Scilight

International Journal of Transportation and Logistics Research

https://www.sciltp.com/journals/ijtlr

Article

An On-Demand Autonomous Modular Ridesharing Service

Xi Cheng [1,†], Amir Shafiee [2,†], Hanieh Rastegar Moghaddam [2,†] and Jane Lin [2,*]

[1] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA
[2] Department of Civil, Materials, and Environmental Engineering, University of Illinois Chicago, Chicago, IL 60607, USA
* Correspondence: janelin@uic.edu
† These authors contributed equally to this work.

How To Cite: Cheng, X.; Shafiee, A.; Moghaddam, H.R.; et al. An On-Demand Ridesharing Service with Autonomous Modular Vehicles. *International Journal of Transportation and Logistics Research* **2025**, *1*(1), 2.

Received: 12 May 2025
Revised: 30 May 2025
Accepted: 30 May 2025
Published: 5 June 2025

Abstract: Traffic congestion significantly impacts urban environments, costing billions annually and contributing notably to greenhouse gas emissions and air pollution. Ridesharing services, facilitated by widespread smartphone adoption, have emerged as a promising mobility-on-demand (MoD) solution to alleviate these issues. With the advent of Autonomous Modular Vehicle Technology (AMVT), characterized by autonomy and modularity, these vehicles (also known as pods) can dynamically connect to form a pod train (platooning) and consolidate passengers en-route, potentially enhancing energy efficiency beyond conventional ridesharing. This study introduces and formulates an AMVT-based autonomous modular ridesharing system (AMRS) that employs a shareability hypergraph approach and integer linear programming (ILP) to optimally match passenger requests and minimize energy consumption. Numerical experiments show that AMRS reduces energy consumption by up to 34% compared to non-ridesharing scenarios, with modular coordination contributing incremental yet scalable savings (up to approximately 1.4% additional reduction over basic ridesharing at high demand levels). Sensitivity analysis indicates AMRS benefits are most pronounced in denser networks and with higher numbers of requests, highlighting conditions under which modular operations effectively complement ridesharing efficiency.

Keywords: autonomous modular vehicle technology (AMVT); autonomous modular ridesharing system (AMRS); in-motion joining and disjoining; shareability hypergraph

1. Introduction

Rapid urbanization and growing demand for mobility have intensified transportation challenges globally, leading to increased traffic congestion, vehicle miles traveled (VMT), and greenhouse gas (GHG) emissions [1]. These issues are particularly pronounced in the United States, where congestion-related costs have increased significantly, from approximately $15 billion in 1982 to $190 billion annually by 2019 (in 2020 dollars), as reported by the 2021 Urban Mobility Report [2]. Additionally, transportation is the largest contributor to total GHG emissions in the U.S., accounting for roughly 29% of national emissions [3], with vehicular traffic also recognized as a major source of urban air pollution [4].

These challenges highlight the need for an efficient and sustainable transportation solutions. One such solution is ridesharing, defined as the practice of multiple passengers sharing their rides in the same vehicle either entirely or partially, has emerged as a promising strategy for improving vehicle occupancy, lowering travel costs, and reducing environmental impacts [5]. Originated from informal carpooling, ridesharing has evolved into an increasingly popular Mobility-on-Demand (MoD) service through the integration of mobile and GPS technologies.

In 2016, ridesharing accounted for roughly 1% of total vehicle miles traveled (VMT) in the United States [6], and its usage has rapidly expanded over the past decade. Today, ridesharing forms an important component of MoD systems, complementing public transportation and offering a scalable means of reducing congestion and emissions in urban areas [7]. Globally, the ridesharing market is expected to continue its rapid growth, with a projected compound annual growth rate (CAGR) of 16.3% from 2021 through 2028 [8].

The emergence of Autonomous Modular Vehicle Technology (AMVT) (see, e.g., [9,10]) holds significant promise for accelerating the growth of the MoD market, particularly for ridesharing services. AMVT is distinctive in two combined features—autonomy and modularity—that enable autonomous driving with adjusted capacity in real time with real-time en-route pod joining and separation to better accommodate demand, potentially leading to more efficient utilization of pod capacity and energy savings [11,12]. This study investigates the feasibility of an AMVT based autonomous modular ridesharing system (AMRS) in reducing vehicle trips, VMT, and energy consumption by leveraging pod connections analogous to vehicle platooning. To the best of our knowledge, this research represents one of the earliest studies exploring the application of AMRS.

To this end, this study seeks to answer the question of whether AMRS can improve the energy efficiency compared to traditional MoD models, particularly under high-demand and dense network conditions. AMVT allows pods to join or split en route, and to consolidate passengers, and thus offers new possibilities for saving energy and improving efficiency. The contributions of this work are as follows:

- We propose an AMVT based AMRS with features such as pod joining and passenger consolidation.
- We develop a mathematical framework using a shareability hypergraph and integer programming to match requests while minimizing energy use.
- We quantify how much additional energy can be saved through modular coordination compared to basic ridesharing.
- We run sensitivity analyses to understand when and where AMRS is most effective—for example, in networks of different sizes or with varying pod capacities.

The remainder of this paper is structured as follows: Section 2 reviews recent advances in ridesharing literature, while Section 3 clearly defines the research problem and outlines the system under investigation. Section 4 details our proposed modeling approach and key methodological components. The solution approach is discussed in Section 5, followed by numerical experiments and results in Section 6. Finally, conclusions and future directions are presented in Section 7.

2. Literature Review of Ridesharing

Ridesharing systems have been studied extensively since the concept of ridesharing or carpooling was originated from the oil crisis and energy crisis in 1970s [13]. This literature review focuses on the research efforts in mathematical formulation of ridesharing.

Mathematically speaking, the ridesharing problem is more related to the vehicle routing problem [14,15] and dynamic pickup and delivery problem [16], where demand must be picked up and delivered within predetermined time window [17]. These problems are intractable NP-hard problems. Thus, most previous studies focused on small-scale routing problems [18], such as within airport perimeters [19]. Hsieh et al. [20] addressed the carpooling optimization problem by formulating it as an integer programming model and proposing a Discrete Cooperative Coevolutionary Particle Swarm Optimization (DCCPSO) algorithm. Their approach captured essential operational constraints such as vehicle capacity, time windows, and bid compatibility between drivers and passengers. While the algorithm demonstrated promising results, its evaluation was limited to small-scale scenarios with a restricted number of taxis and passengers, leaving its scalability and applicability to real-world urban systems largely untested. In [21], authors first considered dynamic ridesharing in the context of large urban taxi systems, and designed a system called T-Share to investigate the potential of taxi ridesharing in serving more taxi ride queries. By evaluating a GPS trajectory dataset generated by 33,000 taxis over 3 months with over 10 million queries extracted, the platform T-Share proposed in [21] served additional 25% taxi users than no ridesharing.

In [22], the authors proposed an innovative shareability network method to translate spatio-temporal ridesharing problems into a graph-theoretic framework that provides efficient solutions. Applying this shareability network method to a dataset of 150 million taxi trips in New York City, ref. [22] reported up to 80% of the taxi trips in Manhattan could be shared by two riders, with a delay of a few minutes in the travel time. However, the method in [16] is limited to two riders in one shared trip for the optimal allocation. Ref. [23] extended the work in [22] by proposing a more general mathematical model for high-capacity ridesharing problems. In [23], the authors first generated a request–trip-vehicle graph (RV-graph) considering various time and capacity constraints and then formulated an integer linear programming problem to calculate the optimal assignment plan. Ref. [23] also

validated the model with New York City taxi data and finds that 2000 vehicles (15% of the taxi fleet) of capacity 10 or 3000 of capacity 4 can serve 98% of current demand within a mean waiting time of 2.8 min and mean trip delay of 3.5 min. However, due to computational tractility, ref. [23] set limits on the number of edges in the computation of the RV-graph. In [24], the authors lowered the computational burden in [23] by restricting at most one passenger assigned to a vehicle at each optimization epoch. To further ease the computational burden of large-scale matching, ref. [25] introduced a trip-based uniform graph-partitioning (TUGP) algorithm that decomposes the ride-matching graph into balanced sub-graphs using a shareability-driven dissimilarity measure. By solving the resulting sub-problems in parallel, TUGP delivers near-optimal vehicle-mile savings while cutting run-times by an order of magnitude on New York City taxi data. The authors also demonstrate extensions to one-to-many and role-flexible matching, though the method's effectiveness still depends on the accuracy of its proxy graph and is primarily validated under one-to-one assumptions. Walking is also considered in the ridesharing model as a second mode [26]. Rapid advancements in autonomous vehicle technology have spurred growing interest in shared autonomous platforms, encompassing both passenger ridesharing and automated last-mile package delivery [23,27–29].

Current ridesharing literature focuses on utilizing traditional standalone vehicles (passenger cars or vans) for ridesharing/car/vanpooling. By introducing AMVs (pods) to an MoD ridesharing service, this study makes the following contributions:

(i) We propose an AMV based ridesharing system, which differs from the current ridesharing systems in two operational features: en-route joining of pods to form a pod train (platooning) and en-route consolidation of passengers between pods.
(ii) We formulate a mathematical model to investigate the energy consumption of this AMV based MoD ridesharing system, in particular, the potential energy savings through en-route joining and passenger consolidation.
(iii) We adapt a shareability hypergraph approach to represent sharable rides (requests).
(iv) We devise an Integer Linear Programming (ILP) model to determine an optimal match in the shareability hypergraph that minimizes total energy use.

3. Problem Definition

Consider a company, referred to as A, that employs a fleet of pods with a pod capacity K to provide on-demand ridesharing service. A passenger requests a ride via the company's mobile platform. For simplicity, the pick-up and drop-off locations are pinpointed at the nearest intersections respectively. Suppose that the platform dispatches pods at a pre-determined time interval (e.g., every 30 min) after a set of requests are pooled within the interval. If the requests are open to ridesharing, the platform determines the ridesharing cohorts among those rides before dispatching the pods. In that case, a pod picks up shared-ride passengers at their origins and drops them off accordingly to their destinations to best serve the passengers in the most efficient manner.

So far what is described is a typical MoD ridesharing scheme. What differs in this study is the following two AMV specific capabilities. With AMVT, the pods in operation possess the capability of joining and disjoining en-route, both stationary and in motion. This feature offers potential energy savings similar to platooning. Additionally, if two pods share a portion of the journey leading to a common destination and there is extra capacity in one or both pods, theoretically speaking they can join en-route and consolidate their passengers into one pod. The other pod could be released for new requests, further improving efficiency.

Specifically, these two en-route operations are described as follows:

(i) En-route joining/disjoining for opportunistic pod platooning

As illustrated in Figure 1, if two pods v_1 and v_2 arrive at node A heading to node B within a time window tw, the two pods will be connected as one pod train on arc AB. This operation brings (i) time delay due to joining and (ii) energy savings due to platooning. If the two pods have different destinations, they part (disjoin) at some point down the road and resume two separate pod operations. En-route joining and disjoining may take place when the pods are stationary or in-motion.

(ii) Consolidation of passengers

When two pods v_1 and v_2 are heading to the same final destination D with a shared final portion of the journey, AD, as shown in Figure 2, it is possible to consolidate the passengers of the two pods into one if there is enough remaining capacity in either pod or both. The pod carrying the consolidated passengers will complete the final leg of the journey and the other one will be released for other tasks. Ideally, consolidation of passengers can take place whenever possible during the journey. However, this study is limited to consolidation of passengers on the final shared portion of the journey between the two pods for simplicity.

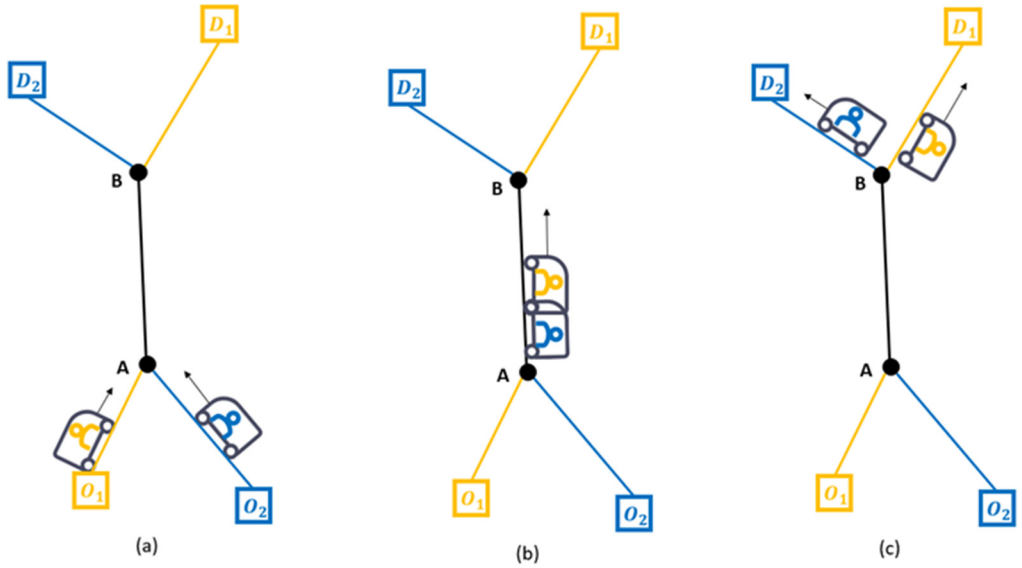

Figure 1. Movements of two pods in join/disjoin en-route: (**a**) the two pods are approaching point A from their respetive paths within a small time window *tw*; (**b**) the two pods join together to form a pod train; and (**c**) the two pods part at point B to go on to their respective destinations.

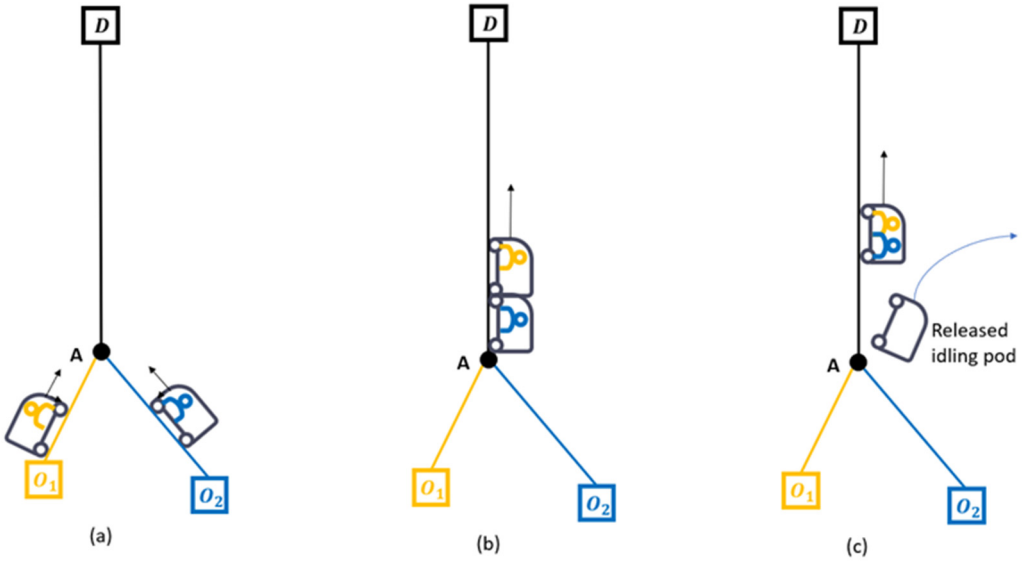

Figure 2. Movements of two pods with consolidation of passengers: (**a**) the two pods are approaching point A from their respetive paths within a small time window *tw*; (**b**) the two pods join together to form a pod train; and (**c**) the passengers will move into one pod if capacity permits and the other pod will be released from the journey.

Let us first define an individual request r as

$$r = \{o_r, d_r, t_r^r, t_r^*, p_r, s_r, \Delta_r, \delta_r, t_r^p, t_r^d\} \tag{1}$$

that contains origin o_r, destination d_r, requested pick-up time t_r^r, the earliest possible drop-off time t_r^*, number of passengers p_r in the party for request r, ridesharing acceptance indicator s_r (1 = yes and 0 = no), maximum tolerable delay Δ_r, and maximum tolerable waiting time for pickup δ_r. All of the above information is provided by the passenger. On the platform end, two other elements are added to r for its record, actual pick-up time t_r^p and actual drop-off time t_r^d.

The current state of a pod v is defined as

$$v = \{q_v, t_v, P_v\} \tag{2}$$

for current location q_v, current time t_v, and existing passengers P_v, respectively.

Consider a traffic network represented by a strongly connected graph $G = (V, A)$. The vertices of the graph are the street intersections. Arcs are the connections (street segments) between vertices. Consider a set of requests $\mathcal{R} = \{r_1, r_2, \ldots, r_m\}$ and a set of pods $\mathcal{V} = \{v_1, v_2, \ldots, v_k\}$ with their current states; $T = \{r_1, r_2, \ldots, r_{n_T}\}$ is a subset

of requests that can be combined and served by a single pod on the same journey. The problem in this study is defined as to assign a set of requests to a set of pods to minimize the total energy consumption of the system. It is hypothesized that the proposed service incurs energy saving from a non-ridesharing counterpart. The potential energy saving stems from ridesharing (through reduced VMT) and from the pod joining and the route consolidation operations as described above.

4. Model Formulation

A static model is considered in this study, where requests are known within a pre-determined time interval and pods are dispatched at the end of each interval. As such, the procedure of the static ridesharing model can be illustrated as follows:

Step 1. During a given time interval Γ, passengers submit trip requests to the ridesharing platform via its mobile app, $\mathcal{R} = \{r_1, r_2, \dots r_p, \dots, r_m\}$ for all m passengers. Each request specifies whether to agree to rideshare, and if so, a maximum tolerable waiting time at pickup δ and a maximum tolerable total delay at dropoff Δ.

Step 2. At the end of Γ, the platform determines the ridesharing plan among the eligible passengers and dispatches pods accordingly to serve both the single- and shared-ride trips.

Step 3. During the journey, a pod v may join/disjoin with one or more other pods en-route to save energy or consolidate its passenger(s) with another pod, as illustrated in Figures 1 and 2.

The key technical element in Step 2 is the identification of sharable requests (or rides). This is accomplished by employing a shareability hypergraph technique to represent sharable rides, which is modified from the shareability graph from [16]. We begin by constructing a shareability hypergraph, in which nodes correspond to individual ride requests and hyperedges signify that connected requests can rideshare. The decision to utilize a hypergraph is informed by the real-world observation that a common trip often encompasses more than just two requests. We then assign weights to the hyperedges. These weights are the energy consumption (kWh) of the sharable rides. We apply integer linear programming (ILP) to find an optimal matching scheme in the shareability hypergraph that minimizes total energy use.

In Step 3, we account for potential energy savings from pod platooning and consolidation. For consolidation, energy saving incurs when passengers in two separate pods are consolidated into one towards the same final destination of both pods. These are formulated into the ILP model.

4.1. Building Shareability Hypergraph (Combining Requests to Shared Ride Trips)

4.1.1. Defining a Shareability Hypergraph

In mathematics, a hypergraph is a generalization of a graph in which an edge can join any number of vertices. In our model, one single request without ridesharing is a trip. Thus, in a more general context, we define a shareability hypergraph $\mathcal{H}$=$(\mathcal{R}, \mathcal{L})$, where $\mathcal{R} = \{r_1, r_2, \dots, r_m\}$ is a set of nodes representing all m requests, and $\mathcal{L}$ is a set of connections of $\mathcal{R}$ called hyperedges. In our definition, each hyperedge, corresponding to one possible ridesharing trip among the connected nodes (requests), may contain a single or two or more requests (nodes). The order of hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ is the number of nodes in $\mathcal{R}$, while the size of the hypergraph is the number of edges in $\mathcal{L}$. A k-uniform hypergraph is a hypergraph such that all its hyperedges have a size of k, i.e., all hyperedges connecting k nodes respectively in the k-uniform hypergraph. Because each pod accommodates up to K passengers, the size of an arbitrary hyperedge in $\mathcal{L}$ is at most K.

All possible trips are a combination of single ride trips—each trip contains only one request, and shared ride trips—each trip contains two or more requests. Thus, a set of all possible trips is defined as $\mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_2 \cup \dots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1}$, where $\mathcal{L}_0 = \{(r_i), \text{for a subset of } i \in \{1,2, \dots, m\}\}$ is a set of single ride requests; $\mathcal{L}_1 = \{(r_i, r_j), \text{for a subset of } i, j \in \{1,2, \dots, m\}\}$ is a set of pairs for ridesharing; and $\mathcal{L}_k$ $(\forall\, k, 1 \le k \le K-1)$, is a set of trips combining $k+1$ requests. Thus, the hyperedge set $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_2 \cup \dots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1}$. We denote the size of $\mathcal{L}$ (i.e., the total number of all possible trips) by $|\mathcal{L}|$, and $|\mathcal{L}| = n$.

4.1.2. Building a 2-Uniform Hypergraph on $\mathcal{R}$

To build a shareability hypergraph $\mathcal{H}$, we first describe the construction of a 2-uniform hypergraph $(\mathcal{R}, \mathcal{L}_1)$ on $\mathcal{R}$ (i.e., a hypergraph in which each hyperedge contains exactly 2 vertices). Given a set of requests $\mathcal{R} = \{r_1, r_2, \dots, r_m\}$, two arbitrary requests $r_i \in \mathcal{R}$ and $r_j \in \mathcal{R}$, defined in Equation (1) can be combined into one trip $(r_i, r_j) \in \mathcal{L}_1$ if and only if the following conditions, called by Φ, are satisfied:

$$t_{r_i}^r \le t_{r_i}^p \le t_{r_i}^r + \delta_{r_i} \tag{3}$$

$$t_{r_j}^r \le t_{r_j}^p \le t_{r_j}^r + \delta_{r_i} \tag{4}$$

$$t_{r_i}^d \le t_{r_i}^* + \Delta_{r_i} \tag{5}$$

$$t_{r_j}^d \le t_{r_j}^* + \Delta_{r_i} \tag{6}$$

$$p_{r_i} + p_{r_j} \le K \tag{7}$$

$$D_{o_i d_i} + D_{o_j d_j} \ge \min\left(D_{o_i o_j d_j d_i}, D_{o_i o_j d_i d_j}, D_{o_j o_i d_i d_j}, D_{o_j o_i d_j d_i}\right) \tag{8}$$

where $D_{o_i d_i}$ denotes the shortest path distance of request r_i between its origin o_i and its destination d_i in the traffic network $G = (V, A)$. When r_i and r_j can be combined into one trip, the order of picking up and dropping off request r_i and r_j must be determined. There are four possible combinations: $o_i o_j d_j d_i$, $o_i o_j d_i d_j$, $o_j o_i d_i d_j$, or $o_j o_i d_j d_i$. The one that yields the shortest distance is the final pickup and drop-off order. For example, if the order combination of $o_i o_j d_j d_i$ yields the minimal distance, then passenger i is picked up first then j, and passenger j is dropped off first followed by i. The arrival time of any node q on path $o_i o_j d_j d_i$ is denoted by $t_q^{(r_i, r_j)}$. If the travel time to the first passenger's location o_i is ignored, then $t_q^{(r_i, r_j)}$ is computed as follows:

$$t_q^{(r_i, r_j)} = \begin{cases} t_{r_i}^p + \dfrac{D_{o_i q}}{u} + \tau + \tau_p \cdot p_{r_i} & \text{if } q \text{ is between } o_i \text{ and } o_j \\ t_{r_i}^p + \dfrac{D_{o_i o_j} + D_{o_j q}}{u} + 2\tau + \tau_p \cdot p_{r_i} + \tau_p \cdot p_{r_j} & \text{if } q \text{ is between } o_j \text{ and } d_j \\ t_{r_i}^p + \dfrac{D_{o_i o_j} + D_{o_j d_j} + D_{d_j q}}{u} + 3\tau + \tau_p \cdot p_{r_i} + 2\tau_p \cdot p_{r_j} & \text{if } q \text{ is between } d_j \text{ and } d_i \end{cases} \tag{9}$$

where τ is the time loss due to stopping for pickup or dropoff, and τ_p is the time loss due to boarding or alighting the pod per passenger, and u is the average cruise speed of pods, assumed constant.

4.1.3. Building a k-Uniform ($2 \le k \le K$) Hypergraph on $\mathcal{R}$

To generalize a 2-uniform hypergraph on $\mathcal{R}$ described in Section 4.1.2, a k-uniform ($\forall\, k, 2 \le k \le K$) hypergraph on $\mathcal{R}$ is denoted as $(\mathcal{R}, \mathcal{L}_1 \cup \ldots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1})$. Given the request set $\mathcal{R} = \{r_1, r_2, \ldots, r_m\}$, any k requests ($2 \le k \le K$) in $\mathcal{R}$ can be combined into one trip $\left(r_{i_1}, r_{i_2}, \ldots, r_{i_k}\right) \in \mathcal{L}_{k-1}$, if and only if the following conditions are satisfied.

- Any two requests in these k requests can be combined into a trip (i.e., any two nodes are in the 2-uniform hypergraph $(\mathcal{R}, \mathcal{L}_1)$).
- Total number of passengers in k requests is less than K.

The pickup and drop-off order of the involved ridesharing passengers, as well as the arrival time at node q of the trip $\left(r_{i_1}, r_{i_2}, \ldots, r_{i_k}\right)$, $t_q^{\left(r_{i_1}, r_{i_2}, \ldots, r_{i_k}\right)}$, can be similarly determined as in Section 4.1.2 Building a 2-uniform hypergraph.

4.1.4. Building a Shareability Hypergraph

The final shareability hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ combines the single ride requests $\mathcal{L}_0 = \{(r_i), \forall\, i \in \{1,2,\ldots,m\}\}$ and all k-uniform hypergraphs ($2 \le k \le K$), $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \ldots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1}$. In other words, $\mathcal{R} = \{r_1, r_2, \ldots, r_m\}$, and $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_2 \cup \ldots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1}$. This shareability hypergraph can be represented in a binary matrix $\mathcal{S}_{|\mathcal{R}| \times |\mathcal{L}|}$ as shown in Figure 3. Recall that the size of $\mathcal{L}$, $|\mathcal{L}|$ is n. Thus, in matrix $\mathcal{S}$, there are m rows and n columns. The element $s_{ij} \in \mathcal{S}_{|\mathcal{R}| \times |\mathcal{L}|}$ ($\forall i \in \{1,2,\ldots,m\}, j \in \{1,2,\ldots,n\}$) is given by

$$s_{ij} = \begin{cases} 1, & \text{if request } r_i \text{ is included in a hyperedge in } \mathcal{L}_j \\ 0, & \text{o.w.} \end{cases} \tag{10}$$

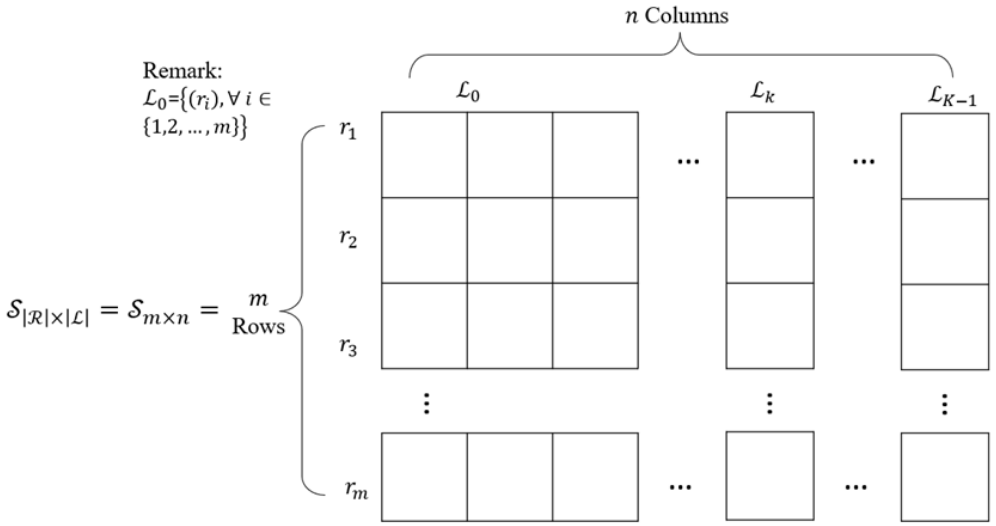

Figure 3. Shareability hypergraph matrix.

4.2. Assigning Weights to Hyperedges including Single Requests

The weights to all the hyperedges $\mathcal{L}$ in hypergraph $\mathcal{H}$ are assigned as follows. The weight of each hyperedge is the energy consumption of the trip (hyperedge). According to the matrix $\mathcal{S}_{|\mathcal{R}|\times|\mathcal{L}|}$, the total number of hyperedges, including all single requests in $\mathcal{L}_0$, in $\mathcal{L}$ is $|\mathcal{L}|$. All hyperedges in $\mathcal{L}$ are ordered from $\mathcal{L}_0$ to $\mathcal{L}_{K-1}$, denoted by T_i $(i \in \{1,2,\dots,|\mathcal{L}|\})$. Thus, for each hyperedge T_i, $i \in \{1,2,\dots,|\mathcal{L}|\}$, the weight assigned to this hyperedge T_i is w_i, given by

$$w_i = c_{T_i} = e \cdot D_{T_i}, \quad \forall\, i, i \in \{1,2,\dots,|\mathcal{L}|\} \tag{11}$$

where c_{T_i} is the total energy consumption (kWh) of the combined trip T_i, which is calculated by multiplying the energy consumption rate per unit distance (e) with the shortest path distance D_{T_i} of trip T_i.

4.3. Examining Potential Energy Savings for Pod Platooning and Consolidation

Examining the potential opportunities for pod platooning and consolidation takes place in the road network $G = (V, A)$ defined in Section 3. The two en-route operations of pods are already described in Section 3 as well. Here, the opportunities for the in-motion joining/disjoining and consolidation must satisfy both the spatial and temporal conditions as follows:

Step 1: Identify the overlapping segments between trips.

Step 2: Check if the arrival time to an overlapping segment is within a small time window tw. If yes, the two pods are candidates for joining. If no, it is temporally infeasible for the two pods to join.

To identify the overlapping segments between two trips, we define a trip (route) as a set of sequential street segments (arcs) on the traffic network $G = (V, A)$, i.e., trip T_i={$[T_i^1, T_i^2], [T_i^2, T_i^3], \dots\dots, [T_i^{m_i-1}, T_i^{m_i}]$}.

4.3.1. Determine the Effective Overlap Segments between Two Trips

We present an algorithm in Algorithm 1 to determine the overlapping segments between any two trips $T_i, T_j \in T$.

Algorithm 1 Algorithm for determining the effective overlapping segments spatially and temporally between two trips T_i, T_j

Input: Initialize $\text{OverlapIntervals}_{\text{ij}} = \{\}, \text{ConsoliOverlapInterval}\ s_{ij} = \{\},\ \forall\, T_i, T_j \in T$

1: Let trip $T_i \in T$ as $\{[T_i^1, T_i^2], [T_i^2, T_i^3], \dots, [T_i^p, T_i^{p+1}], \dots, [T_i^{m_i-1}, T_i^{m_i}]\}$, trip $T_j \in T$ as $\{[T_j^1, T_j^2], [T_j^2, T_j^3], \dots, [T_j^q, T_j^{q+1}], \dots, [T_j^{m_j-1}, T_j^{m_j}]\}$

2: Enumerate each street segment (arc) $[T_i^p, T_i^{p+1}] \in T_i$ in each trip $T_i \in T$

3: Enumerate each street segment (arc) $[T_j^q, T_j^{q+1}] \in T_j$ in trip $T_j \in T$

4: If street segments (arcs) $[T_i^p, T_i^{p+1}]$ and $[T_j^q, T_j^{q+1}]$ overlap, they are the same arc; and if the arrival time at node T_i^p (same as T_j^p) between the two trips $|t_{T_i^p}^{T_i} - t_{T_i^p}^{T_j}| \le tw$:

5: Add street segment $[T_i^p, T_i^{p+1}]$ (i.e., $[T_j^q, T_j^{q+1}]$) to $\text{OverlapIntervals}_{\text{ij}}$.

6: If trip T_i and T_j share the same destination and a common final leg of the journey, and if the arrival time at the final leg $|t_{T_i^p}^{T_i} - t_{T_i^p}^{T_j}| \leq tw$:

7: Add the final common leg $\{[T_i^p, T_i^{p+1}], \dots, [T_i^{m_i-1}, T_i^{m_i}]\}$ to $\text{ConsoliOverlapIntervals}_{\text{ij}}$

Output: list of $\text{OverlapIntervals}_{\text{ij}}$, and $\text{ConsoliOverlapIntervals}_{\text{ij}}$.

4.3.2. Computing Energy Saving From in-Motion Joining/Consolidation between Two Combined Trips

For any two trips T_i and T_j, energy savings from in-motion joining s_{ij} is computed as

$$s_{ij} = \eta \cdot e \cdot \sum \text{OverlapLength}_{\text{ij}} \tag{12}$$

where η is the percentage of energy saving from two pods joining (platooning).

For consolidation, passengers in two pods are consolidated into one pod. Thus, the energy savings from consolidation of passengers of trips T_i and T_j in the last leg of the journey, cs_{ij} is computed as

$$cs_{ij} = e \cdot \sum \text{ConsoliOverlapIntervals}_{\text{ij}} \tag{13}$$

4.4. Examining Potential Energy Savings for Pod Platooning and Consolidation

With the shareability hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ established, a set of all possible trips is $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_2 \cup \dots \cup \mathcal{L}_{K-2} \cup \mathcal{L}_{K-1}$. We call a subset $\mathcal{L}^F (\subset \mathcal{L})$ a set of all feasible ridesharing trips if and only if all requests in $\mathcal{R}$ appear exactly once in $\mathcal{L}^F$. A matching plan of a hypergraph is a set of hyperedges, in which every two hyperedges are disjoint. Thus, a matching plan, denoted $\mathcal{M}$, in hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ is in fact $\mathcal{L}^F (\subset \mathcal{L})$, in which every two hyperedges e_1 and e_2 are disjoint.

We define a set of binary variables χ_i as

$$\chi_i = \begin{cases} 1 & \text{if hyperedge } T_i (i \in \{1,2,\dots,|\mathcal{L}|\}) \text{ is included in a matching plan} \\ 0 & \text{o.w.} \end{cases} \tag{14}$$

The overall objective is to determine an optimistic matching $\mathcal{M}^*$ from the shareability hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$, so that the total energe consumption of all trips in $\mathcal{M}^*$ is minimized. The total energy consumption is composed of two terms: (i) total energy consumed from all trips in $\mathcal{M}^*$ without considering in-motion joining, and (ii) total energy savings due to in-motion joining and consolidation between trips in $\mathcal{M}^*$. Thus, the problem (P) is formulated into an integer linear programming (ILP) as follows.

$$(\mathbf{P}) \text{ minimize } \sum_{\forall i, i \in \{1,2,\dots,|\mathcal{L}|\}} c_{T_i} \cdot \chi_i - \sum_{\forall i, i \in \{1,2,\dots,|\mathcal{L}|\}} \sum_{\forall j, j \in \{i, i+1,\dots,|\mathcal{L}|\}} \chi_i \cdot \chi_j \cdot (s_{ij} + cs_{ij}) \tag{15}$$

$$\text{subject to: } \sum_{\forall j, j \in \{1,2,\dots,|\mathcal{L}|\}} \mathcal{S}_{ij} \cdot \chi_j = 1, \qquad \forall i, i \in \{1,2,\dots,|\mathcal{R}|\} \tag{16}$$

$$\chi_i = \{0,1\}, \qquad \forall i, i \in \{1,2,\dots,|\mathcal{L}|\} \tag{17}$$

5. Solution Approach

In this section, we introduce an algorithmic framework, as illustrated in Algorithm 2, to solve for problem P. Our goal is to find the optimal matching $\mathcal{M}^*$ within the shareability hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ to minimize total energy use. To construct $\mathcal{H}$, we aggregate ride requests into combined trips if they meet the Φ constraint (Equations (3)–(8)). We establish the shortest path between any two nodes in traffic network $G = (V, A)$ using Dijkstra's algorithm, considering all alternatives for the shortest path. Using the constructed hypergraph, we calculate the energy consumption for each trip—the first term in problem P. We also assess potential energy savings, s_{ij} and cs_{ij}, between two trips T_i and T_j. Finally, we solve the ILP problem P using the Gurobi solver.

Algorithm 2 Algorithm for ridesharing with pods

Input: Traffic network $G = (V, A)$, a set of requests $\mathcal{R} = \{r_1, r_2, \dots, r_m\}$, pod speed u, pod capacify K, energy consumed rate e, percentage of energy saving from two pods joining (platooning) η, time loss due to stopping by τ, and time loss from boarding and alighting the pod per passenger τ_p.

Build Shareability Hypergraph as described in Section 4.1
Store the shareability hypergraph $\mathcal{H} = (\mathcal{R}, \mathcal{L})$ in $\mathcal{S}_{|\mathcal{R}|\times|\mathcal{L}|}$
Optimalistic Matching from ILP P
    Compute $c_{T_i} = e \cdot D_{T_i}, \ \forall i, i \in \{1,2,\dots,|\mathcal{L}|\}$
    For any two trips $T_i, T_j \in T, \ \forall i, i \in \{1,2,\dots,|\mathcal{L}|\},, \forall j, j \in \{1,2,\dots,|\mathcal{L}|\}$
        Apply **Algorithm 1** to return $\text{OverlapLength}_{\text{ij}}$ and $\text{ConsoliOverlapIntervals}_{\text{ij}}$
        Compute $s_{ij} = \eta \cdot e \cdot \sum \text{OverlapLength}_{\text{ij}}$
        Compute $cs_{ij} = e \cdot \sum \text{ConsoliOverlapIntervals}_{\text{ij}}$
    Solve for ILP P
        minimize $\sum_{\forall i, i\in\{1,2,\dots,|\mathcal{L}|\}} c_{T_i} \cdot \chi_i - \sum_{\forall i, i\in\{1,2,\dots,|\mathcal{L}|\}} \sum_{\forall j, j\in\{i,i+1,\dots,|\mathcal{L}|\}} \chi_i \cdot \chi_j \cdot (s_{ij} + cs_{ij})$
        subject to: $\sum_{\forall j, j\in\{1,2,\dots,|\mathcal{L}|\}} \mathcal{S}_{ij} \cdot \chi_j = 1, \quad \forall i, i \in \{1,2,\dots,|\mathcal{R}|\}$
        $\chi_i = \{0,1\}, \ \forall i, i \in \{1,2,\dots,|\mathcal{L}|\}$

Output: Optimistic matching $\mathcal{M}^*$

6. Experiments and Results

6.1. Experiment Design

We consider a square $N \times N$ grid network, where each node represents a street intersection, and each edge denotes a connecting road segment. As illustrated in Figure 4, the network is mapped onto a Cartesian coordinate system for clarity. All N^2 nodes are uniquely indexed with consecutive integers from 0 to $N^2 - 1$, following a consistent labeling scheme shown in the Figure 4. The distance between any two adjacent nodes in the network is constant and denoted by λ. Consequently, the shortest path distance between any two nodes i and j ($\forall i, j \in \{0,1,\dots,N^2-1\}$) using the Manhattan distance formula ($D_{ij} = \lambda \cdot (|x_i - x_j| + |y_i - y_j|)$), where (x_i, y_i) represents the Cartesian coordinates of node i. In this grid structure, the coordinates can be derived directly from the node index using the expressions $\text{x}_\text{i} = i \bmod \text{N}$ and $\text{y}_\text{i} = N, \lfloor i/N \rfloor$).

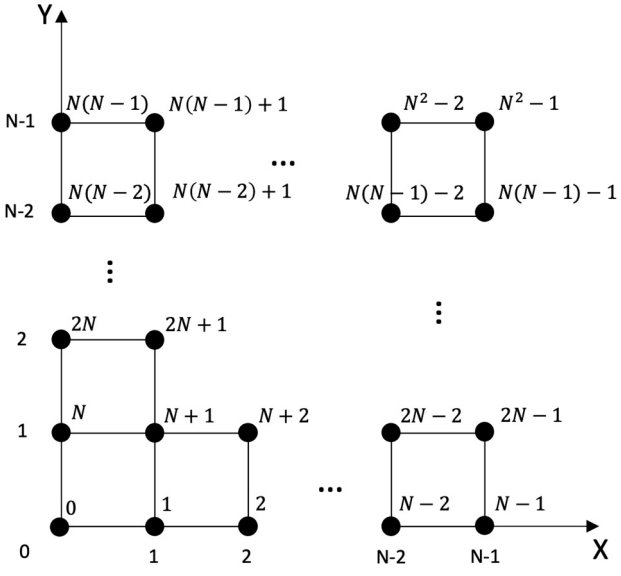

Figure 4. Grid network representation.

A static set of m ride requests is randomly generated at 10-s intervals over a 30-min period on the $N * N$ network. Each request corresponds to a single passenger trip. Requests can be grouped into a single ridesharing trip up to the maximum capacity of the pod. Simulation results are collected over multiple time periods spanning a full 8-h operational window, which includes a 1-h morning peak and a 1-h afternoon peak to capture variation in demand patterns. During peak hours, the request generation rate is increased by 50% to reflect elevated travel demand. To evaluate the system's performance, we examine three distinct scenarios:

- RS0 (No Ridesharing): A baseline scenario in which no ridesharing is allowed; each request is served independently.
- RS1 (Basic Ridesharing): A simple ridesharing model in which requests can be paired, but in-motion joining and route consolidation between pods are not permitted.
- AMRS: The proposed model described in Section 4, which allows for dynamic in-motion joining and route consolidation among modular autonomous vehicles.

In numerical experiments, the grid network size is set to 20 * 20. Figure 5 illustrates an example of 100 randomly generated requests within a 30-min period on this network. In Figure 5, red dots represent the origins of the requests, while green dots indicate their corresponding destinations. Model parameter values used in the simulation are summarized in Table 1. All experiments are implemented in Python 3.8.3 and executed on a MacBook Pro with the Apple M2 Pro chip.

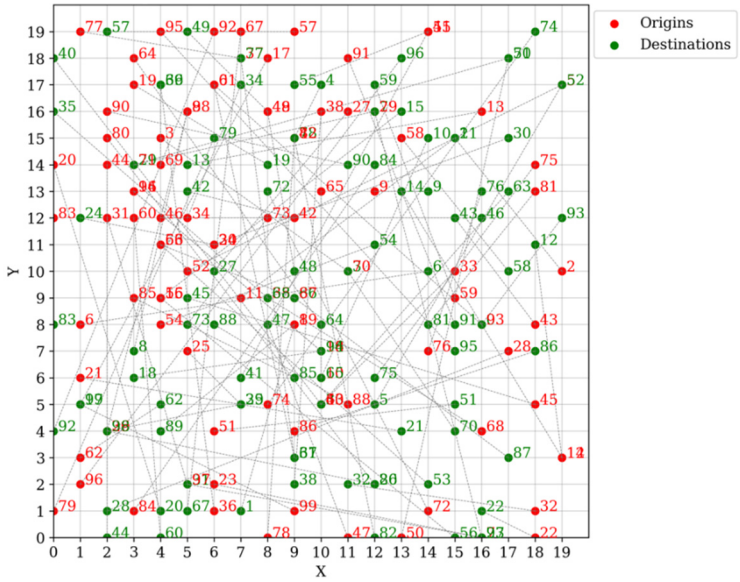

Figure 5. Requests representation on a 20 × 20 network.

Table 1. Parameter values.

Parameter	Description	Value	Source
$N \times N$	Traffic network size (N is the street number).	20 × 20	-
Γ	Pre-determined time interval of analysis	30 min	-
m	Number of requests (single rides in RS0, ridesharing requests in RS1 and AMRS) within T	{2,5,8,10,20,50,80,100}	-
λ	Link length between any two adjacent nodes	0.25 mi	-
K	Pod capacity	3	-
u	Average link cruising speed	25 mph	[30]
τ	Time loss due to deceleration and acceleration at a stop	12 s	[31]
τ_p	Time loss due to boarding/alighting the pod	3 s/passenger	-
e	Energy consumption rate	0.35 kWh/mi	[32]
η	Percentage of energy saving from two pods joining (i.e., platooning)	6.4%	[33]
δ	Maximum tolerable waiting time for pickup (assumed the same for all passengers)	3 min/passenger	-
Δ	Maximum tolerable total delay on the trip (assumed the same for all passengers)	5 min/passenger	-
tw	Time window allowed to realize in-motion joining	30 s	-

6.2. Results

In our numerical experiments, we vary the number of requests across the set {2,5,8,10,20,50,80,100}. For each request size, 10 groups of randomly generated requests are independently solved to ensure result stability. We define the matching rate as the percentage of requests that are successfully grouped into shared rides. As illustrated in Figure 6, the matching rates for both RS1 and AMRS generally improve. However, their relative

performance varies although AMRS aims to maximize route consolidation, it may occasionally result in a lower matching rate than RS1. This occurs when excessive consolidation efforts leave more requests unmatched, especially when the request pool does not align well for efficient grouping.

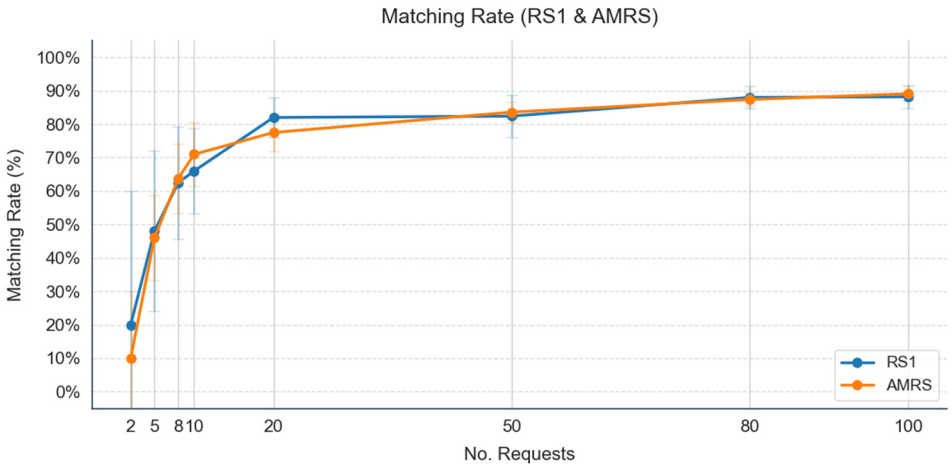

Figure 6. Matching rate of RS1 and AMRS.

To assess the impact of each strategy on energy usage, the energy consumption results are summarized in Table 2, which presents both the average energy usage and the standard deviation across 10 instances per request size. These values are aggregated over a full 8-h operational period, divided into sixteen 30-min horizons. This period includes two 1-h peak intervals, one in the morning and one in the afternoon, during which the request generation rate increases by 50% to simulate elevated demand conditions.

Table 2. Energy consumption comparison among three scenarios across different number of requests**.**

Avg Number of Off-Peak Requests	RS0 (kWh)		RS1 (kWh)		AMRS (kWh)	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
2	2.32	0.8	2.28	0.8	2.27	0.8
5	5.83	1.3	5.38	1.1	5.34	1.1
8	9.32	1.7	8.23	1.2	8.15	1.2
10	11.59	1.9	10.04	1.3	9.92	1.3
20	23.2	2.7	18.79	1.8	18.53	1.8
50	58.22	4.0	42.51	2.3	41.75	2.3
80	93.54	5.2	64.86	2.7	63.59	2.7
100	116.83	6.1	78.75	3.0	77.14	3.0

As shown in Figure 7, RS1 reduces energy consumption by up to 33% compared to the RS0, with greater savings observed as the number of requests increases. Notably, the most rapid improvement in energy efficiency occurs when the number of requests is fewer than 20. Beyond this point, the rate of improvement slows. Moreover, the AMRS model achieves further energy reductions beyond RS1 by allowing for in-motion pod joining and route consolidation. These mechanisms enable multiple requests to be serviced with higher spatial and temporal efficiency when conditions allow. In particular, the benefits of AMRS are most evident when request locations and time windows align in a way that supports effective modular grouping. Although the differences between RS1 and AMRS are modest in percentage terms, they become more substantial at higher demand levels. For instance, AMRS results in approximately a 0.24% reduction in energy use over RS1 when there are only 2 requests per horizon, but this improvement grows steadily, reaching 1.37% at 100 requests. This demonstrates the growing potential of modular coordination to enhance energy efficiency, particularly under high-demand scenarios.

Building on the previous analysis, Table 3 further disaggregates the average energy savings by source. Ridesharing accounts for the majority of energy savings in AMRS, accounts for over 90% of total energy savings in all cases. Meanwhile, additional savings attributable to platooning and consolidation range from 0.24% at 2 requests to 1.37% at 100 requests. Although modest, these gains steadily increase with the number of requests, highlighting the scalable advantage of modular coordination under high-demand conditions. As the number of requests increases, the energy savings attributed to modular operations also grows, reinforcing the increasing impact of AMRS under higher demand scenarios.

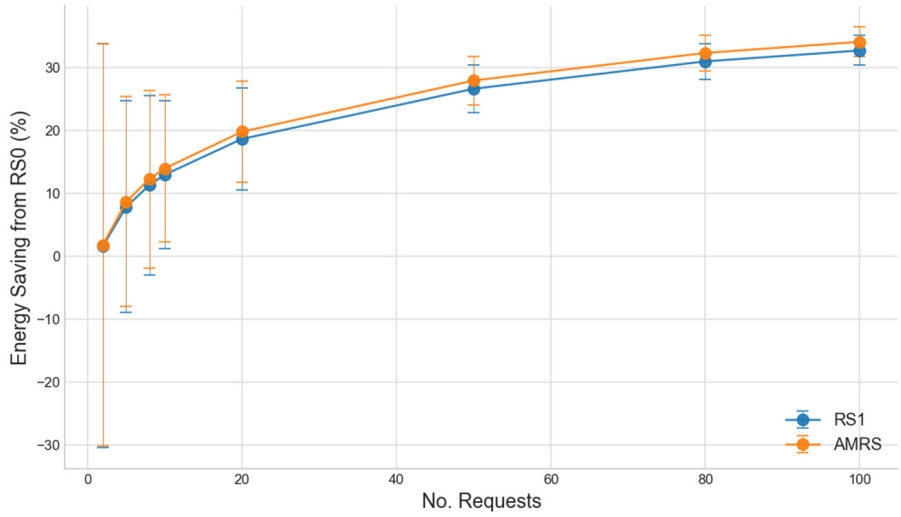

Figure 7. Energy savings of RS1 and AMRS over RS0 WRT No. Requests.

Table 3. Energy saving breakdown in AMRS relative to the baseline RS0 across different number of requests.

Average Number of Off-Peak Requests	Avg. Total Energy Saving of AMRS to RS0 (%)	Avg. Energy Saving Due to Ridesharing (%)	Avg. Additional Energy Saving from Platooning and Consolidation (%)
2	2.15	1.91	0.24
5	8.39	7.67	0.72
8	12.57	11.67	0.90
10	14.41	13.42	0.99
20	20.16	19.01	1.15
50	28.28	26.98	1.31
80	32.01	30.66	1.36
100	33.97	32.60	1.37

6.3. Sensitivity Analysis

To assess the robustness of our findings and identify key influencing factors, we conduct a sensitivity analysis on the network size (N: 10, 15, 20, 25, 30) and pod capacity (2, 3, 4, 5). For each parameter, we vary one setting while keeping the others constant and report the average energy savings under RS0, RS1, and AMRS, along with the corresponding breakdown between ridesharing and modular operations. This analysis is based on a single 30-min planning horizon with 50 randomly generated ride requests, representing a moderate-demand scenario that offers a consistent basis for comparison across all sensitivity cases.

6.3.1. Varying Network Size

As shown in Table 4, increasing the network size from 10 * 10 to 30 * 30 leads to a rise in energy consumption across all scenarios. For instance, RS0 consumption increases from 28.87 kWh to 87.36 kWh, while AMRS consumption grows from 19.92 kWh to 61.87 kWh. Despite this increase in total energy usage, Table 5 shows that the relative energy savings of AMRS compared to RS0 remain relatively stable, ranging between 29.18% and 31%. Ridesharing continues to account for the vast majority of these savings while modular coordination contributes a decreasing share, from 2.58% to 0.89%. These findings suggest that smaller, denser networks facilitate more opportunities for pod coordination due to shorter distances and higher spatial density, enhancing the impact of modular features.

Table 4. Energy consumption comparison among three scenarios Across Different Network Sizes**.**

Network Size	RS0 (kWh)		RS1 (kWh)		AMRS (kWh)	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
10 * 10	28.90	1.8	20.79	1.0	20.04	1.0
15 * 15	43.71	3.1	31.81	1.7	31.04	1.6
20 * 20	57.60	4.0	42.25	2.2	41.50	2.2
25 * 25	72.16	5.5	53.36	3.2	52.60	3.2
30 * 30	87.41	6.0	64.06	3.7	63.29	3.7

Table 5. Energy saving breakdown in AMRS relative to the baseline RS0 Across Different Network Sizes.

Network Size	Avg. Total Energy Saving of AMRS to RS0 (%)	Avg. Energy Saving Due to Ridesharing (%)	Avg. Additional Energy Saving from Platooning and Consolidation (%)
10 * 10	30.65	28.07	2.58
15 * 15	28.99	27.24	1.75
20 * 20	27.95	26.64	1.31
25 * 25	27.11	26.05	1.06
30 * 30	27.60	26.72	0.89

The impact of network size on modular coordination is further illustrated through Figures 8 and 9. Figure 8 shows that although the total energy savings of both RS1 and AMRS slightly decline as the network becomes larger, AMRS consistently provides greater savings than RS1 across all network sizes; however, the gap between them narrows as the network size increases. Figure 9 presents a multilevel radial breakdown of energy savings by network size, where each concentric ring represents a different grid—ranging from 30 × 30 at the center to 10 × 10 on the outermost ring. In these visualizations, the orange segments indicating modular savings are clearly more prominent in smaller networks. This pattern confirms that denser networks allow for more frequent vehicle encounters and better consolidation opportunities, resulting in a higher share of energy savings from platooning and modular coordination.

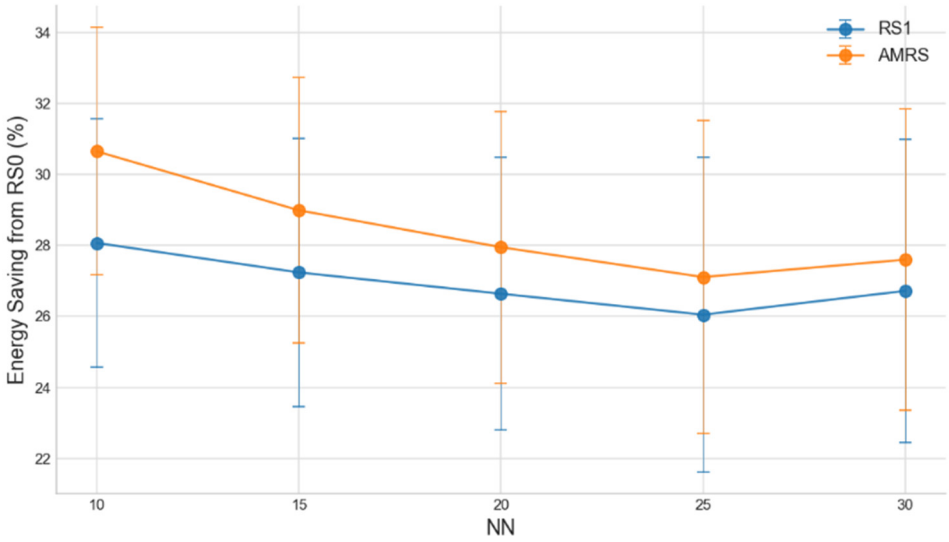

Figure 8. Energy savings of RS1 and AMRS over RS0 WRT network size.

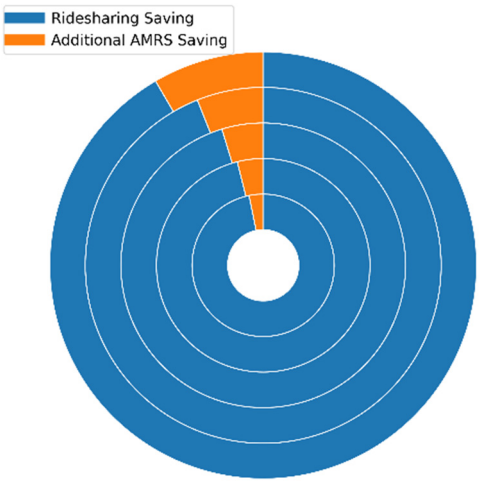

Figure 9. Multilevel Energy Saving Breakdown by pod network size innermost 30 * 30, Outermost 10 * 10.

6.3.2. Varying Pod Capacity

Larger pod capacities allow more requests to be grouped together, which improves energy efficiency across all system configurations. As shown in Table 6, increasing the pod capacity from 2 to 5 reduces AMRS energy consumption from 43.95 kWh to 41.91 kWh, resulting in higher total energy savings—from 24.4% to 28.33% relative to RS0.

Table 6. Energy consumption comparison among three scenarios across different pod capacity.

Pod Capacity	RS0 (kWh)		RS1 (kWh)		AMRS (kWh)	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
2	58.14	4.0	44.96	2.3	43.95	2.3
3	57.60	4.0	42.25	2.2	41.50	2.2
4	57.58	4.1	42.01	2.3	41.35	2.2
5	58.47	3.9	42.50	1.9	41.91	1.9

Table 7 provides further insight into the breakdown of these savings. Ridesharing (RS1) benefits significantly from higher capacities, with energy savings increasing from 22.68% to 27.33%. However, the additional savings attributable to AMRS decline from 1.73% at pod capacity 2 to just 1.00% at capacity 5. This pattern aligns with expected system behavior. As pod capacity increases while demand remains fixed, it becomes easier to match and group passengers, making RS1 more effective. Since RS1 alone can already capture most of the potential energy savings through ridesharing, there is less room for AMRS to deliver additional improvements.

Table 7. Energy saving breakdown in AMRS relative to the baseline RS0 across different pod capacity.

Pod Capacity	Avg. Total Energy Saving of AMRS to RS0 (%)	Avg. Energy Saving Due to Ridesharing (%)	Avg. Additional Energy Saving from Platooning and Consolidation (%)
2	24.40	22.68	1.73
3	27.95	26.64	1.31
4	28.18	27.04	1.14
5	28.33	27.33	1.00

The trends in energy savings with increasing pod capacity are also clearly illustrated in Figures 10 and 11. Figure 10 demonstrates that as pod capacity grows, both RS1 and AMRS yield greater energy savings due to improved ridesharing efficiency. However, the gap between them narrows, reflecting the reduced marginal benefit of modular operations when more requests can already be grouped in a single pod. Figure 11, a multilevel radial visualization, reinforces this observation: the share of energy savings from platooning and consolidation (orange segments) is most prominent at lower capacities (e.g., pod capacity 2 in the outermost ring) and steadily diminishes toward the center as capacity increases to 5.

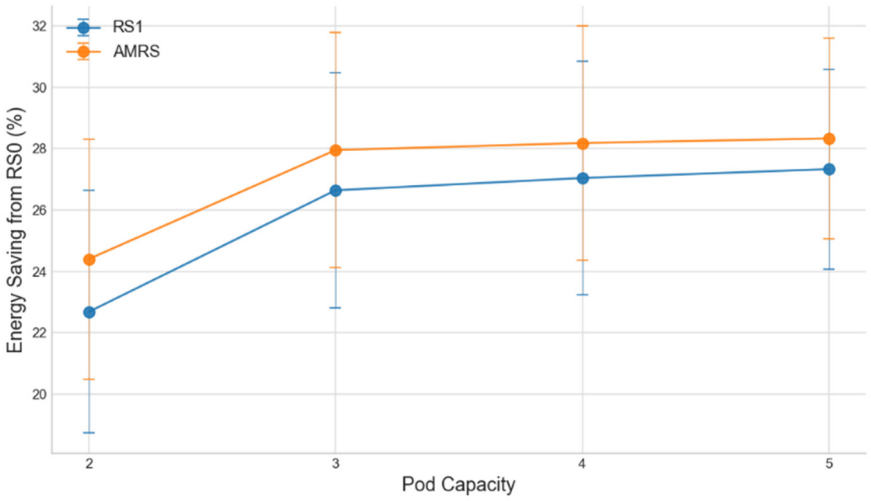

Figure 10. Energy savings of RS1 and AMRS over RS0 WRT Pod Capacity.

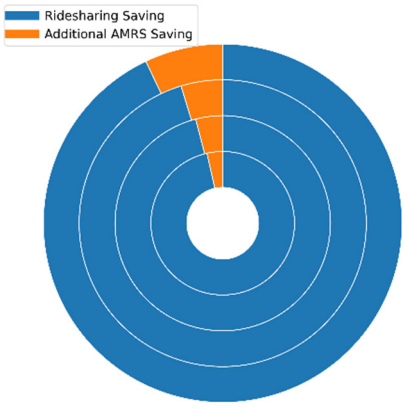

Figure 11. Multilevel Energy Saving Breakdown by pod capacity Innermost 5, Outermost 2.

The results demonstrate benefits of AMRS, especially in reducing energy consumption relative to traditional and basic ridesharing models. Importantly, while the incremental gains from modular platooning and passenger consolidation (around 1–2%) seem modest in isolation, their cumulative impact could be considerable at the urban scale, particularly under high demand or dense network conditions. Our sensitivity analyses highlight the scenarios under which AMRS performs best. For instance, smaller network grids (e.g., 10×10) and moderate pod capacities (3–4 passengers) provided optimal conditions for frequent vehicle interactions and effective passenger consolidation. This insight suggests that urban planners and mobility providers aiming to maximize AMRS benefits might prioritize deployment in central, densely populated zones or during peak commuting hours when demand densities are highest. Additionally, while energy savings are a primary benefit explored here, the operational efficiencies from AMRS could also translate into substantial reductions in urban congestion and emissions—further amplifying societal benefits. However, actual realization of these advantages depends heavily on practical factors such as passenger acceptance, realistic operational constraints, and regulatory frameworks to address safety and reliability concerns inherent to autonomous systems.

7. Conclusions

This study developed and tested an AMRS to enhance traditional ridesharing efficiency. A novel mathematical model based on a shareability hypergraph approach was proposed and solved using integer linear programming, incorporating unique AMVT operations such as in-motion pod joining (platooning) and passenger consolidation.

Numerical analyses demonstrated significant energy savings, with AMRS achieving approximately 28–34% reductions in energy consumption compared to scenarios without ridesharing. Although additional savings specifically attributable to modular coordination (such as platooning and passenger consolidation) averaged around 1%, these benefits increased notably under favorable operational conditions. For example, modular operations accounted for up to 2.58% of additional energy savings in a dense 10×10 network, and 1.73% when pod capacity was limited to two passengers, a scenario that increases the number of pods on the road and thus raises opportunities for joining and consolidation. Furthermore, as customer demand grew, the incremental savings from modular coordination steadily rose—from just 0.24% at the lowest tested demand level (two requests) to approximately 1.37% at higher demand (100 requests). These results highlight the scalability and strategic advantage of modular operations, emphasizing their effectiveness in enhancing ridesharing performance under higher demand and tighter capacity constraints.

In practice, the findings suggest several insights for transportation providers and policymakers considering AMRS adoption. For instance, AMRS is most beneficial in dense urban areas with moderate-to-high passenger demand (e.g., above 20 requests per dispatch horizon), as these conditions maximize modular operations and energy savings. Operationally, maintaining short time-windows (around 30 seconds for pod joining) and pod capacities of 3 to 4 passengers provides an effective balance between efficiency and passenger convenience. From a policy perspective, supporting modular ridesharing through incentives or dedicated infrastructure (such as lanes for modular platooning) could reduce vehicle miles traveled, energy usage, and urban congestion. However, real-world deployment must carefully evaluate upfront and ongoing implementation costs, including vehicle acquisition, infrastructure retrofitting, operational management, and maintenance. Additionally, due to the autonomous nature of AMVT, safety and reliability become paramount. Adequate safety standards, rigorous testing protocols, and clear regulatory frameworks must be established to ensure public acceptance and safe operations. Policymakers must proactively develop or update legislation governing autonomous vehicles, addressing issues such as liability, data privacy, and cybersecurity, to enable safe and reliable deployment at scale.

A comprehensive cost-benefit and risk analysis, encompassing economic, operational, and regulatory dimensions, is therefore crucial before widespread adoption.

Despite mentioned results, the study has several limitations. The model adopted a static approach, where requests are assumed to be known at the start of each horizon, and each horizon is solved independently without considering interactions across horizons. This discrete-interval approach may limit practical applicability in fully dynamic, real-world scenarios, where requests continuously arrive, and real-time adjustments are essential. Future research could explore dynamic modeling approaches, real-time decision-making, and the incorporation of more realistic constraints (e.g., varying speeds, traffic conditions, dynamic passenger preferences). Expanding the analysis to heterogeneous pod capacities, multimodal integration (e.g., walking), and broader environmental impacts beyond energy use, such as emissions and cost analysis, would further enhance understanding of AMVT's potential role in urban mobility solutions.

Author Contributions

J.L. and X.C.: conceptualization, methodology, software; X.C..: data curation, writing—original draft preparation; X.C., A.S. and H.R.M.: visualization, investigation; J.L.: supervision; A.S. and H.R.M.: software, validation; J.L. and A.S.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Science Foundation grant number CMMI 2127677.

Data Availability Statement

The data used or generated in this study will be made publicly available after publication via the University of Illinois Chicago data repository.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ngo, N.S.; Zou, Z.; Yang, Y.; et al. The impact of urban form on the relationship between vehicle miles traveled and air pollution. *Transp. Res. Interdiscip. Perspect.* **2024**, *28*, 101288. https://doi.org/10.1016/j.trip.2024.101288.
2. Texas A&M Transportation Institute. 2021 Urban Mobility Report. Available online: http://tti.tamu.edu/documents/mobility-report-2021.pdf (accessed on 10 May 2025).
3. United States Environmental Protection Agency, Fast Facts: U.S. Transportation Sector GHG Emissions 1990–2021. Available online: https://www.epa.gov/system/files/documents/2023-06/420f23016.pdf (accessed on 10 May 2025).
4. Caiazzo, F.; Ashok, A.; Waitz, I.A.; et al. Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005. *Atmos. Environ.* **2013**, *79*, 198–208. https://doi.org/10.1016/j.atmosenv.2013.05.081.
5. Morency, C. The ambivalence of ridesharing. *Transportation* **2007**, *34*, 239–253. https://doi.org/10.1007/s11116-006-9101-9.
6. McKinsey & Company. Cracks in the Ridesharing Market—And How to Fill Them. Available online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/cracks-in-the-ridesharing-market-and-how-to-fill-them#0 (accessed on 10 May 2025).
7. Furuhata, M.; Dessouky, M.; Ordóñez, F.; et al. Ridesharing: The state-of-the-art and future directions. *Transp. Res. Part B Methodol.* **2013**, *57*, 28–46. https://doi.org/10.1016/j.trb.2013.08.012.
8. Fortune Business Insights. Ride Sharing Market Size, Share & COVID-19 Impact Analysis. Available online: https://www.fortunebusinessinsights.com/ride-sharing-market-103336 (accessed on 10 May 2025).
9. NEXT Future Transportation—Modular Autonomous Vehicles. Available online: https://www.next-future-mobility.com/ (accessed on 10 May 2025).
10. CNN. Dubai's Futuristic Transport Concepts. Available online: https://www.cnn.com/2017/11/29/middleeast/gallery/dubai-future-transport/index.html (accessed on 10 May 2025).
11. Lin, J.; Nie, Y.M.; Kawamura, K. An Autonomous Modular Mobility Paradigm. *IEEE Intell. Transp. Syst. Mag.* **2023**, *15*, 378–386. https://doi.org/10.1109/MITS.2022.3159484.

12. Shafiee, A.; Moghaddam, H.R.; Lin, J. Using Autonomous Modular Vehicle Technology as an Alternative for Last-Mile Delivery. In Proceedings of the 2024 Forum for Innovative Sustainable Transportation Systems (FISTS*)*, Riverside, CA, USA, 26–28 February 2024; pp. 1–6. https://doi.org/10.1109/FISTS60717.2024.10485532.
13. Handke, V.; Jonuschat, H. *Flexible Ridesharing*; Springer: Berlin/Heidelberg, Germany, 2013.
14. Shawe-Taylor, J.; Cristianini, N. *Kernel Methods for Pattern Analysis*; Cambridge University Press: Cambridge, UK, 2004.
15. Toth, P.D. Vehicle Routing. *Public Transp.* **2014**, *1*, 6–7.
16. Berbeglia, G.; Cordeau, J.F.; Laporte, G. Dynamic pickup and delivery problems. *Eur. J. Oper. Res.* **2010**, *202*, 8–15. https://doi.org/10.1016/j.ejor.2009.04.024.
17. Mahmoudi, M.; Zhou, X. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state-space-time network representations. *Transp. Res. Part B Methodol.* **2016**, *89*, 19–42. https://doi.org/10.1016/j.trb.2016.03.009.
18. Cordeau, J.F. A branch-and-cut algorithm for the dial-a-ride problem. *Oper. Res.* **2006**, *54*, 573–586. https://doi.org/10.1287/opre.1060.0283.
19. Marín, Á.G. Airport management: Taxi planning. *Ann. Oper. Res.* **2006**, *143*, 191–202. https://doi.org/10.1007/s10479-006-7381-2.
20. Hsieh, F.S.; Zhan, F.M.; Guo, Y.H. A solution methodology for carpooling systems based on double auctions and cooperative coevolutionary particle swarms. *Appl. Intell.* **2019**, *49*, 741–763. https://doi.org/10.1007/s10489-018-1288-x.
21. Ma, S.; Zheng, Y.; Wolfson, O. T-share: A large-scale dynamic taxi ridesharing service. In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, QLD, Australia, 8–12 April 2013; pp. 410–421. https://doi.org/10.1109/ICDE.2013.6544843.
22. Santi, P.; Resta, G.; Szell, M.; et al. Quantifying the benefits of vehicle pooling with shareability networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13290–13294. https://doi.org/10.1073/pnas.1403657111.
23. Alonso-Mora, J.; Samaranayake, S.; Wallar, A.; et al. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 462–467. https://doi.org/10.1073/pnas.1611675114.
24. Simonetto, A.; Monteil, J.; Gambella, C. Real-time city-scale ridesharing via linear assignment problems. *Transp. Res. Part C Emerg. Technol.* **2019**, *101*, 208–232. https://doi.org/10.1016/j.trc.2019.01.019.
25. Tafreshian, A.; Masoud, N. Trip-based graph partitioning in dynamic ridesharing. *Transp. Res. Part C Emerg. Technol.* **2020**, *114*, 532–553. https://doi.org/10.1016/j.trc.2020.02.008.
26. Lin, J.; Sasidharan, S.; Ma, S.; et al. A model of multimodal ridesharing and its analysis. In Proceedings of the 2016 17th IEEE International Conference on Mobile Data Management (MDM), Porto, Portugal, 13–16 June 2016; pp. 164–173. https://doi.org/10.1109/MDM.2016.34.
27. Ma, S.; Zheng, Y.; Wolfson, O. Real-Time City-Scale Taxi Ridesharing. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 1782–1795. https://doi.org/10.1109/TKDE.2014.2334313.
28. Martinez, J.M.; M-Correia, L.; HA-Viegas, G. An agent-based simulation model to assess the impacts of introducing a shared-taxi system: An application to Lisbon (Portugal). *J. Adv. Transp.* **2015**, *49*, 475–495. https://doi.org/10.1002/atr.
29. Shafiee, A.; Asgharpour, S.; Askari, S.; et al. Understanding Characteristics of Crowdshipping Trip Production: Evidence from Atlanta. In Proceedings of the International Conference on Transportation and Development 2024, Atlanta, Georgia, 15–18 June 2024; pp. 62–71. https://doi.org/10.1061/9780784485521.006.
30. Guarino, M. Uber, Lyft Worsen Traffic in Chicago's Loop. Available online: https://www.chicagobusiness.com/article/20180316/ISSUE01/180319945/uber-lyft-worsen-traffic-in-chicago-s-loop (accessed on 10 May 2025).
31. Chen, P.W.; Nie, Y.M. Analysis of an idealized system of demand adaptive paired-line hybrid transit. *Transp. Res. Part B Methodol.* **2017**, *102*, 38–54. https://doi.org/10.1016/j.trb.2017.05.004.
32. U.S. Department of Energy. eGallon Methodology. Available online: https://www.energy.gov/sites/prod/files/2013/06/f1/eGallon-methodology-final.pdf (accessed on 10 May 2025).
33. Lammert, M.P.; Duran, A.; Diez, J.; et al. *Effect of Platooning on Fuel Consumption of Class 8 Vehicles Over a Range of Speeds, Following Distances, and Mass 2014-01-2438*; SAE International: Warrendale, PA, USA, 2014. https://doi.org/10.4271/2014-01-2438.