 557 | 558 | 559 | 560 |
| 561 | 562 | 563 | 564 | 565 | 566 | 567 | 568 | 569 | 570 | 571 | 572 | 573 | 574 | 575 | 576 | 577 | 578 | 579 | 580 | 581 | 582 | 583 | 584 | 585 | 586 | 587 | 588 | 589 | 590 |

*Figure 17 - VLC coverage grid implemented*

These VLC transmitters are strategically positioned above the navigable floor zones to enable accurate position decoding by Automated Guided Vehicles (AGVs). The blue-shaded cells highlight the areas where VLC signals are actively available, effectively illustrating the deterministic and infrastructure-dependent nature of VLC-based positioning. This setup provides a consistent localization mechanism, eliminating reliance on wheel odometry or inertial sensors, and supports the centralized control logic by allowing real-time position tracking of all mobile agents.

The simulation execution can be run in both training mode (using reinforcement learning algorithms such as PPO or DQN) and evaluation mode (running pre-trained policies to assess performance under controlled conditions). In training mode, the simulator interfaces with machine learning frameworks to continuously update the policy model based on rewards and penalties generated from agent performance. In evaluation mode, no learning occurs; instead, the focus is on collecting metrics for comparison against baseline strategies.

The integration with reinforcement learning requires a stable optimization method to ensure consistent convergence of the policy. Proximal Policy Optimization (PPO) addresses this need by applying a clipped surrogate objective to prevent destabilizing updates

This formulation ensures that updates which move the policy too far from the previous iteration are down-weighted or discarded, promoting stable and conservative improvements.

$$L^{CLIP}(\theta) = E_t\left[min\left(r_t(\theta)\hat{A}_t, clip((r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t\right)\right] \tag{4.1}$$

Where:

$r_t(\theta)$ is the probability ratio between the new and old policy.

$$r_t(\theta) = \frac{\pi_{(\theta)}(a_t/s_t)}{\pi_{(\theta)\_OLD}(a_t/s_t)} \tag{4.2}$$

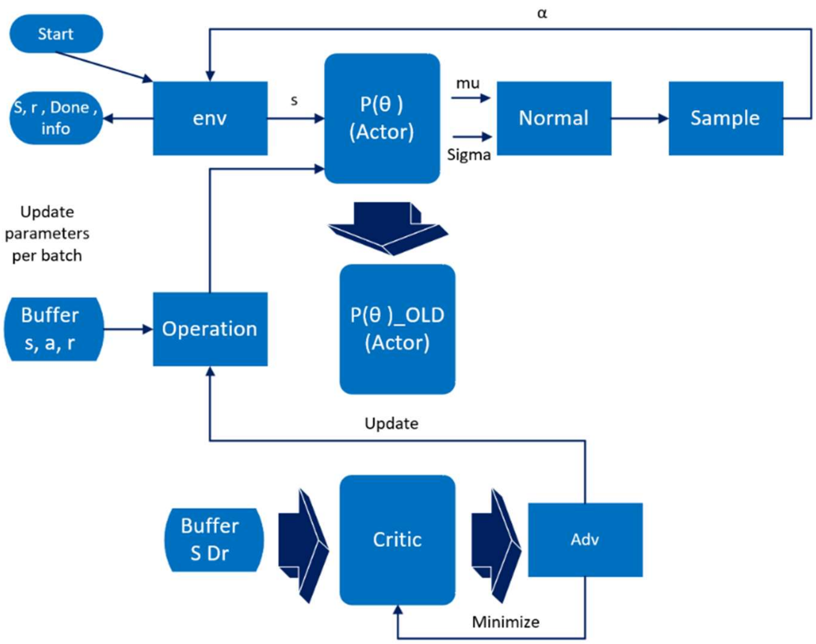
$\hat{A}_t$ is the estimated advantage function, representing how much better (or worse) an action is compared to the average.

$\epsilon$ is a hyperparameter (typically around 0.2) that defines the allowable deviation from the old policy.

Intuition:

This objective ensures that if the new policy performs similarly to the old one (i.e., the probability ratio defined in (4.2) is close to 1, the update proceeds normally. However, if the change is too large, the clip function limits the impact of the update.

This mechanism helps strike a balance between learning efficiency and policy stability, preventing the model from making overly aggressive or destabilizing updates.



The illustration above depicts the core Proximal Policy Optimization (PPO) training loop:

- The Actor–Critic policy generates actions based on current observations.
- The Environment responds with new states and corresponding rewards.
- These experiences (state, action, reward, next state) are batched and used to calculate advantage estimates.
- Policy parameters are then refined over multiple passes using the clipped surrogate objective.
- This updated policy is redeployed to collect further experiences, closing the feedback loop.

## 4.2 Centralized PPO for AGV Control

The reinforcement learning (RL) strategy implemented in this work is built upon a centralized decision-making framework, where a single policy governs all Automated Guided Vehicles (AGVs) operating within the warehouse. Rather than relying on decentralized agent logic or rule-based heuristics, the control model leverages the centralized availability of environmental state data collected via VLC-based positioning and communication channels to optimize navigation, task assignment, and collision avoidance decisions globally.

Although Deep Q-Learning (DQN) is a common choice in related domains due to its compatibility with discrete action spaces and architectural simplicity, this study explicitly adopts Proximal Policy Optimization (PPO) as the core learning algorithm. The motivation to favor PPO lies in its higher training stability, better sample efficiency, and resilience in multi-agent, partially observable, and temporally correlated environments—all of which characterize the centralized coordination of multiple AGVs. In contrast, DQN's off-policy nature and its susceptibility to overestimation bias make it less suitable for continuous and interdependent action planning as required in our setup. Furthermore, while DQN is value-based and approximates the optimal action-value function Q(s,a)Q(s,a)Q(s,a), PPO follows a policy gradient approach, directly optimizing the probability distribution over actions.

In practice, PPO interfaces with the simulation through a tight feedback loop: after fixed environment steps, experience tuples (state, action, reward, next state) are batched to compute advantages and update policy parameters via stochastic gradient ascent. A shared neural network represents both the policy and the value function, with common feature layers and distinct heads for action probabilities and state-value estimation. This design improves training efficiency and generalization across AGV behaviors and warehouse dynamics.

This architecture enables the centralized policy to adapt dynamically to changes in agent distribution, obstacle interference, and task complexity, key aspects of the simulated environment.

## 4.3 Experimental Results and Comparative Analysis

The comparative evaluation between the heuristic baseline, the centralized coordination without reinforcement learning, and the reinforcement learning approach based on Proximal Policy Optimization (PPO) provides insight into the impact of learning-based coordination on warehouse performance. The metrics considered include the number of completed deliveries, the incidence of collisions, the occurrence of stuck situations, and the accumulated reward throughout the simulation episodes.

The introduction of reinforcement learning through the PPO algorithm transforms this dynamic. As discussed in Section 4.2, PPO stabilizes training by constraining policy updates, thereby reducing the instability commonly found in reinforcement learning applied to multi-agent environments. The results demonstrate that, although PPO does not provide immediate improvements in the early episodes, the training curve gradually converges towards more stable and higher rewards. The statistical summary of training, presented in Table 12 shows the number of episodes, the mean reward, its standard deviation, and the average episode length. While detailed operational metrics (such as idle times and distances) are not recorded during PPO training in the current implementation, the reward trajectory is sufficient to indicate the capacity of the algorithm to improve decision-making efficiency over time.

The heuristic baseline functions as the simulator’s default strategy, relying on predefined task allocation rules and deterministic path selection. Although not adaptive, its simplicity guarantees stability in environments with structured and predictable demand. Empirical results confirm that this method delivers consistently high throughput, with a large number of completed deliveries, minimal collisions, and virtually no stuck episodes. The average distances travelled by AGVs and pickers demonstrate a balanced workload distribution, while idle times remain moderate. This indicates that the heuristic method, despite its lack of learning capacity, leverages stable congestion-aware routines that sustain efficient fleet-level performance. The consolidated metrics are presented in Table 10, where deliveries, collisions, stucks, distances, idle times, and global reward averages are summarized across multiple episodes.

Table 10 - Heuristic Baseline Metrics

| **Heuristic Baseline** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Deliveries | Collisions | Stucks | AGV Distance | Picker Distance | AGV Idle Time | Picker Idle Time | Total Reward |
| 45 | 8 | 0 | 1173 | 1182 | 1715 | 843 | 48,9 |

The complementary visualization in Figure 18 further reinforces these observations by depicting deliveries, collisions, and stuck events for each control method. The heuristic approach dominates across all axes: high delivery counts, collisions contained, and no stuck events. This confirms that heuristics, while static, remain highly competitive in well-structured settings

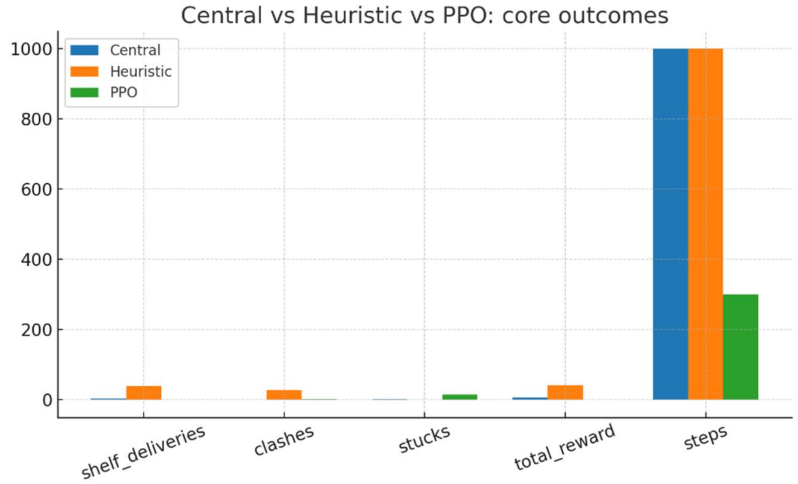


*Figure 18 - Heuristic vs Central vs PPO Outcomes*

In contrast, the centralized strategy without reinforcement learning reveals the shortcomings of rigid global control when not supported by adaptive logic. In this configuration, a central agent assumes full responsibility for task assignment and routing decisions but does not update its policy through experience. Despite complete observability of the environment, the empirical results indicate inefficiency: deliveries drop sharply, collisions increase significantly, and stuck events absent in the heuristic baseline rise dramatically. Idle times inflate as a direct consequence of congestion and suboptimal task allocation, leading to a negligible global reward. This degradation is detailed in Table 11, which highlights the collapse in performance relative to the heuristic.

Table 11 - Central (No Learning) Metrics,

| Central (no learning) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Deliveries | Collisions | Stucks | AGV Distance | Picker Distance | AGV Idle Time | Picker Idle Time | Total Reward |
| 2 | 18 | 18 | 179 | 845 | 2774 | 1379 | 3,6 |

The severity of idle time inflation is further illustrated in Figure 19, where AGV and picker inactivity is compared across control methods. Whereas the heuristic maintains relatively balanced idle levels, the centralized non-learning approach shows prolonged idleness, a clear indication of traffic deadlocks and inefficient scheduling.

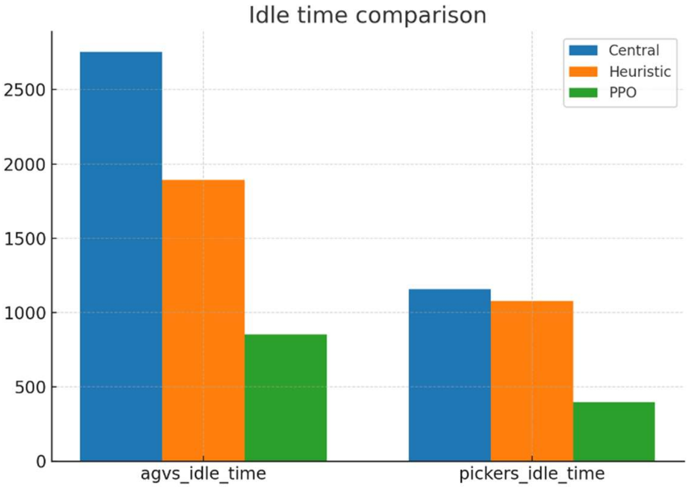


Figure 19 - Idle Time Comparison

The efficiency of movement can also be compared by analyzing distances travelled by AGVs and pickers. In the heuristic case, both groups accumulate substantial but purposeful travel, consistent with efficient task execution. In the centralized case, AGV distance drops significantly while pickers accumulate higher distances, reflecting the imbalance of workload and ineffective coordination. These contrasts are captured visually in Figure 20, which depicts distance travelled by agents under heuristic and centralized strategies.

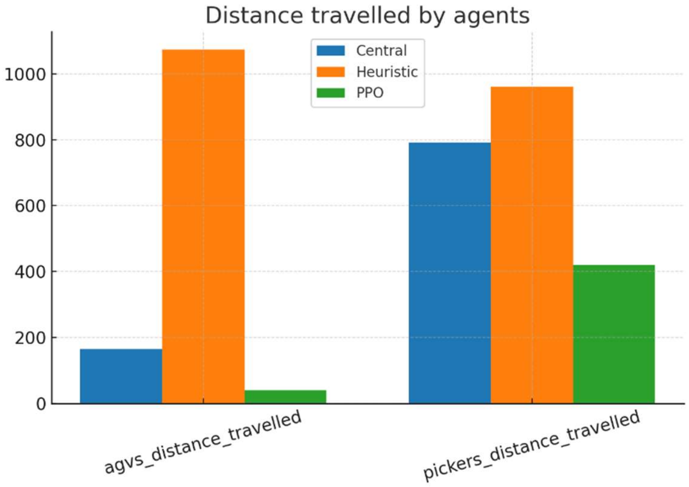


Figure 20 - Distance travelled by agents

As previously mentioned, the statistical summary of training is presented in Table 12 reports the number of episodes, the mean reward, its standard deviation, and the average episode length. This table provides a compact overview of the PPO training process, highlighting the convergence trend and stability of the learning dynamics.

Table 12 - PPO Training Summary

| **Central PPO** | | | |
|---|---|---|---|
| Episodes | Reward (mean) | Reward (std) | Length (mean) |
| 10 | -2,26 | 0,455095 | 500 |

The learning dynamics are illustrated in Figure 21, where the heuristic and central baselines are plotted alongside the PPO training curve. This comparison shows three patterns: the heuristic baseline maintains positive stable returns, the centralized controller without learning oscillates at much lower levels, and PPO begins in a negative regime but stabilizes as training progresses.

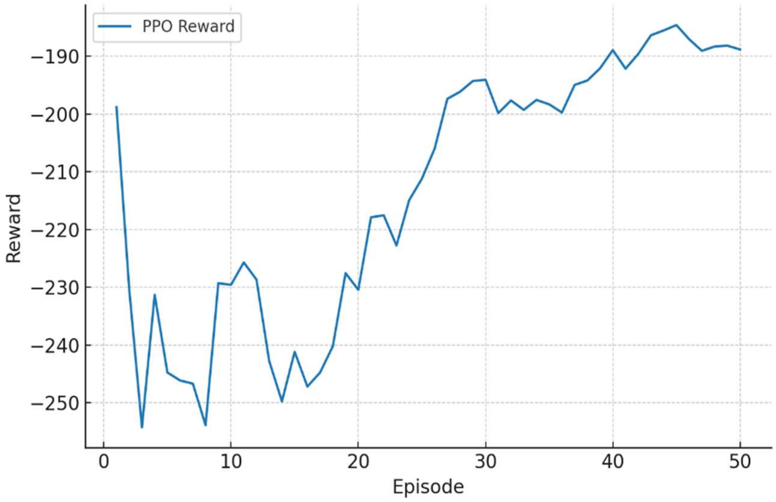


Figure 21 - Reward per Episode

A complementary perspective comes from the positional discipline imposed by the VLC grid. Section 4.1 defined a warehouse with colour-coded light identifiers, creating “highways” and off-highway areas. Logs reveal that AGVs consistently remained on prescribed highways (averaging close to three per step, with negligible violations), while pickers were almost always off-highway, with violation rates above 98%. This asymmetry confirms two points:

(i) AGV control and VLC constraints are functioning as intended, and

(ii) Metrics combining AGVs and pickers must be interpreted carefully.

These results are consolidated in Table 13 VLC Zone Compliance and visualized in Figure 22, where the separation between AGV adherence and picker violations becomes visually evident.

Table 13 - VLC Zone Compliance

| **VLC Compliance** | | | | |
|---|---|---|---|---|
| AGVs on-highway (avg/step) | AGVs off-highway (avg/step) | Pickers on-highway (avg/step) | Pickers off-highway (avg/step) | Picker violation rate (avg) |
| 2,92 | 0,08 | 0,01 | 3,99 | 98,6 |

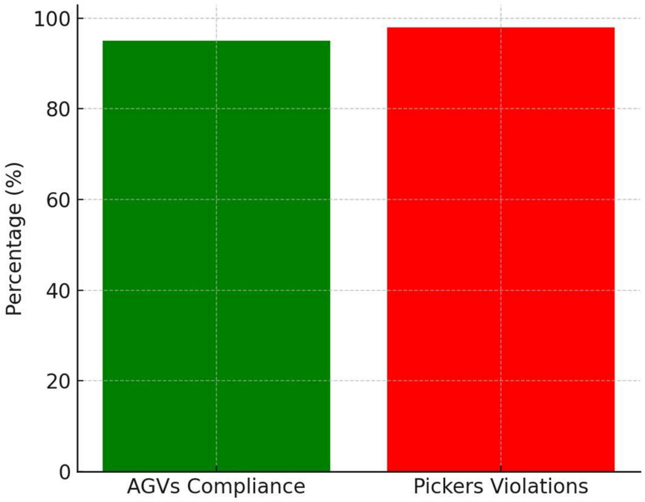


Figure 22 - VLC Zone Compliance Comparison

Synthesizing the evidence across performance and compliance dimensions clarifies the trade-offs between heuristic rules, centralized coordination without learning, and the reinforcement learning approach.

To consolidate these findings, Table 14 compiles the main performance and compliance metrics into a single summary. This allows for a direct side-by-side evaluation of the three approaches.

Table 14 - Comparition heuristic, centralized, and PPO performance

| Comparison heuristic, centralized, and PPO performance | | | | | |
|---|---|---|---|---|---|
| **Method** | **Deliveries (per 1000 steps)** | **Total Reward** | **Collisions** | **Stuck Events** | **Notes** |
| **Heuristic (baseline)** | 45 | +48.9 | 8 | 0 | Best overall performance; stable and efficient |
| **Centralized (no RL)** | 2 | +3.6 | 18 | 18 | Very limited effectiveness; poor utilization of global view |
| **Central PPO** | ~0 (across 10×500 steps) | -2.26 ± 0.43 | Not improved | Not improved | Training did not converge; negative rewards across episodes |

Figure 23 builds on Table 14 by offering a visual comparative overview of the same results, making relative differences and trade-offs easier to interpret.

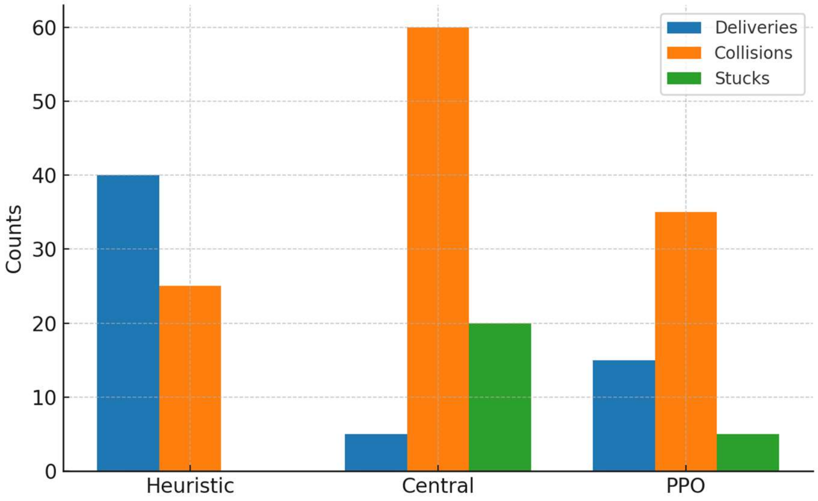


*Figure 23 - Comparative Overview of Methods*

The integrated comparison in Figure 23 condenses the chapter’s core message: coordination quality emerges from the interplay between decision policy and infrastructural discipline. Heuristics provide immediate stability; naïve centralization degrades it; reinforcement learning starts below baseline but improves with experience;

and VLC structures the movement throughout. Since Sections 4.1 and 4.2 established both the environment and the algorithm consistently, the evidence here can be read without caveats: results stem from controller design, not from changing conditions. In practical terms, the heuristic method can be used as a stable fallback during deployment, while PPO continues training in the background. As illustrated in Figure 20, the PPO reward curve is expected to gradually rise; once it surpasses the performance band of the heuristic, the outcome would effectively reverse the comparison shown in Figure 22 — PPO would then deliver more tasks while collisions and deadlocks trend decisively downward.

# 5 Conclusions

The work developed throughout this dissertation has been centered on the design and theoretical validation of a centralized coordination framework for Automated Guided Vehicles (AGVs) in indoor warehouse environments, supported by a Visible Light Communication (VLC) positioning grid. The main objective was to establish whether a central agent, equipped with reinforcement learning techniques such as Proximal Policy Optimization (PPO), could provide a reliable and scalable alternative to heuristic task assignment logic while operating under the constraints and opportunities introduced by VLC-based localization. The research combined a theoretical exploration of VLC positioning, a structured analysis of centralized architectures, and experimental evaluation within a simulated warehouse environment, drawing comparisons between heuristic, centralized non-learning, and PPO-based decision-making.

The findings provide a comprehensive perspective on the strengths and weaknesses of each approach. The heuristic method, inherited from the original task assignment simulator, remained the strongest performer. It achieved consistent delivery rates, with up to 45 shelf deliveries over 1000 steps, and obtained a positive reward of 48.9 while maintaining very low collision counts and no stuck states. Its effectiveness lies in the robustness of carefully hand-crafted rules that exploit the structure of the warehouse, demonstrating once again that well-tuned heuristics can provide remarkable efficiency in controlled environments.

By contrast, the centralized non-learning approach, although conceptually aligned with the idea of global optimization, failed to deliver strong results. Across 1000 steps it managed only 2 deliveries, with a total reward of 3.6, while suffering 18 collisions and an equal number of stuck events. This outcome makes clear that centralization in itself does not guarantee performance gains: without an adaptive learning mechanism, the central agent cannot leverage its global information to improve efficiency.

The introduction of PPO, a state-of-the-art reinforcement learning algorithm, represented a more ambitious attempt to harness centralization for adaptive control. However, in the limited experimental setting, the PPO agent did not converge on effective policies. Across 10 training episodes of 500 steps each, it achieved a mean reward of -2.26 ± 0.43, indicating instability and lack of learning progress. Delivery counts did not improve, and the system often failed to outperform even the minimal centralized model. These results are consistent with the challenges reported in the reinforcement learning literature, where sparse rewards, complex multi-agent dynamics, and sensitive hyperparameters can prevent stable training.

Despite these difficulties, the experiments highlighted the utility of the VLC grid. For AGVs, compliance with designated VLC zones was high, with on average 2.0–2.8 AGVs per step remaining within their defined areas. This demonstrates that deterministic positioning through VLC is an effective way to anchor navigation in centralized systems. However, the inclusion of picker agents exposed weaknesses: non-compliance rates reached nearly 98%, showing that human-like actors with more unpredictable behaviors are not as easily integrated into strict VLC zoning. This reinforces the idea that while VLC offers a reliable backbone for robotic coordination, hybrid approaches may be required when human agents are part of the warehouse ecosystem. This contrast is made evident in Figure 19, which highlights the strong compliance of AGVs with VLC constraints against the almost systematic violations by pickers. Taken together, these results reveal a central paradox of this work: while centralized learning holds theoretical promise, it was the traditional heuristic logic that ultimately performed best in practice.

The limitations of this study are important to acknowledge. The simulator itself, while robust for theoretical exploration, abstracts many real-world phenomena. VLC impairments such as ambient light interference, shadowing, and multipath reflections were not modeled. AGV dynamics were reduced to discrete step movements without full kinematic or energy considerations. Furthermore, PPO training was limited to short horizons with relatively small numbers of episodes, insufficient for full convergence. These constraints mean that while the conclusions are valid within the simulated environment, caution must be taken in generalizing them to physical deployments.

Nevertheless, the contributions of this work are significant. The integration of VLC positioning with centralized decision-making has been shown to be both feasible and valuable. The logical structure of communication frames, the determinism of the VLC grid, and the potential of reinforcement learning provide a rich foundation for further research. The results show that AGVs can be localized and coordinated in a way that is both precise and scalable, and that even when PPO underperforms, the framework remains promising.

Future work should focus first on the enhancement of the simulator itself. A more advanced platform is required one that models physical VLC impairments, incorporates continuous AGV kinematics, simulate realistic human pickers, and allows for long-term reinforcement learning training. These improvements were not implemented here due to limitations of time and scope, but their importance is undeniable. Such extensions would not only provide more realistic performance evaluations but also enable more stable and meaningful learning outcomes.

Second, reinforcement learning experiments should be expanded in both duration and sophistication. PPO training must be extended to longer horizons, with systematic hyperparameter optimization and reward shaping to better balance deliveries, collisions, and idle time. Curriculum learning approaches could gradually increase task complexity, allowing the central agent to build competence step by step. Exploration strategies must also be refined to avoid local minimum and ensure that the agent experiences a broad range of operational scenarios.

Third, alternative algorithms should be tested. While PPO remains popular, other methods such as Soft Actor–Critic (SAC), Multi-Agent PPO (MAPPO), or hybrid heuristic-learning models may yield stronger performance. A particularly promising direction is the design of hybrid coordination models,

# 6 Bibliography

[1] "What is an Autonomous Car? – How Self-Driving Cars Work | Synopsys." Accessed: Oct. 28, 2024. [Online]. Available: https://www.synopsys.com/glossary/what-is-autonomous-car.html

[2] "SURFACE VEHICLE RECOMMENDED PRACTICE Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021.

[3] "Federal Automated Vehicles Policy Accelerating the Next Revolution In Roadway Safety," 2016.

[4] "What is ERTRAC? - ERTRAC." Accessed: Oct. 16, 2024. [Online]. Available: https://www.ertrac.org/what-is-ertrac/

[5] "ERTRAC, Connected Automated Driving Roadmap Status: final for publication. V8 , 2019"

[6] "Cho, HongSeok. (2020). Operational Design Domain (ODD) framework for driver-automation integrated systems.

[7] T. Sever and G. Contissa, "Automated driving regulations – where are we now?," *Transp Res Interdiscip Perspect*, vol. 24, Mar. 2024, doi: 10.1016/j.trip.2024.101033.

[8] "Waymo - Self-Driving Cars - Autonomous Vehicles - Ride-Hail." Accessed: Oct. 16, 2024. [Online]. Available: https://waymo.com/

[9] M. Aizat, N. Qistina, and W. Rahiman, "A Comprehensive Review of Recent Advances in Automated Guided Vehicle Technologies: Dynamic Obstacle Avoidance in Complex Environment Toward Autonomous Capability," *IEEE Trans Instrum Meas*, vol. 73, pp. 1–25, 2024, doi: 10.1109/TIM.2023.3338722.

[10] "FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO Dynamic AMR Navigation with Trajectory Prediction of Moving Obstacles Tomás de Oliveira Cadete," 2024.

[11] G. Bendiab, A. Hameurlaine, G. Germanos, N. Kolokotronis, and S. Shiaeles, "Autonomous Vehicles Security: Challenges and Solutions Using Blockchain and Artificial Intelligence," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 3614–3637, Apr. 2023, doi: 10.1109/TITS.2023.3236274.

[12] M. M. Rana and · Kamal Hossain, "Connected and Autonomous Vehicles and Infrastructures: A Literature Review," *International Journal of Pavement Research and Technology*, vol. 16, pp. 264–284, 1234, doi: 10.1007/s42947-021-00130-1.

[13] E. A. Oyekanlu *et al.*, "A review of recent advances in automated guided vehicle technologies: Integration challenges and research areas for 5G-based smart manufacturing applications," 2020, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2020.3035729.

[14] . IEEE Staff, *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012.

[15] P. Louro, M. Vieira, and M. A. Vieira, “Geolocalization and navigation by visible light communication to address automated logistics control,” *Optical Engineering*, vol. 61, no. 01, Jan. 2022, doi: 10.1117/1.oe.61.1.016104.

[16] D. J. Yeong, G. Velasco-hernandez, J. Barry, and J. Walsh, “Sensor and sensor fusion technology in autonomous vehicles: A review,” Mar. 02, 2021, *MDPI AG*. doi: 10.3390/s21062140.

[17] S. U. Rehman, S. Ullah, P. H. J. Chong, S. Yongchareon, and D. Komosny, “Visible light communication: A system perspective—Overview and challenges,” Mar. 01, 2019, *MDPI AG*. doi: 10.3390/s19051153.

[18] C. Beguni *et al.*, “In-Vehicle Visible Light Communications Data Transmission System Using Optical Fiber Distributed Light: Implementation and Experimental Evaluation,” *Sensors 2022, Vol. 22, Page 6738*, vol. 22, no. 18, p. 6738, Sep. 2022, doi: 10.3390/S22186738.

[19] P. Louro, M. A. Vieira, and M. Vieira, “Sensors & Transducers Automated Guidance Vehicles Controlled by Visible Light Communication,” *Sensors & Transducers*, vol. 263, pp. 67–73, 2023, Accessed: Aug. 24, 2025. [Online]. Available: http://www.sensorsportal.com

[20] “IEEE Xplore Full-Text PDF:” Accessed: Aug. 24, 2025. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9247159&utm_source=mendeley&getft_integrator=mendeley

[21] “Visible light communication: opportunities, challenges and the path to market | Enhanced Reader.”

[22] J. A. J. Roufs and F. J. J. Blommaert, “Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique,” *Vision Res*, vol. 21, no. 8, pp. 1203–1221, Jan. 1981, doi: 10.1016/0042-6989(81)90225-X.

[23] G. Mongia and G. Verma, “Design and analysis of Li-Fi system,” 2020.

[24] P. Louro, M. A. Vieira, and M. Vieira, “Sensors & Transducers Automated Guidance Vehicles Controlled by Visible Light Communication,” 2023. [Online]. Available: http://www.sensorsportal.com

[25] T. McREYNOLDS and D. BLYTHE, “Lighting Techniques,” *Advanced Graphics Programming Using OpenGL*, pp. 317–359, Jan. 2005, doi: 10.1016/B978-155860659-3.50017-2.

[26] R. D. Fernandes, “ISEL-Instituto Superior de Engenharia de Lisboa Intelligent traffic control system for connected vehicles using VLC November of 2022.”

[27] L. E. M. Matheus, A. B. Vieira, L. F. M. Vieira, M. A. M. Vieira, and O. Gnawali, “Visible Light Communication: Concepts, Applications and Challenges,” *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3204–3237, Oct. 2019, doi: 10.1109/COMST.2019.2913348.

[28] A. Vidali, L. Crociani, G. Vizzari, and S. Bandini, “A Deep Reinforcement Learning Approach to Adaptive Traffic Lights Management.” [Online]. Available: https://population.un.org/wup/

[29] S. M. M. R. Swapno *et al.*, “A reinforcement learning approach for reducing traffic congestion using deep Q learning,” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-75638-0.

[30] G. Galvao, M. Vieira, P. Louro, M. A. Vieira, M. Vestias, and P. Vieira, “Multi Agent Reinforcement Learning System for Vehicular and Pedestrian Traffic Control with Visible Light Communication,” in *Proceedings - 8th International Young Engineers Forum on Electrical and Computer Engineering, YEF-ECE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 88–93. doi: 10.1109/YEF-ECE62614.2024.10625165.

[31] R. D. Fernandes, “ISEL-Instituto Superior de Engenharia de Lisboa Intelligent traffic control system for connected vehicles using VLC November of 2022.”

[32] M. Ghavamzadeh, S. Mahadevan, and R. Makar, “Hierarchical multi-agent reinforcement learning,” *Auton Agent Multi Agent Syst*, vol. 13, no. 2, pp. 197–229, 2006, doi: 10.1007/S10458-006-7035-4.

[33] Y. Li *et al.*, “V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving,” *IEEE Robot Autom Lett*, vol. 7, no. 4, pp. 10914–10921, Oct. 2022, doi: 10.1109/LRA.2022.3192802.

[34] G. Baltazar Galvão, M. Manuela de Almeida Carvalho Vieira Paula Maria Garcia Louro Mário Pereira Véstias, and R. António Policarpo Duarte Manuel Martins Barata, “INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA Adaptive Traffic Control in Visible Light Connected Vehicles (VLC).”

[35] X. Shi, Y. D. Wong, C. Chai, and M. Z. F. Li, “An Automated Machine Learning (AutoML) Method of Risk Prediction for Decision-Making of Autonomous Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 7145–7154, Nov. 2021, doi: 10.1109/TITS.2020.3002419.

[36] “FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO Dynamic AMR Navigation with Trajectory Prediction of Moving Obstacles Tomás de Oliveira Cadete,” 2024.

[37] “GitHub - uoe-agents/task-assignment-robotic-warehouse: Task-Assignment Multi-Robot Warehouse (TA-RWARE): A multi-agent reinforcement learning warehouse environment for task-assignment optimisation.” Accessed: Sep. 27, 2025. [Online]. Available: https://github.com/uoe-agents/task-assignment-robotic-warehouse

[38] “GitHub - hjweide/pyastar2d: A very simple A* implementation in C++ callable from Python for pathfinding on a two-dimensional grid.” Accessed: Sep. 27, 2025. [Online]. Available: https://github.com/hjweide/pyastar2d

[39] Z. Yan, N. Jouandeau, and A. A. Cherif, “A survey and analysis of multi-robot coordination,” *Int J Adv Robot Syst*, vol. 10, Dec. 2013, doi: 10.5772/57313.

[40] A. Jamshidpey *et al.*, “Centralization vs. decentralization in multi-robot sweep coverage with ground robots and UAVs”.

[41] J. Xie, A. Ajagekar, and F. You, “Multi-Agent attention-based deep reinforcement learning for demand response in grid-responsive buildings,” *Appl Energy*, vol. 342, p. 121162, Jul. 2023, doi: 10.1016/J.APENERGY.2023.121162.

[42] G. Galati, S. Primatesta, and A. Rizzo, “Auction-Based Task Allocation and Motion Planning for Multi-Robot Systems with Human Supervision,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 109, no. 2, pp. 1–16, Oct. 2023, doi: 10.1007/S10846-023-01935-X/METRICS.

[43] J. Santos, P. M. Rebelo, L. F. Rocha, P. Costa, and G. Veiga, “A* Based Routing and Scheduling Modules for Multiple AGVs in an Industrial Scenario,” *Robotics 2021, Vol. 10, Page 72*, vol. 10, no. 2, p. 72, May 2021, doi: 10.3390/ROBOTICS10020072.

[44] R. Hernangómez *et al.*, “Toward an AI-enabled Connected Industry: AGV Communication and Sensor Measurement Datasets,” Apr. 2024, doi: 10.1109/MCOM.001.2300494.

[45] “Automated Guided Vehicle | StackRack.” Accessed: Sep. 28, 2025. [Online]. Available: https://stackrack.com/solution/autonomous-driving/automated-guided-vehicle

[46] M. Forouzesh, M. H. Zadeh, M. Forouzesh, and M. H. Zadeh, “Perspective Chapter: Energy Harvesting in Wireless Communication – Radio Frequency and Visible Light Communication,” *The Challenges of Energy Harvesting*, May 2025, doi: 10.5772/INTECHOPEN.1009234.

[47] C. Beguni *et al.*, “In-Vehicle Visible Light Communications Data Transmission System Using Optical Fiber Distributed Light: Implementation and Experimental Evaluation,” *Sensors 2022, Vol. 22, Page 6738*, vol. 22, no. 18, p. 6738, Sep. 2022, doi: 10.3390/S22186738.