Transportation Research Part C 191 (2026) 105829

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A network flow optimisation model for operational planning of modular autonomous public transport services with in-motion transfers

Neema Nassir[a,*], Andres Fielbaum[b], Sapan Tiwari[c], Avishai (Avi) Ceder[d]

[a] University of Melbourne, Australia
[b] University of Sydney, Australia
[c] Centre for Urban Research, RMIT University, Melbourne, Australia
[d] Technion Israel Institute of Technology, Israel

ARTICLE INFO

Keywords:
Modular autonomous vehicles (MAV)
In-motion-transfer (IMT)
Stop-less service
Vehicle scheduling
Minimum cost flow (MCF)
Time expanded network

ABSTRACT

Modular Autonomous Vehicles (MAVs) enable public transport services in which small pods couple into platoons on trunk segments and decouple to serve stops or diverging branches. This supports in-motion transfers (IMT) between intersecting platoons while passengers remain onboard, creating transferless connectivity in many-to-many origin-destination networks. Such flexibility can deliver substantial passenger benefits by combining high-frequency trunk performance with expanded coverage on branches. However, it also introduces operational challenges not present in conventional systems, including the need to coordinate pod exchanges at junctions, model decouple-dwell-recouple behaviour at stops, and maintain feasible platoon sizes across space-time.

We develop a network flow mathematical formulation and network transformation for vehicle scheduling and fleet sizing of MAV network services. We adopt a discrete time-space representation and construct a time-expanded network with two new transformations: (i) junction-coupling links that synchronise left, through, and right pod swaps when concurrent platoons meet, and (ii) stop transformations that represent decouple-dwell-recouple movements enabling stop-less operations while the through-platoon continues uninterrupted. Embedding these transformations in the time-expanded network allows all operational requirements to be represented as link flows with integer bounds. The resulting optimisation model is a minimum-cost circulation problem that minimises fleet size (primary objective) and non-revenue movements (secondary objective), subject to timetable feasibility, platoon-size bounds, and depot and junction capacities. Repositioning is represented explicitly through deadhead links with non-zero costs. The proposed formulation is linear, totally-unimodular (due to integer constraints), and highly scalable.

The proposed framework supports evaluation of capital and operating requirements for alternative service designs. We demonstrate its effectiveness on a small public-transport network with modular vehicles. We extend the case study network in Khan and Menéndez, 2024 to accommodate asymmetric service, where different routes or directions operate at unequal frequencies. This generalisation enables the evaluation of heterogeneous corridor patterns and directional peak services. Through a set of case study scenarios, we demonstrate how the model quantifies service

* Corresponding author.
E-mail address: neema.nassir@unimelb.edu.au (N. Nassir).

https://doi.org/10.1016/j.trc.2026.105829
Received 28 November 2025; Received in revised form 11 June 2026; Accepted 19 June 2026
Available online 1 July 2026

performance and cost-effectiveness metrics, including revenue hours (or kilometres) per vehicle hours (kilometres), and the resulting fleet requirements under alternative asymmetric service patterns.

1. Introduction

Modular autonomous vehicles (MAVs) represent a promising and emerging technology for public transport (PT) systems. Instead of relying on a single large vehicle, MAVs consist of multiple smaller modules (pods) that can couple to form a platoon on trunk segments and decouple to operate independently on branches or at stops. This capability increases operational flexibility and spatial coverage by benefitting from the network effect that pod coupling and decoupling can create. Although not yet implemented at scale, the underlying technology has been demonstrated. For instance, NEXT Future Transportation Inc. (2018) developed MAVs capable of in-motion transfers (IMT). IMT, a key feature of MAVs (Wu et al., 2021), allows a pod to recouple between intersecting platoons while passengers can walk freely among pods in the platoon, enabling seamless route changes without alighting. Because transfers are often perceived as a barrier to PT use (Jara-Diaz et al., 2022), eliminating platform-based transfers can improve perceived service quality and potentially boost ridership.

Beyond a single corridor, the value of modularity unfolds at the network level. With IMT-enabled recoupling at scheduled meet points, multiple origin stops get connected to multiple destinations through the IMT network effect. Consequently, modular services can deliver network connectivity without stops at intermediate locations.

Such a network effect scales rapidly with the number of covered branches. With n branches feeding into IMT-enabled trunks, the effective connectivity of the system grows on the order of n^2. In other words, each origin branch can be linked to every other destination branch through modular coupling and in-motion transfers, without requiring platform-based transfers or intermediate delays/dwell at stops. This combinatorial growth in connections highlights how modularity and IMT can transform a corridor service into a highly connected network (see Fig. 1).

Akin to a network of ski lifts, as different lifts coordinate to move skiers across slopes with minimal waiting, modular pods couple into fast platoons on trunks, decouple to serve branches and stops, and recouple at scheduled meets. The same features that make the service attractive also make it hard to deliver at scale; each trip may require multiple pods acting in concert. Pods must swap between intersecting platoons at junctions, stop operations require precise decouple-dwell-recouple timing while the main platoon continues, and depots/terminals must supply and absorb pods under corridor, junction, and berth capacities.

1.1. Conventional service and operational planning decisions

Fixed-schedule PT systems are typically planned and operated through a hierarchical and sequential decision-making framework that reflects both academic practice and the standard procedures adopted by transit agencies in real-world applications (Ceder, 2016, Desaulniers and Hickman, 2007). Within this context, it is common practice to distinguish between service planning and operational planning, which address two complementary but distinct objectives.

Service planning decisions focus on defining the service as experienced by passengers. It includes the stages of network and route design, stop locations, service frequency setting, and timetable development. These stages determine the spatial and temporal structure of the transit system, including route layouts, service coverage, headways, and departure times. Importantly, these decisions are explicitly driven (as displayed in Fig. 2) by passenger demand information, load profiles, infrastructure capacity, and service policies. Transit agencies rely on demand data, historical usage patterns, and policy targets (e.g., coverage standards, minimum service levels, maximum on-board congestion levels, etc.) to design services that maximise utilisation and accessibility.

Once the service plan is established, operational planning decisions address how to efficiently deliver the designed service using available resources (vehicles and crew). This includes vehicle scheduling and crew scheduling, which are typically formulated as downstream optimisation problems aimed at minimising operational (and fleet) costs while ensuring service feasibility (Desaulniers and Hickman, 2007; Bunte and Kliewer, 2009).

A key feature of current industry practice is that these decisions are made sequentially rather than jointly. While integrated optimisation of service and operations is theoretically desirable, transit agencies typically adopt a “design-then-operate” approach; first developing a service plan that adequately reflects demand, and then optimising its delivery (Ceder, 2016). As a result, passenger demand is incorporated upstream in service planning, and is not treated as a direct input in the operational planning stage, where it is already implicitly embedded in the timetable and service specifications.

This study adopts the same decomposition. The focus is on the development of an operational modelling framework for MAV network-based service with IMT, including vehicle scheduling, coupling and decoupling processes, and the associated network representation. Modelling demand explicitly would require additional assumptions on demand estimation, assignment, and behavioural response, which would substantially expand the scope of the study and detract from the primary objective. By treating demand as embedded in the given service specification, the analysis isolates the operational problem and enables a clear examination of the feasibility and efficiency of operations.

While this decomposition is consistent with established practice, it is important to recognise its implications for the interpretation of results. In particular, the estimated fleet size and operational feasibility obtained from the proposed model are conditional on the adequacy of the service plan in representing underlying passenger demand. In this framework, capacity is specified explicitly through the number of pods assigned to each service movement, including through movements, stop operations, and turning flows at junctions.

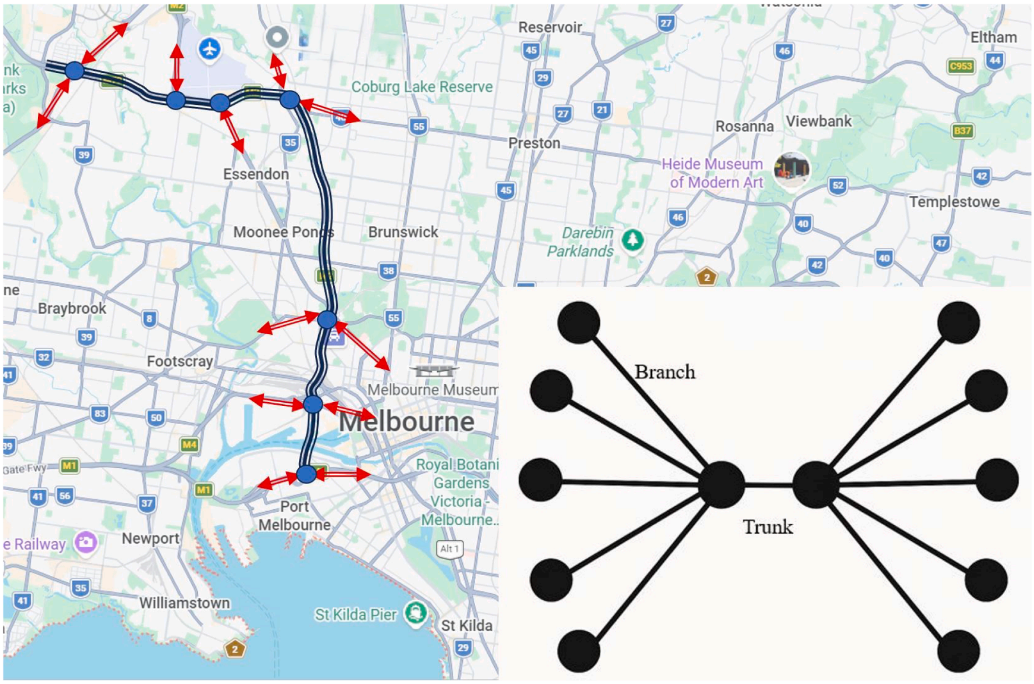

Fig. 1. Illustrative example of network connectivity with IMT.

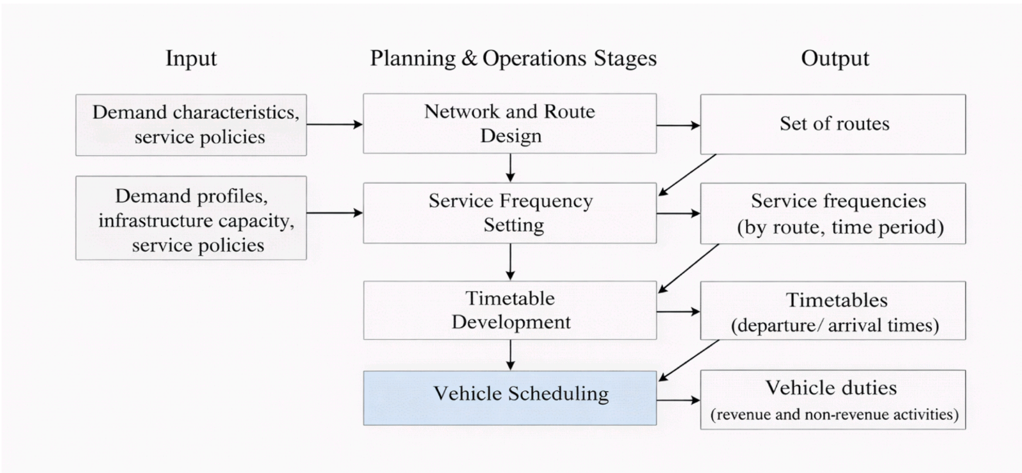

Fig. 2. Scope of planning decisions for the proposed MAV service.

These capacity values are assumed to be determined based on conservative estimates to ensure that passenger demand is accommodated. If demand levels deviate from those implicitly embedded in the service plan, adjustments to these capacity specifications-such as changes in platoon sizes or service frequencies-may be required, which could affect the resulting fleet requirements. Therefore, the proposed framework is best interpreted as an operational evaluation tool for a given service design, while extensions that explicitly integrate demand estimation and over-crowding constitute an important direction for future research.

1.2. Operational planning for MAV service networks

The proposed model assumes that the service plan details, comprising routes, frequencies, and timetables, along with number of pods allocated to each part of service are given, and planning decisions concentrate on optimising vehicle movements and fleet utilisation subject to operational constraints. Given the context of modular autonomous vehicle (MAV) systems, crew scheduling is not relevant, as vehicle operations would not require human drivers.

We cast the MAV operational decisions (fleet size and vehicle routing) as a deterministic minimum cost flow problem on a time-expanded network with junction-coupling and stop transformations that represent pod manoeuvres in space-time.

The resulting minimum-cost circulation delivers all timetabled trips and capacities as inputs, while minimising fleet size (primary) and non-revenue movements (secondary) subject to platoon-size requirements, turning limits, and depot/corridor capacities. The solution returns executable pod itineraries and platoon profiles. In the operational model, the number of pods performing each manoeuvre (e.g., the left-turn pod and the right-turn pod) is not chosen by the optimiser but is specified directly in the service plan and enforced through link-level minimum-flow constraints.

This paper makes the following contributions:

1. We develop the first network-scale scheduling method specifically designed for modular PT with IMT. The method embeds junction-coupling and stop transformations into a time-expanded network so that coupling, decoupling, and recoupling are directly represented in space-time.
2. We cast the planning problem as a minimum-cost flow circulation that minimises fleet size (primary) and non-revenue movements (secondary), subject to timetable feasibility, per-trip platoon requirements, depot and corridor capacities, and operational limits on coupling events. Repositioning is represented explicitly via deadhead links with nonzero costs.
3. We apply the framework to the general grid network introduced by Khan and Menéndez (2024), relaxing several uniformity assumptions (e.g., symmetric frequencies) to study asymmetric, directionally peaked services and their implications for synchronisation and fleet requirements.
4. The optimiser returns executable module itineraries and platoon plans, enabling planners to compare alternative designs (service layout, headways, platoon and depot capacities, etc.) before implementation.

Collectively, these elements move modular PT from concept toward an implementable operational plan, linking passenger-facing benefits of IMT with the backend planning and scheduling decisions required to deliver them at network scale.

This paper focuses on grid networks as a structured and analytically tractable setting. The proposed framework is general and can be extended to complex network topologies with more asymmetric configurations.

2. Literature review

2.1. Modular autonomous vehicles

Research on the operational benefits of MAV has gained attention in recent years, focusing on both single-route and network-wide PT systems (Dai et al., 2020; Tang et al., 2024; Shi and Li, 2021). Several early MAV concepts have been proposed and tested in practice. One of the first was developed by Next Future Transportation Inc., which introduced Patented Self-driving Pods (PSP) capable of autonomously connecting and disconnecting during operations (Gecchelin and Spera, 2017, 2022). Prototype demonstrations of these pods were conducted in Dubai in 2018 (Fahmy, 2018), showcasing the feasibility of such systems in real-world conditions. A second notable initiative is the Dynamic Autonomous Road Transit (DART) system, designed by the research platform TUMCREATE. This concept, based on modular autonomous mini-buses, has been evaluated extensively through deployment planning and simulation studies of urban networks in Singapore (Liu et al., 2018; Rau et al., 2019; Sun et al., 2020). A third concept, the Duo-Module Units (DMU) developed at the German Aerospace Center (DLR), consists of vehicles with a driveboard and interchangeable passenger capsules, allowing flexible composition and capacity adjustments during operation (Friedrich et al., 2019; Ulrich et al., 2019). Beyond passenger transport, MAVs have been applied to freight and bi-modal systems integrating passenger-freight operations (Shafiee et al., 2024; Shi et al., 2024; Hatzenbühler et al., 2023; Lin et al., 2022). Studies also highlight energy and environmental advantages of MAV platoons over conventional fleets (Yi et al., 2025; Zhang et al., 2022, 2024). At the microscopic scale, research has addressed in-motion docking and transfers, including trajectory optimisation for pods and passengers at intersections and docking in mixed traffic environments (Wu et al., 2021; Li and Li, 2022; Han et al., 2024; Ye and Yamamoto, 2024). More recent work has further extended MAV research toward real-time operational control, adaptive scheduling, and safety assurance, where rolling-horizon and learning-based approaches enable dynamic adjustment of vehicle formations and service schedules under fluctuating demand, while adaptive control strategies ensure safe docking and undocking operations in uncertain environments (Tian et al., 2025; Ma et al., 2025).

Recent studies highlight MAVs' ability to adapt capacity to fluctuating demand. Guo et al. (2018) first examined flexible feeder services with modular pod combinations, while Chen et al. (2019, 2020) optimised shuttle convoys of up to five pods under oversaturation. Later works introduced station-wise docking and boarding-priority rules to better utilise capacity (Chen and Li, 2021; Chen et al., 2022; Shi and Li, 2021), considered bidirectional and rebalancing strategies (Liu et al., 2023; Tian et al., 2023), and applied skip-stop docking to reduce waiting (Zhang et al., 2024). Most recently, Luo et al. (2025) proposed an optimal design model for Modular Feeder Transit (MFT), where MAV platoons decouple into individual vehicles for door-to-door feeder services and recouple for return trips, demonstrating through analytical and numerical studies that MFT can yield significant cost savings compared to traditional fixed-capacity feeder systems. Recent work further shows that MAVs can also be integrated with conventional bus systems through hybrid operational strategies, where modular units complement fixed-route services to improve passenger assignment, fleet utilization, and sustainability performance while reducing operating and environmental costs (Huang et al., 2025). Subsequent studies further demonstrate that modular operations benefit from jointly optimizing timetables, vehicle formations, and route layouts under time-dependent and uncertain demand, enabling reductions in passenger waiting time, empty-seat operations, and overall operational costs (Zhao et al., 2026; Cong et al., 2025; Yao et al., 2025). Complementary analytical frameworks also show that corridor-based heterogeneous transit systems using modular vehicles can be designed using continuum approximation methods, enabling scalable optimization of station spacing, service frequencies, and vehicle coupling strategies for large networks (Luo et al., 2026).

Research on IMT has primarily aimed to reduce or eliminate passengers' out-of-vehicle transfers and the associated delays. Zhang et al. (2020) were among the first to incorporate IMT into modular transit design. They formulated a vehicle routing problem involving two types of pods-main pods and trailer pods-within a time-expanded network, allowing both vehicle and rider trajectories to be optimized simultaneously. This model demonstrated how riders can transfer between pods en route, but it did not explicitly restrict the number of transfers, raising concerns about excessive inconvenience. To address this limitation, Caros and Chow (2021) restricted riders to a maximum of two IMTs and explored the impact of modular operations across different service types, including door-to-door urban mobility, commuter first/last mile services, and hub-and-spoke networks. Their approach used an insertion heuristic to dynamically identify vehicle routes and IMT events. Recent simulation studies show that MAVs can reduce vehicle demand and costs in FMLM PT while maintaining low passenger waiting times (Bao et al., 2025). Similarly, Fu and Chow (2023) reformulated the dial-a-ride problem (DARP) to account for IMT in modular systems. A heuristic neighbourhood search algorithm was developed to solve large-scale instances, demonstrating the scalability of such models. In a related study, Zermasli et al. (2023) integrated IMT within the feeder bus network design problem by embedding synchronization constraints, highlighting the potential of IMT in feeder-trunk systems.

Beyond these general formulations, some studies have concentrated on single-line modular services. Romea and Estrada (2021) and Khan and Menendez (2022) independently introduced the concept of stopping-less operations based on IMT. In their design, the main convoy bypasses intermediate stops while only a tail pod detaches to board or alight passengers, requiring users to pre-position themselves in the detaching module. This approach substantially increases commercial speed, making modular buses competitive with private cars. Romea and Estrada (2021) provided analytical optimization of stop spacing, headways, and vehicle composition, while Khan and Menendez (2022) showed that stopping-less operations help mitigate bus bunching, as the convoy is unaffected by individual pod dwell times. Building on this, Khan et al. (2023) and Khan and Menéndez (2023) proposed splitting and holding strategies, in which IMT is leveraged before modules split for skip-stop operations, further stabilising headways. More recently, Lin et al. (2024) and Liu et al. (2024) extended this research by optimising passenger assignments among pods to alleviate modular bus bunching. Wang et al. (2025a) further analysed autonomous-vehicle timetabling using coupling and decoupling tactics. Their work demonstrates how modular buses can synchronise entry and exit of vehicle units along corridors to optimise headway adherence and operational efficiency, reinforcing the relevance of controlled coupling operations for network-level MAV services. Recent extensions also show that adaptive module coupling and stopping-less strategies can significantly reduce both operator costs and passenger travel times in urban deployments, while network-wide fleet allocation and module sharing across lines can further reduce required fleet sizes and improve occupancy levels (Filippi et al., 2025; Torres et al., 2025).

IMT strategies have also been embedded into network design. Khan and Menéndez (2024) applied stopping-less IMT operations to grid-based modular transit networks. Their designs employed uniform decision rules for stop spacing, line spacing, and headways. By contrast, Cheng et al. (2024) focused on fixed-route and fixed-stop systems, incorporating IMT to allow a share of passengers to transfer within vehicles without disembarking, thus blending traditional design with modular flexibility. More recent work further extends modular vehicle deployment to hub-and-spoke PT systems, jointly optimizing routing, scheduling, and vehicle repositioning decisions to reduce passenger travel times and operational costs under heterogeneous demand conditions (Wang et al., 2025b).

Other scholars have considered infrastructure-specific implementations. Tian et al. (2022) limited in-motion docking and un-docking to specially designated stations while treating other stops as conventional bus halts. They formulated a location problem to determine optimal station placement and capacities for storing detached pods. This hybrid approach balances the operational advantages of IMT with realistic infrastructure requirements, offering a more pragmatic pathway for implementation. More recent studies have begun to address real-time operations of modular-vehicle transit systems. For example, Tian et al. (2025) propose a rolling horizon control framework combined with a learning-based optimisation proxy to dynamically adjust vehicle schedules and formations in response to time-varying demand. These approaches highlight the potential of integrating real-time data and adaptive control into modular transit operations. In contrast, the present study focuses on the offline operational planning problem, providing a tractable network-scale formulation for vehicle scheduling and fleet sizing under a given service plan.

2.2. Multi-depot vehicle scheduling problem

In PT planning, the vehicle scheduling problem (VSP), particularly multi-depot VSP (MDVSP), has been extensively studied to optimize vehicle duties and minimize costs. Traditionally, the VSP aims to assign a fleet of vehicles to a set of predefined trips while minimizing total non-revenue movements such as deadheading, pull-out/pull-in operations, and idle times. While the Single Depot VSP (SDVSP) assumes a homogeneous fleet operating from a single depot, real-world applications often involve MDVSP that introduce added complexities including heterogeneous fleets, depot capacity constraints, and inter-depot consistency requirements. The SDVSP can be solved in polynomial time using the minimum cost flow algorithm. Bertossi et al. (1987) established that the MDVSP is NP-hard, indicating that it is likely that no polynomial-time solution exists.

Various exact and heuristic approaches have been proposed to address the MDVSP. Early contributions include the branch-and-bound algorithm of Carpaneto et al. (1989), later improved by Fischetti et al. (2001) with a branch-and-cut procedure. One of the most widely used approaches is the time-space network formulation (Kliewer et al., 2006), which discretises time into events representing trip start and end times. While this formulation can become large in terms of variables, it can be solved with branch-and-price algorithms. Kulkarni et al. (2018) proposed an inventory-based variation that also uses a time-space network, solving the LP relaxation with column generation and applying heuristics to obtain high-quality integer solutions. Several exact methods have been proposed to tackle real-world instances of the MDVSP, including Column Generation (CG) combined with Lagrangian Relaxation (Löbel, 1999), as well as hybrid approaches that integrate CG with Branch-and-Bound techniques (Hadjar et al., 2006).

In addition, various heuristic strategies have been explored, such as schedule-first/cluster-second methods, shortest-path-based heuristics, and iterated local search (ILS) algorithms (Daduna and Pinto Paixão, 1995; Laurent and Hao, 2009). A comparative study by Pepin et al. (2009) found that truncated CG was the most effective when extended computation times were permissible, while Large Neighbourhood Search (LNS) provided fast, high-quality solutions under time constraints.

Research on MDVSP has evolved to include network flow approaches, which model vehicle movements through time-space networks (Bunte and Kliewer, 2009; Kliewer et al., 2006; Kulkarni et al., 2018). These methods connect compatible trips through time and space, reducing the problem size and computational effort while maintaining robust solutions (Bunte and Kliewer, 2009). Kliewer et al. (2006) demonstrated how time-space networks could be applied to solve MDVSP by linking trips via nodes representing stops and depots over time.

Further developments in network flow-based MDVSP formulations have explored alternative representations and solution enhancements to improve scalability and solution quality. The multi-commodity flow (MCF) model introduced by Desrosiers et al. (1995) treats vehicles from different depots as distinct commodities, explicitly incorporating depot-specific constraints within the network structure. Although the MCF model often results in a cubic number of variables with respect to the number of trips and depots, subsequent work by Hadjar et al. (2006) examined its polyhedral properties and introduced strategies such as reduced-cost fixing and branch-and-cut algorithms with odd-cycle cuts to tighten the formulation and accelerate convergence.

Computational efficiency has also been improved through arc reduction techniques, which limit the inclusion of deadhead arcs to those within a specified travel time threshold or to arcs directly adjacent to service trips. These methods, applied in both connection-based and time-space formulations, can significantly reduce model size without sacrificing optimality (Bunte and Kliewer, 2009; Kliewer et al., 2006). Other studies have investigated aggregation of trips or time intervals to reduce the number of nodes in the network, and decomposition strategies, such as column generation, to handle very large-scale problem instances (Kulkarni et al., 2018). More recent studies have further extended MDVSP research toward reliability-aware and more flexible formulations. Ricard et al. (2024) incorporated stochastic travel times into the MDVSP to improve schedule reliability, highlighting the importance of robustness in real-world transit operations. Similarly, van Lieshout and van der Schaft (2025) developed a dynamic discretization discovery algorithm for the MDVSP, enabling exact solution of large-scale instances without explicitly enumerating all possible shifted trips. In addition, recent work has extended MDVSP formulations to emerging transit technologies, such as battery-electric bus systems, where vehicle scheduling must be jointly optimized with charging coordination under infrastructure and energy constraints (Jia and An, 2025).

The integration of MAVs introduces a new layer of complexity to the scheduling problem. Unlike conventional vehicles, MAVs involve not only the assignment of trips but also the coordination of pods that may split, merge, and move throughout the network. This modularity creates additional challenges, such as determining optimal platoon configurations, synchronizing pod attachments and detachments. These complexities push the limits of traditional MDVSP methodologies and make it necessary to develop new optimisation frameworks that can address the unique characteristics of MAV operations. In this paper, we focus on operational planning decisions for a modular PT system. Specifically, we propose a novel method to optimise vehicle scheduling, determine fleet size, and generate detailed itineraries for pods within such a system. Our approach builds upon the time-expanded network transformation for MDVSP by Davari et al. (2025), which optimizes daily vehicle itineraries to minimize fleet size, and total (non-revenue) distance travelled. We adapt this framework by considering all critical operational elements including pull-out trips from depots, service trips, attach and detach events, required deadheading trips, idle times, and finally pull-in trips back to depots. This comprehensive adaptation ensures that the distinct characteristics of modular vehicles are fully integrated into the scheduling process. The method is applied to a more general version of the transferless network (Fig. 1) proposed by Khan and Menéndez (2024). Unlike the original framework, which assumes perfect symmetry in demand and service frequency, our expanded model accommodates asymmetrical demand patterns and variable service frequencies. Furthermore, it incorporates depot-related constraints and optimizes deadheading trips, ensuring the operational feasibility and efficiency of the entire system. By addressing these complexities, our approach provides a robust foundation for planning and managing modular PT systems in real-world, heterogeneous conditions.

3. Methodology

We propose a novel network flow model for operational planning of modular PT services. We adopt the basic time-expanded transformation of the single-depot vehicle scheduling problem (SDVSP) by Kliewer et al. (2006) together with the relaxation in Davari et al. (2025) to address the multi-depot VSP (MDVSP) and fleet size determination (FSDP) as a minimum-cost flow (MCF) on a time-expanded network. The MCF problem ensures all timetabled trips are delivered while minimising non-revenue movements (deadheads, pull-outs/ins, idling) and costs, and we extend this representation in Section 3.2 to the complex system of MAV's services in network operation with IMT (Khan and Menéndez, 2024).

The proposed methodology is presented in a general and systematic manner, and is applicable to arbitrary network structures and service configurations.

3.1. Basic time-expanded network

The basic time-expanded network is a widely adopted approach for modelling time-dependent transport systems, in which the temporal dimension is discretised to transform the problem into a tractable space-time representation (Nassir et al., 2014a,b; Masoud

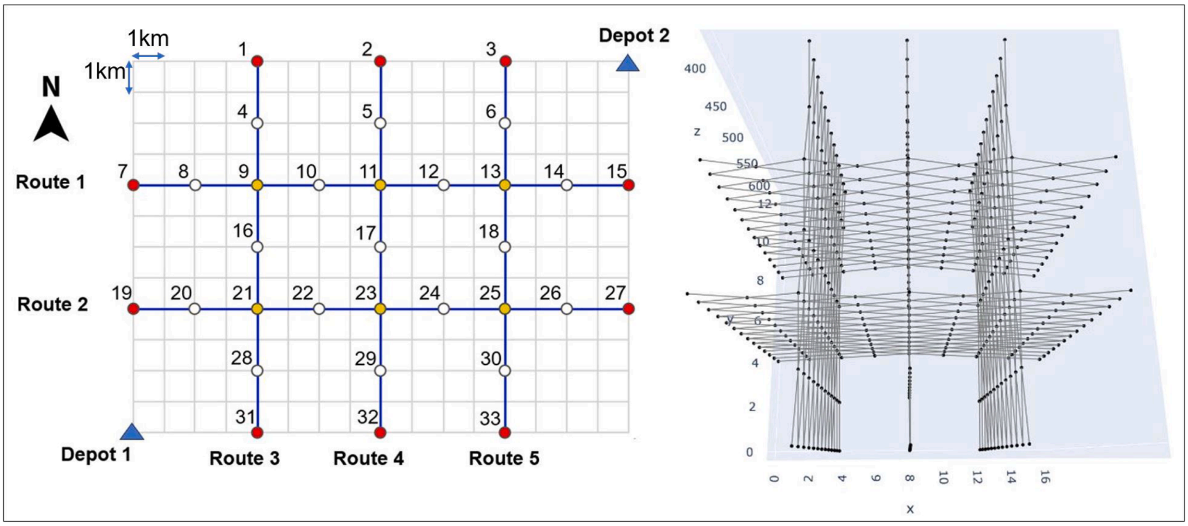

Fig. 3. Network by Khan and Menéndez (2024), and basic time-expanded network.

and Jayakrishnan, 2017). This representation constitutes the foundational structure employed in this study. It provides a discrete space-time formulation of MAV operations, where each node represents a physical location at a specific point in time and each directed link represents a feasible vehicle movement or an operational activity (coupling/decoupling). By embedding temporal ordering directly within the network, this approach enables the transformation of continuous scheduling, coordination, and synchronisation decisions into a network flow problem that can be solved using standard optimisation techniques.

The physical transport system is represented as a directed network $G = (V, A)$, where V denotes the set of physical locations and A denotes the set of directed arcs connecting them. Each node $i \in V$ corresponds to a stop, junction, terminal, or depot. The physical node set includes the subsets of stops (V^S), terminals (V^T), junctions (V^J) and depots (V^D).

$$V = V^S \cup V^J \cup V^T \cup V^D \tag{1}$$

Each arc $[i, j] \in A$ is associated with a distance $d_{ij} \geq 0$ and an average traversal speed $v_{ij} > 0$, yielding a travel time:

$$\tau_{ij} = \frac{d_{ij}}{v_{ij}} \tag{2}$$

Time is discretised using a fixed step size $\delta > 0$. Over an operational day horizon consisting of H time steps, the set of discrete time indices is defined as follows.

$$\mathcal{K} = \{0, 1, \dots, H\} \tag{3}$$

To ensure feasibility in the discrete-time representation, travel times are conservatively approximated by rounding upward. The discrete traversal time of arc $[i, j]$ is therefore:

$$\bar{\tau}_{ij} = \left\lceil \frac{\tau_{ij}}{\delta} \right\rceil \tag{4}$$

The discretisation of time entails an approximation when continuous travel times are not integer multiples of the selected time-step. In such cases, travel times are represented by rounding upward to the nearest multiple of the time-step, which may overestimate actual travel durations. The resulting approximation error is strictly less than one time-step. Importantly, this error can be reduced to any desired level of accuracy by selecting a smaller time interval, and therefore reflects a modelling choice rather than a limitation of the framework.

The service plan is assumed to be exogenously specified as an input through a timetable. Let $\mathcal{R}$ denote the set of routes, $\mathcal{D}_r$ the set of operating directions for route r, and $\mathcal{S}$ the set of timetabled service trips. Each trip $s \in \mathcal{S}$ is characterised by route $r(s)$, direction $d(s)$, and start time k_s^{start}.

For each trip s, let $\Pi_s = (i_1^s, \dots, i_{m_s}^s)$ denote the ordered sequence of visited stops, where index $h = 1, \dots, m_s$ identifies the position of a stop along the trip. We assume that dwell times at stops are negligible relative to the discretisation scale. Consequently, arrival and departure events at each stop are represented by a single time index. Each stop i_h^s is therefore associated with a single event time $k_{s,h}$, eliminating the need to distinguish between arrival and departure times.

In the basic time-expanded network, each service trip s is represented by a single aggregated link connecting its origin terminal i_1^s to its destination terminal $i_{m_s}^s$. This abstraction is consistent with classical vehicle scheduling formulations for conventional bus services, where trips are treated as indivisible units and assigned to vehicles in a holistic manner. As a result, intermediate stops are not explicitly represented, since they do not influence the vehicle scheduling decisions at this level of modelling.

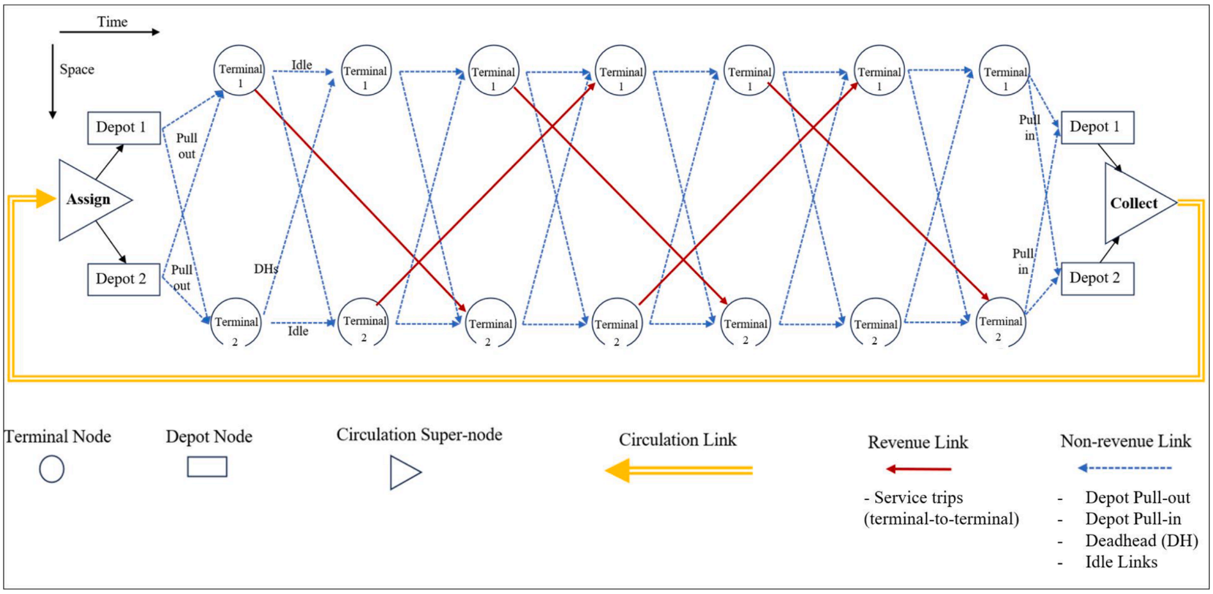

Fig. 4. Illustration for super-nodes used to transform MDVSP to SDVSP.

Under the assumption of homogeneous vehicle types, vehicles may originate from any depot, serve any scheduled trip, reposition freely, and terminate at any depot. This enables a circulation-based reformulation of the scheduling problem.

To make flow circulation possible, two virtual super-nodes are introduced: 1) an assign node o, and 2) a collect node d as dispyaed in Fig. 4. The assign node distributes vehicles to depots at the beginning of the service day, while the collect node gathers vehicles from depots at the end of the day. A circulation (virtual) link connects the collect node back to the assign node, ensuring continuity of vehicle flows from one day to another.

The basic time-expanded network is defined as $G^{TE} = (N, L^0)$, where the basic time-expanded node set is

$$N = \{(i,k) : i \in V,\ k \in \mathcal{K}\} \cup \{o,d\} \tag{5}$$

The basic time-expanded link set L^0 is composed of several subsets, each representing a distinct class of feasible vehicle movements or operational actions:

- L^{asg} (assignment links): connect the source node o to depot nodes (i,k), where i corresponds to a depot. These links represent the assignment of vehicles to enter the system at specific depots and times $k \in \mathcal{K}$.
- L^{svc} (service links): connect departure terminal nodes (i^s_1, k) to arrival terminal nodes $(i^s_{m_s}, k')$ for each scheduled service trip s. These links represent the execution of revenue service between terminals according to fixed departure times and durations.
- L^{idle} (idle links): connect consecutive event nodes (i,k) and $(i,k+1)$ of the same node i at a terminal or depot, where $k+1 \le H$. These links represent vehicles waiting of one unit at Node i.
- L^{dh} (deadheading links): connect terminal nodes (i,k) to (j,k'), where $i \ne j$, and $k' = k + \bar{\tau}_{ij}$. These links represent non-revenue repositioning movements of vehicles between terminals.
- L^{out} (pull-out links): connect depot nodes (i,k) to departure terminal nodes (j,k'), where $k' \ge k + \bar{\tau}_{ij}$. These links represent the dispatch of vehicles from depots to the starting terminal of a service trip, accounting for travel time between depot and terminal.
- L^{in} (pull-in links): connect arrival terminal nodes (i,k) to depot nodes (j,k'), where $k' \ge k + \bar{\tau}_{ij}$. These links represent the return of vehicles from service terminals to depots.
- L^{col} (collection links): connect depot nodes (i,k) to the sink node d, representing the completion of a vehicle's duty and its removal from the operation.
- L^{cir} (circulation links): connect the sink node d back to the source node o, enabling flow conservation and supporting a circulation-based network flow formulation.

Together, these subsets define all feasible operational transitions of vehicles within the time-expanded network and enable the vehicle scheduling problem to be formulated as a network flow optimisation problem.

$$L^0 = L^{asg} \cup L^{svc} \cup L^{idle} \cup L^{dh} \cup L^{out} \cup L^{in} \cup L^{col} \cup L^{cir} \tag{6}$$

The basic time-expanded network defined above serves as a reference representation to facilitate the presentation of the model extensions introduced in Section 3.2. The basic network corresponds to a conventional vehicle scheduling formulation and does not capture modular interactions such as junction transfers or stop-based decoupling and recouping. These features are introduced through the transformations presented in Section 3.2.

Table 1
Procedure for constructing basic time-expanded links.

Step	Description
Step 1	For each depot $m \in V^D$, create assignment link $[o,(m,0)] \in L^{asg}$
Step 2	For each service trip $s \in \mathcal{S}$, create service link $[(i_1^s, k_{s,1}),(i_{m_s}^s, k_{s,m_s})] \in L^{svc}$
Step 3	For each $i \in V$ and $k = 0,\dots,H-1$, create idle link $[(i,k),(i,k+1)] \in L^{idle}$, with $k+1 \le H$
Step 4	For each $i, j \in V^T$ and each k such that $k + \bar{\tau}_{ij} \le H$, create deadhead link $[(i,k),(j,k+\bar{\tau}_{ij})] \in L^{dh}$
Step 5	For each $m \in V^D$, $j \in V^T$, and each k such that $k + \bar{\tau}_{mj} \le H$, create pull-out link $[(m,k),(j,k+\bar{\tau}_{mj})] \in L^{out}$
Step 6	For each $i \in V^T$, $m \in V^D$, and each k such that $k + \bar{\tau}_{im} \le H$, create pull-in link $[(i,k),(m,k+\bar{\tau}_{im})] \in L^{in}$
Step 7	For each depot $m \in V^D$, create collection link $[(m,H),d] \in L^{col}$
Step 8	Create circulation link $[d,o] \in L^{cir}$

The procedure for constructing the basic time-expanded links is summarised in Table 1.

Link costs are specified to reflect the operational characteristics associated with each type of movement. In the resulting minimum-cost flow formulation, costs are incurred by vehicle flows traversing links in the time-expanded network. Service links are assigned zero cost, as they correspond to mandatory timetabled operations that must be performed. Idle links incur time-dependent holding costs, representing the opportunity cost of keeping vehicles inactive. Movement links, including deadhead, pull-out, and pull-in links, are assigned higher costs to reflect energy consumption and vehicle depreciation associated with repositioning. Assignment, collection, and circulation links are assigned zero cost, as they serve to initialise and balance vehicle flows rather than represent operational decisions.

Link capacities are defined on each link to regulate the feasible flow of vehicles (pods). Service links have fixed flow requirements corresponding to the prescribed platoon size of each trip. Specifically, for each service link $\ell \in L^{svc}$, the flow is fixed as $x_\ell = q_s$, ensuring that the required number of pods is allocated to each service trip. For all other links, including idle, deadhead, pull-out, and pull-in links, upper bounds may be set to sufficiently large values to avoid restricting feasible movements. However, where relevant, capacity limits can be imposed to reflect physical or operational constraints, such as limited storage capacity at nodes or restricted throughput on certain movements. Assignment, collection, and circulation links are similarly assigned capacities consistent with fleet availability or depot handling limits. These capacity definitions ensure that the resulting minimum-cost circulation remains both operationally feasible and physically consistent.

The resulting time-expanded network facilitates a minimum-cost circulation problem in which the flow variables represent vehicle assignments over space and time. The optimisation determines a feasible circulation that satisfies all service requirements while minimising the total operational cost induced by vehicle movements. In particular, the formulation simultaneously ensures that all timetabled service trips are covered and that the number of vehicles and non-revenue movements are minimised.

The basic network does not capture the additional coordination requirements arising in modular services, such as junction-based transfers and stop-level decouple–dwell–recouple operations. These interactions introduce dependencies between vehicle flows that cannot be represented within the basic structure. In Section 3.2, we introduce network transformations that extend the time-expanded network to explicitly model these operational features while preserving the tractability of the minimum-cost flow formulation.

3.2. *Transformed time-expanded network*

The basis time-expanded network introduced in Section 3.1 represents each service trip as a single aggregated link and does not capture the operational interactions required for modular services. In particular, it does not represent decoupling and recoupling of pods at service stops, nor transfers of pods between concurrent service platoons at junctions. These interactions require a structural modification of the time-expanded network.

To address this, a transformed time-expanded network $\hat{G}^{TE} = (\hat{N}, \hat{L})$ is constructed directly from the physical network and the service timetable. In this representation, selected physical nodes are expanded into structured sets of nodes corresponding to different operational stages, and service movements are represented through interconnected link sequences rather than single aggregated arcs.

Fig. 5 illustrates the proposed transformation. The physical network is shown in Fig. 5a, the corresponding time-expanded representation in Fig. 5b and d, and the detailed transformations at junctions and stops in Fig. 5c and e.

The transformed node set $\hat{N}$ is defined by expanding each stop and junction node at every time step. For each stop $i \in V^S$ and each time $k \in \mathcal{K}$, the node is replaced by four nodes:

$$N^{stop}(i,k) = \{(i,k)^{brd},\ (i,k)^{dec},\ (i,k)^{alg},\ (i,k)^{cou}\} \tag{7}$$

representing boarding, decoupling, alighting, and recoupling stages, respectively. The boarding and decoupling nodes correspond to sightly earlier (upper) stage, while the alighting and recoupling nodes correspond to the later (lower) stage (Fig. 5e). Internal links between these nodes follow a crossing structure, where pods that decouple from the service move to the alighting node, and pods that complete boarding move from the boarding node to the recoupling node, as illustrated in Fig. 5e.

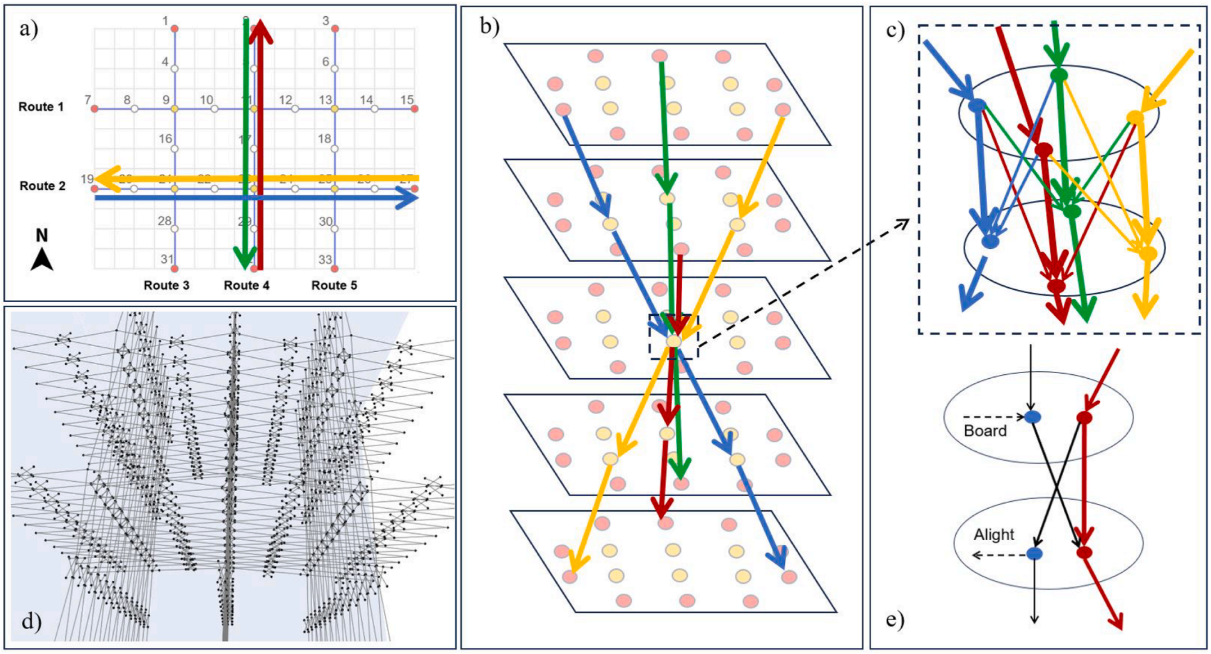

Fig. 5. Proposed transformation for junctions and stops in the time-expanded network.

For each junction $j \in V^J$ and each time $k \in \mathcal{K}$, the node is expanded according to the set of route-directions intersecting at that junction. Let $Q(j)$ denote this set. Then,

$$N^{junc}(j,k) = \{(j,k)_q^{dec},\ (j,k)_q^{cou}\ :\ q \in Q(j)\} \tag{8}$$

which yields eight nodes in the common case of four intersecting route-directions. Incoming service flows reach the decoupling nodes, from which pods may either continue along the same route or transfer to other routes. Transfers are represented by links connecting decoupling nodes of one route-direction to recoupling nodes of another, as illustrated in Fig. 5c.

Nodes associated with depots, terminals, and circulation structure are retained in the same form as in the base network. The complete transformed node set is therefore

$$\hat{N} = N^{dep} \cup N^{term} \cup N^{stop} \cup N^{junc} \cup \{o, d\} \tag{9}$$

The transformed time-expanded link set $\hat{L}$ is defined hierarchically as

$$\hat{L} = L^{svc} \cup L^{non} \cup L^{cir} \tag{10}$$

where L^{svc} denotes service-related links, L^{non} denotes non-revenue links, and L^{cir} denotes assignment and circulation links.

Service links are decomposed into stop service links and junction transfer links as

$$L^{svc} = L^{svc\text{-}stop} \cup L^{svc\text{-}junc} \tag{11}$$

The junction transfer links $L^{svc\text{-}junc}$ are further decomposed as

$$L^{svc\text{-}junc} = L^{pre\text{-}junc} \cup L^{mid\text{-}junc} \cup L^{tr\text{-}junc} \cup L^{post\text{-}junc} \tag{12}$$

The subsets are defined as follows:

- $L^{pre\text{-}junc}$ (pre-junction links): connect pods arriving from upstream stop or terminal nodes (i,k) to decoupling nodes at an upcoming junction (j,k'), where $k' = k + \tau_{ij}$. These links represent the movement of coupled pods towards a junction prior to any reconfiguration.
- $L^{mid\text{-}junc}$ (mid-junction links): connect decoupling nodes to recoupling nodes within the same route direction at a junction. These links represent traversal through the junction for pods that remain on their original route without transferring.
- $L^{tr\text{-}junc}$ (junction transfer links): connect decoupling nodes associated with one route direction to recoupling nodes associated with a different route direction within the same junction. These links represent the transfer of pods across routes, enabling dynamic re-routing and redistribution of vehicle units within the network.
- $L^{post\text{-}junc}$ (post-junction or pre-stop links): connect recoupling nodes at a junction (i,k) to downstream stop (or terminal) nodes (j,k'), where $k' = k + \tau_{ij}$. These links represent the movement of (re)coupled pods from a junction towards the next stop.

Table 2
Procedure for constructing the transformed time-expanded network.

Step	Description
Step 1	For each stop $i \in V^S$ and time $k \in \mathcal{K}$ create nodes $(i,k)^{brd}, (i,k)^{dec}, (i,k)^{alg}, (i,k)^{cou}$
Step 2	For each junction $j \in V^J$, time $k \in \mathcal{K}$, and route-direction $q \in \mathcal{Q}(j)$ create nodes $(j,k)^{dec}_q$ and $(j,k)^{cou}_q$
Step 3	For each service trip $s \in \mathcal{S}$, define the ordered sequence of service-relevant locations $$\Sigma_s = \left(n^s_1, n^s_2, \dots, n^s_{m_s}\right)$$ where $n^s_1 \in V^T$ is the initial terminal, $n^s_{m_s} \in V^T$ is the final terminal, and each intermediate node n^s_h is either a stop or a junction, listed in downstream order along the trip
Step 4	For each consecutive pair (n^s_h, n^s_{h+1}) in Σ_s, create a service-movement link from the recoupling node associated with n^s_h to the decoupling node associated with n^s_{h+1}, using the appropriate subset according to the type of downstream location $$[(n^s_h,k)^{cou}, (n^s_{h+1}, k+\bar{\tau}_{n^s_h n^s_{h+1}})^{dec}] \in L^{pre\text{-}stop} \quad \text{if } n^s_{h+1} \in V^S$$ $$[(n^s_h,k)^{cou}, (n^s_{h+1}, k+\bar{\tau}_{n^s_h n^s_{h+1}})^{dec}_q] \in L^{pre\text{-}junc} \quad \text{if } n^s_{h+1} \in V^J$$ $$[(n^s_h,k)^{cou}, (n^s_{h+1}, k+\bar{\tau}_{n^s_h n^s_{h+1}})] \in L^{svc\text{-}term} \quad \text{if } n^s_{h+1} \in V^T$$
Step 5	For each stop $i \in V^S$ and time $k \in \mathcal{K}$ create stop-operation links: $$[(i,k)^{dec},(i,k)^{alg}] \in L^{alight}, \quad [(i,k)^{brd},(i,k)^{cou}] \in L^{board}, \quad [(i,k)^{dec},(i,k)^{cou}] \in L^{stop}$$
Step 6	At each junction $j \in V^J$ and time $k \in \mathcal{K}$, for each $q \in \mathcal{Q}(j)$ create continuation and transfer links $$[(j,k)^{dec}_q,(j,k)^{cou}_q] \in L^{mid\text{-}junc}, \quad [(j,k)^{dec}_q,(j,k)^{cou}_{q'}] \in L^{tr\text{-}junc}, \quad \forall q' \in \mathcal{Q}(j),\ q' \neq q$$
Step 7	For all nodes $(i,k) \in \hat{N}$ that admit vehicle holding, create idle links $$[(i,k)^{cou},(i,k+1)^{brd}] \in L^{idle}, \quad k+1 \le H$$
Step 8	For all terminal nodes $i \in V^T$, create deadheading links $$[(i,k)^{cou},(j,k+\bar{\tau}_{ij})^{brd}] \in L^{dh}, \quad j \in V^T,\ k+\bar{\tau}_{ij} \le H$$
Step 9	Create depot and circulation links ($L^{out}, L^{in}, L^{asg}, L^{col}, L^{cir0}$) using transformed nodes and travel times (as in Table 1)

The stop service links $L^{svc\text{-}stop}$ are further decomposed as

$$L^{svc\text{-}stop} = L^{pre\text{-}stop} \cup L^{alight} \cup L^{stop} \cup L^{board} \cup L^{pre\text{-}junc} \tag{13}$$

The subsets are defined as follows:

- $L^{pre\text{-}stop}$ (pre-stop links): connect upstream post-junction recoupling nodes (i,k) to decoupling nodes at stop j, (j,k'), where $k' = k + \tau_{ij}$. These links represent the movement of coupled pods approaching a stop prior to any stopping decision.
- L^{alight} (alighting links): connect decoupling nodes to alighting nodes at the same stop and time (j,k). These links represent the separation of pods from the platoon to enable passenger alighting.
- L^{stop} (mid-stop links): connect decoupling nodes directly to recoupling nodes at the same stop and time (j,k). These links represent pods that bypass the stop.
- L^{board} (boarding links): connect boarding nodes to recoupling nodes at the same stop and time (j,k). These links represent the completion of passenger boarding and the recoupling of pods into the platoon.
- $L^{pre\text{-}junc}$ (post-stop / pre-junction links): connect recoupling nodes at stop i, (i,k), to downstream decoupling nodes (j,k'), where $k' = k + \tau_{ij}$. These links represent the movement of the recoupled platoon from a stop towards the next junction, stop, or terminal.

Non-revenue and circulation links are defined as

$$L^{non} = L^{idle} \cup L^{dh} \cup L^{out} \cup L^{in} \tag{14}$$

$$L^{cir} = L^{asg} \cup L^{col} \cup L^{cir0} \tag{15}$$

The transformed time-expanded network is constructed according to the procedure summarised in Table 2.

This construction replaces each relevant stop and junction node at every time step with a structured subnetwork and embeds the operational processes of decoupling, transfer, boarding, alighting, and recoupling directly into the network topology. The resulting network is therefore substantially more complex than the base time-expanded network.

Despite this added structural complexity, the transformed network remains compatible with a minimum-cost circulation formulation. Each link carries a flow representing the number of pods assigned to that space–time movement, and lower-bound constraints on selected links enforce the prescribed service interactions. Solving the resulting minimum-cost circulation problem yields feasible and optimal pod trajectories that are consistent with both the timetable and the operational requirements of modular services.

Each link $\ell \in \hat{L}$ is associated with a cost c_ℓ and a capacity u_ℓ, which together define the optimisation structure of the model. Costs are primarily assigned to links representing vehicle movements between locations, such as pre-stop, pre-junction, post-junction, and deadheading links, and are typically proportional to travel time.

Operational manoeuvres at stops and junctions, including decoupling, alighting, boarding, recoupling, and junction transfers, are assumed to occur within the same discrete time interval. As a result, these processes do not incur additional time or cost in the model, and the corresponding links (e.g., L^{alight}, L^{stop}, L^{board}, and $L^{tr\text{-}junc}$) are assigned zero cost. This assumption reflects the operational design in which stops and junctions are located along the platoon trajectory, allowing these interactions to occur without significant delay.

Link capacities u_ℓ impose limits on the number of pods that can traverse each link. For service movement links, capacities are determined by platoon size or infrastructure constraints, while junction transfer links may also be capacity-restricted to reflect limits on simultaneous reconfiguration. Other links, such as idle and depot links, are assigned sufficiently large capacities unless specific constraints are imposed.

Under these definitions, the transformed time-expanded network supports a minimum-cost circulation formulation in which pod flows are optimally assigned to space–time trajectories while satisfying operational and capacity constraints.

3.3. Mathematical formulation

The MAVs operational planning problem is formulated as a minimum-cost circulation on the transformed time-expanded network $\hat{G}^{TE} = (\hat{N}, \hat{L})$. Passenger demand is assumed to be satisfied by the service plan and is therefore implicitly embedded in the prescribed service flows.

The link set is partitioned as

$$\hat{L} = L^{svc\text{-}stop} \cup L^{svc\text{-}junc} \cup L^{non} \cup L^{cir} \tag{16}$$

where $L^{svc\text{-}stop}$ and $L^{svc\text{-}junc}$ denote service links at stops and junctions, respectively, $L^{non} = L^{idle} \cup L^{dh} \cup L^{out} \cup L^{in}$ represents non-revenue movements, and $L^{cir} = L^{asg} \cup L^{col} \cup L^{cir0}$ ensures flow circulation.

Let $\mathrm{tail}(\ell)$ and $\mathrm{head}(\ell)$ denote the tail and head nodes of link $\ell \in \hat{L}$. Each link is associated with a capacity $u_\ell \in \mathbb{Z}_+$, a lower bound $l_\ell \in \mathbb{Z}_+$, and a cost $c_\ell \geq 0$. Lower bounds are used to enforce required service operations, including flows through stop processes and junction transfers.

The decision variables are defined as $x_\ell \in \mathbb{Z}_+$, representing the number of pods traversing link ℓ, and $F \in \mathbb{Z}_+$, representing the total fleet size.

The primary objective is to minimise fleet size and, subject to that, minimise non-revenue movements. This is implemented through a lexicographic objective using a sufficiently large constant M:

$$\min \quad MF + \sum_{\ell \in L^{non}} c_\ell x_\ell \tag{17}$$

The objective function minimises the total operational cost, with primary emphasis on fleet size and secondary emphasis on non-revenue movements. The fleet size component is captured through the circulation flow, representing the number of vehicles required to operate the system. The cost associated with non-revenue links (e.g., deadheading, pull-out/pull-in, and idle movements) ensures that unnecessary movements are penalised, thereby improving vehicle utilisation. This lexicographic structure reflects the operational priority of minimising fleet requirements while maintaining efficient system operation. The model is subject to the following constraints:

$$l_\ell \leq x_\ell \leq u_\ell \qquad \forall \ell \in \hat{L} \tag{18}$$

$$\sum_{\ell:\,\mathrm{head}(\ell)=v} x_\ell - \sum_{\ell:\,\mathrm{tail}(\ell)=v} x_\ell = 0 \qquad \forall v \in \hat{N} \tag{19}$$

$$\sum_{\ell \in L^{out}} x_\ell = F \tag{20}$$

$$x_\ell = q_\ell \qquad \forall \ell \in L^{svc\text{-}stop} \cup L^{svc\text{-}junc} \tag{21}$$

$$\sum_{\ell \in L^{out}(m)} x_\ell \leq C_m^{\max}, \qquad \sum_{\ell \in L^{in}(m)} x_\ell \leq C_m^{\max} \qquad \forall m \tag{22}$$

$$\sum_{\ell \in L^{out}(m)} x_\ell = \sum_{\ell \in L^{in}(m)} x_\ell \qquad \forall m \tag{23}$$

Constraints (18) enforce link-level feasibility with respect to platoon size and operational limits. Constraints (19) ensure conservation of pod flows at each space–time node. Constraint (20) defines the fleet size through depot departures. Constraint (21) enforces the service plan by fixing flows on all service-related links. In this equation, q_ℓ is an input defining the number of pods that transfer at junctions or decouple to stop, assumed to be specified in the timetable. Constraints (22)-(23) impose depot m capacity limits and ensure flow balance at depots.

Costs are assigned only to non-revenue movements:

$$c_\ell = \begin{cases} \alpha^{\text{idle}}\tau_\ell, & \ell \in L^{idle} \\ \alpha^{\text{dh}}\tau_\ell, & \ell \in L^{dh} \\ \alpha^{\text{pull}}\tau_\ell, & \ell \in L^{out} \cup L^{in} \\ 0, & \ell \in L^{svc\text{-}stop} \cup L^{svc\text{-}junc} \end{cases}$$

Stop operations and junction transfers, including decoupling, alighting, boarding, recoupling, and routing decisions, are assumed to occur within the same discrete time interval and therefore do not incur additional cost.

The formulation is a minimum-cost circulation problem with integer bounds, which belong to the class of linear problems. The associated constraint matrix is totally unimodular, ensuring that the optimal solution yields integer flows (Ahuja et al., 1994). Each unit of flow corresponds to an individual pod, while aggregated flows on service links represent platoon sizes.

The optimal solution provides aggregate pod flows over the space–time network. Individual pod trajectories are obtained through a path decomposition procedure, in which unit flows are iteratively extracted from depot pull-out links to form feasible paths. The resulting paths define executable pod duty chains, and the optimal value of F corresponds to the minimum fleet size required to operate the system.

The proposed formulation results in a minimum-cost circulation problem on a time-expanded network. The model is linear and can be solved efficiently using standard network flow optimisation algorithms. Due to the structure of the formulation, integer solutions are naturally obtained without requiring additional integer constraints. This ensures computational tractability and scalability for large network instances.

4. Case study and results

A set of case study scenarios are investigated in this paper. All scenarios are designed for the simple network presented in Khan and Menéndez (2024). In our scenarios, we include two depots, one located in southwest and another in northeast corner as depicted in Fig. 3.

4.1. Baseline scenario

A baseline scenario is established based on symmetrical service frequencies (20 min headway), synchronized in both directions of all five routes within the network. The service span is set to 3 h, with a 1-h warm-up (conservatively) allocated for vehicle positioning at start of service trips, and 2-h cool-down period to allow vehicles to complete the final service trips (especially those initiated in the last hour of service) before returning to the depots.

For this model, the time-step is set to 10 min in alignment with the structure of the test network. The underlying spatial grid assumes cells of 1(km)x1(km), forming a network of approximately 16 km by 12 km, as illustrated in Fig. 3. The distance between consecutive stops and junctions is approximately 2 km. Given an assumed operating speed of 12–15 km/h for the pods, this corresponds to travel times on the order of 8–10 min per segment, which are represented as 10 min in the discretised network. This construction ensures consistency between the spatial scale, operating speed, and temporal discretisation, thereby minimising approximation error in the time-expanded representation.

The baseline scenario exemplifies a perfectly symmetrical service, where every 20 min left and right turns occur at every junction between North-South and East-West trips.

When modeling the baseline service, the network transformation model expands to include 31,745 links after time expansion. Constructing the transformation network model takes 74 s, while finding the optimal solutions requires an additional 21 s. Table 3 summarises the link flows in the optimal solution across the 31,745 links, classified by type. It displays the total link flows at each time interval, grouped by categories such as “start links”, “depot idle”, and others.

Fig. 6 displays a visualisation of the resulting pod movements in the baseline scenario. It shows, for every time interval, how many pods occupy each operational state - service, coupling, stop-servicing, idle, deadheading, and others - based on the optimal solution of the time-expanded network flow model.

As shown in Table 4, the baseline service includes 124 service trips, with two modules passing through each junction: one turning left and the other turning right. This assumption is applied consistently across all service scenarios discussed in this paper. For asymmetric services, however, platoon sizes can be adjusted to meet demand requirements and improve efficiency. For the sake of brevity, these adjustments are not covered in this paper but are detailed in the full-length version.

4.2. Asymmetric service scenarios

To illustrate how the proposed model can assist planners in understanding the operational implications of asymmetric services, Scenarios 2–7 are designed with varying service frequencies. As shown in Fig. 7, each scenario identifies specific routes or directions as candidates for higher frequency service, indicated by thick blue arrows. While service spans, schedules and route frequencies in all scenarios generally align with the baseline, the blue arrows indicate routes operating at a 10-min frequency, compared to the uniform 20-min frequency applied to the rest of the service.

Table 3
Optimal link flow summary for baseline scenario.

Link flow type	Start time																																			
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360
Assign	204	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Depot Idle	170	170	170	80	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	40	70	90	120	140	160	170	170	0	0
Pull-out (directly to stops)	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Idle at stops	0	0	1	11	26	34	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	10	34	20	34	27	19	9	0	0	0
Pull-out from depot	0	0	0	90	60	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pre-Stop	0	0	0	0	0	0	170	0	170	0	170	0	170	0	170	0	170	0	170	0	170	0	170	0	170	0	120	0	70	0	20	0	0	0	0	0	0
Mid-Stop	0	0	0	0	0	0	0	136	0	136	0	136	0	136	0	136	0	136	0	136	0	136	0	136	0	136	0	96	0	56	0	16	0	0	0	0	0
Alight link	0	0	0	0	0	0	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	24	0	14	0	4	0	0	0	0	0
Board link	0	0	0	0	0	0	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	34	0	24	0	14	0	4	0	0	0	0	0
Pre-Junction	0	0	0	0	0	0	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	70	0	20	0	0	0	0	0	0	0
Mid-Junction	0	0	0	0	0	0	0	0	72	0	72	0	72	0	72	0	72	0	72	0	72	0	72	0	72	0	72	0	54	0	20	0	0	0	0	0	0
Junction transfer (Left-Turn)	0	0	0	0	0	0	0	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	8	0	0	0	0	0	0	0	0
Junction transfer (Right-Turn)	0	0	0	0	0	0	0	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	8	0	0	0	0	0	0	0	0
Post-stop (to terminal)	0	0	0	0	0	0	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	20	0	0	0	0	0
Pull-in to depot	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	50	0	50	0	20	0	0	0	0
Pull-in (directly from stops)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	12	10	9	0	0
Collect	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	204
Circulation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	204
Deadhead	0	0	0	0	0	0	75	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	50	0	10	0	15	0	10	0	5	0	0	0	0

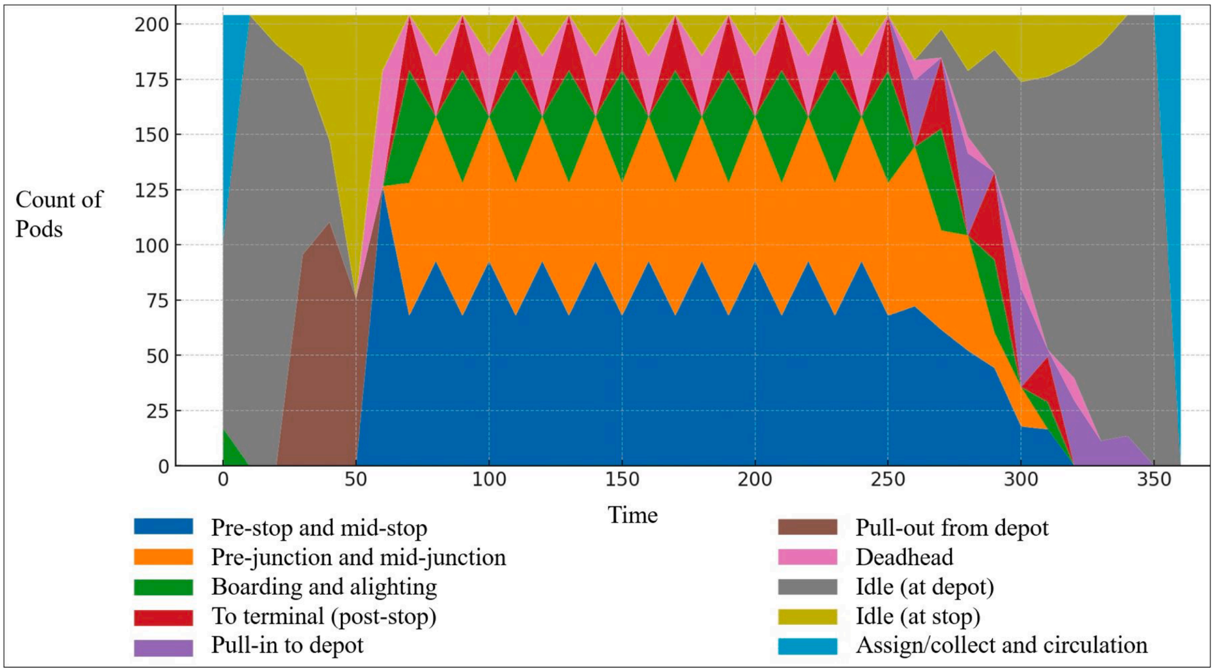

Fig. 6. Pod states in optimal solution-baseline.

The results, summarized in Table 4, include computational measures (network size and computational times) for finding the optimal solutions as well as the operational resources such as fleet size, and total non-revenue times to deliver the service in each scenario. Notable insights from Table 2 pertain to variations in required fleet size across scenarios. For instance, Scenario 2 involves 142 service trips, a 14.5% increase compared to the baseline of 124 trips. To accommodate Scenario 2, a fleet of 234 modules is required, representing 14.7% increase over Scenario 1 which requires 204 modules. Although Scenario 3 also includes 142 service trips, its geographic configuration necessitates 244 modules making it 5% more resource-intensive, than Scenario 2. This difference is likely attributed to the additional deadheading required between service trips in Scenario 3.

The fleet size disparity becomes more pronounced in Scenario 4 where both high-frequency services operate in a north-to-south direction. This configuration requires longer deadhead trips between service segments. As a result, Scenario 4 requires a 25% larger fleet compared to the baseline (254 modules versus 204.) This increase is directly attributable to the spatiotemporal layout of service. Such asymmetry is, in fact, a common characteristic of transport services, as travel demand is often directional peaking in inbound directions during morning periods and outbound directions in the afternoon. An efficient service design should align with these demand patterns, incorporating higher frequencies, larger vehicles, or longer platoons to better accommodate peak directional flows.

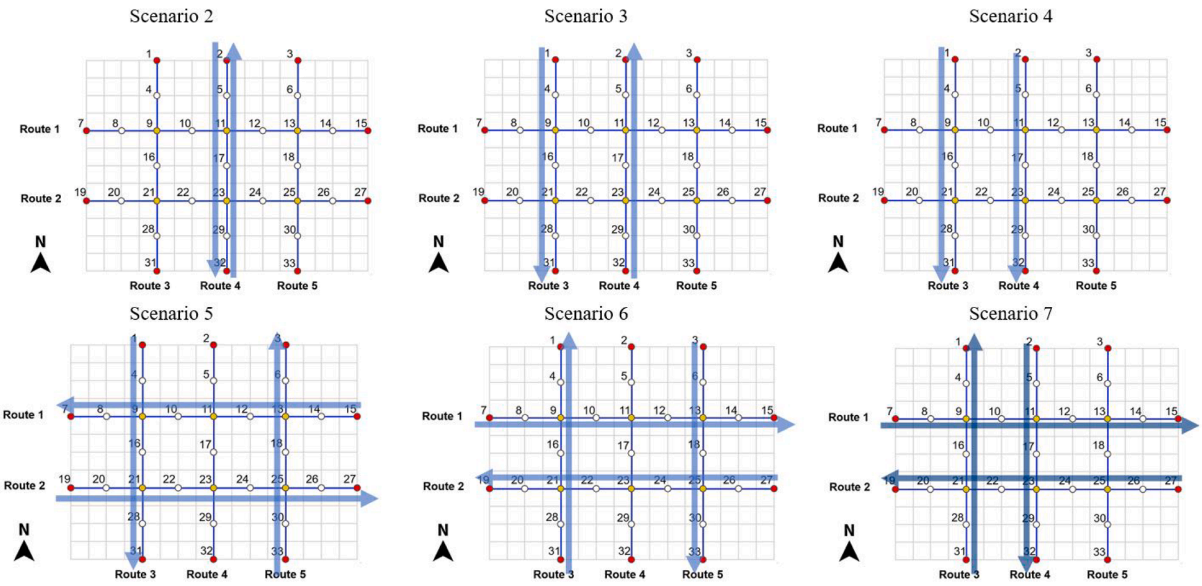

Fig. 7. High frequency services corridors in different scenarios.

Table 4
Result for all scenarios.

Scenario #	1 (baseline)	2	3	4	5	6	7
Number of trips	124	142	142	142	160	160	160
Network size (# of links)	31,745	32,051	32,051	32,051	32,525	32,525	32,529
Computational time for network (sec)	74	89	93	91	95	84	97
Computational time for solver (sec)	21	27	27	26	28	28	30
Optimal fleet size	204	234	244	254	294	294	294
Non-revenue times (minutes)	15,520	16,180	18,580	21,150	20,960	21,060	20,960
Pull-out from Depot 1	99	113	111	131	144	157	144
Pull-out from Depot 2	105	121	133	123	150	137	150

An important consideration for asymmetric services in Scenarios 2 through 7 relates to junction interchanges. Complete interchanges, as illustrated in Fig. 5b and c, are only feasible when north-south and east-west services arrive at junctions simultaneously. In scenarios with asymmetric frequencies, certain services may restrict passenger transfers. In Scenario 2, passengers on Route 4 can only transfer to east-west services if they are travelling on the baseline 20-min frequency service (every other trip). Transfers for the additional 10-min frequency services may not be possible unless these trips are specifically synchronized with a corresponding east-west high-frequency service, as implemented in Scenarios 5 through 7. Such transfer limitations are often unavoidable in real-world service planning. However, their impact can be mitigated by clearly communicating service schedules and transfer opportunities to passengers in advance.

5. Discussions and concluding remarks

In this paper, we have developed a tractable, network-scale operational planning framework for modular autonomous PT services. The proposed approach embeds all IMT-enabled manoeuvres-coupling, decoupling, recoupling, turning, and stop servicing-directly within a time-expanded minimum-cost circulation model. Two network transformations, applied at junctions and stops, enable these operations to be represented explicitly through structured link sequences. This allows in-motion transfers and stop-less servicing to be incorporated within the network while preserving the linear flow-conservation, capacity, and cost structure of classical vehicle scheduling formulations. Platoon sizes along corridors and through manoeuvres are imposed through link bounds, and depot and terminal processes are modelled using pull-out, pull-in, idle, and deadheading links. Solving the resulting formulation yields the minimum fleet required to operate a given timetable while optimally allocating non-revenue movements.

The case study on the grid network of Khan and Menéndez (2024) demonstrates both the computational efficiency and the operational insights provided by the model. The baseline symmetric scenario expands to a time-space network with 31,745 links, is constructed in approximately 74 s, and solved in 21 s, requiring a fleet of 204 pods with 15,520 min of non-revenue movement. The asymmetric scenarios highlight the sensitivity of system performance to spatiotemporal coordination in service plan. Increasing service frequency on a single corridor (Scenario 2) increases fleet requirements to 234 pods, while Scenario 3, with an identical number of trips, requires 244 pods due to additional deadheading. In Scenario 4, where high-frequency services are aligned along a single axis, reduced cross-corridor synchronisation leads to further imbalance and increases fleet size to 254 pods. These results show that fleet requirements are not driven solely by service frequency, but are strongly influenced by the alignment and synchronisation of services across the network. The framework therefore provides a quantitative basis for evaluating trade-offs between connectivity, operational efficiency, and non-revenue movements.

The operational environment of MAV services introduces challenges that differ fundamentally from conventional stop-based systems. Because passengers transfer and alight while vehicles remain in motion, correct allocation of passengers to pods becomes critical. Temporary imbalances in onboard demand may exceed the capacity of pods assigned to turning or alighting manoeuvres. In such cases, passengers may be carried beyond their intended stop or branch, leading to reliability and service-quality issues. To mitigate these risks, the framework can be extended with real-time control mechanisms that dynamically adjust pod allocations. For example, reserve pods can be dispatched from strategically located staging nodes to support overloaded platoons and subsequently reintegrated into the system. In the case study network, centrally located nodes such as 16, 17, and 18 provide suitable candidates for such reserve operations.

Several limitations of the current formulation suggest directions for future research. The service timetable is assumed to be fixed and deterministic, whereas modular systems may benefit from adaptive synchronisation strategies and flexible platoon configurations. Passenger demand is not explicitly modelled and is instead represented indirectly through prescribed service flows and platoon size bounds. Incorporating passenger assignment, demand uncertainty, and crowding effects would enable a more comprehensive assessment of service quality and system performance. Energy consumption and battery dynamics are captured only through cost approximations, and a more detailed treatment would require explicit modelling of charging infrastructure, energy constraints, and vehicle range. The model also abstracts from microscopic operational constraints, such as coupling feasibility, kinematic limits, and safety requirements associated with in-motion operations. Furthermore, key design parameters such as headways, meet windows, depot locations, and reserve capacity are treated as exogenous inputs. Joint optimisation of these elements with the operational model would provide a more integrated framework for system design and planning.

Despite these limitations, the proposed framework provides a significant step toward operationalising modular autonomous PT systems. It offers a scalable and analytically tractable approach for determining fleet requirements and vehicle movements while explicitly capturing the operational complexity of IMT-enabled services. The results demonstrate that system performance is governed by spatiotemporal coordination mechanisms, particularly at junctions and transfer points, rather than by service frequency alone. The framework therefore establishes a foundation for future developments in integrated service design, real-time control, and passenger-oriented modelling for modular transport systems.

CRediT authorship contribution statement

Neema Nassir: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Andres Fielbaum:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization; **Sapan Tiwari:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization; **Avishai (Avi) Ceder:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Data availability

No data was used for the research described in the article.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1994. Network Flows: Theory, Algorithms and Applications. Prentice hall.

Bao, H., Luo, X., Su, Q., Wang, H., 2025. Optimizing last-mile public transit services by leveraging modular autonomous vehicles. J. Transp. Eng. A Syst. 151 (8), 04025056.

Bertossi, A.A., Carraresi, P., Gallo, G., 1987. On some matching problems arising in vehicle scheduling models. Networks 17 (3), 271–281.

Bunte, S., Kliewer, N., 2009. An overview on vehicle scheduling models. Public Transp. 1 (4), 299–317.

Caros, N.S., Chow, J. Y.J., 2021. Day-to-day market evaluation of modular autonomous vehicle fleet operations with en-route transfers. Transp. B Transp. Dyn. 9 (1), 109–133.

Carpaneto, G., Dell'Amico, M., Fischetti, M., Toth, P., 1989. A branch and bound algorithm for the multiple depot vehicle scheduling problem. Networks 19 (5), 531–548.

Ceder, A., 2016. Public Transit Planning and Operation: Modeling, Practice and Behavior. CRC press.

Chen, Z., Li, X., 2021. Designing corridor systems with modular autonomous vehicles enabling station-wise docking: discrete modeling method. Transp. Res. E Logist. Transp. Rev. 152, 102388.

Chen, Z., Li, X., Qu, X., 2022. A continuous model for designing corridor systems with modular autonomous vehicles enabling station-wise docking. Transp. Sci. 56 (1), 1–30.

Chen, Z., Li, X., Zhou, X., 2019. Operational design for shuttle systems with modular vehicles under oversaturated traffic: discrete modeling method. Transp. Res. B Methodol. 122, 1–19.

Chen, Z., Li, X., Zhou, X., 2020. Operational design for shuttle systems with modular vehicles under oversaturated traffic: continuous modeling method. Transp. Res. B Methodol. 132, 76–100.

Cheng, X., Nie, Y.M., Lin, J., 2024. An autonomous modular public transit service. Transp. Res. C Emerg. Technol. 168, 104746.

Cong, Y., Bie, Y., Wang, L., Xu, N., Yang, Y., 2025. Joint timetabling and vehicle formation design for autonomous modular buses under two-level route layout mode. IEEE Trans. Intell. Transp. Syst. 26 (10), 14690–14703.

Daduna, J.R., Pinto Paixão, J.M., 1995. Vehicle scheduling for public mass transit-an overview. In: Computer-Aided Transit Scheduling: Proceedings of the Sixth International Workshop on Computer-Aided Scheduling of Public Transport. Springer, pp. 76–90.

Dai, Z., Liu, X.C., Chen, X., Ma, X., 2020. Joint optimization of scheduling and capacity for mixed traffic with autonomous and human-driven buses: a dynamic programming approach. Transp. Res. C Emerg. Technol. 114, 598–619.
Davari, N., Nassir, N., Ceder, A., Aye, L., 2025. Network-flow practical approach for scheduling of battery electric public transport buses. In: Transportation Research Board Annual Meeting. Washington, USA.
Desaulniers, G., Hickman, M.D., 2007. Public transit. Handb. Oper. Res. Manage. Sci. 14, 69–127.
Desrosiers, J., Dumas, Y., Solomon, M.M., Soumis, F., 1995. Time constrained routing and scheduling. Handb. Oper. Res. Manag. Sci. 8, 35–139.
Fahmy, T., 2018. Dubai tests world's first self-driving pods. Reuters News, March 2018. Accessed: 2025-02-01. https://www.reuters.com/article/us-dubai-transport-automation-idUSKCN1GE2AU.
Filippi, C., Guastaroba, G., Peirano, L., Speranza, M.G., 2025. Exploiting the flexibility of modular buses in an urban transit system. Transp. Res. C Emerg. Technol. 175, 105119.
Fischetti, M., Lodi, A., Martello, S., Toth, P., 2001. A polyhedral approach to simplified crew scheduling and vehicle scheduling problems. Manage. Sci. 47 (6), 833–850.
Friedrich, H.E., Ulrich, C., Schmid, S., 2019. New vehicle concepts for future business model. In: 19. Internationales Stuttgarter Symposium: Automobil-und Motorentechnik. Springer, pp. 815–829.
Fu, Z., Chow, J. Y.J., 2023. Dial-a-ride problem with modular platooning and en-route transfers. Transp. Res. C Emerg. Technol. 152, 104191.
Gecchelin, T., Spera, E., 2017. Connectable road vehicle. US Design Patent.
Gecchelin, T., Spera, E., 2022. Selectively combinable independent driving vehicles. US Patent.
Guo, Q.-W., Chow, J. Y.J., Schonfeld, P., 2018. Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options. Transp. Res. C Emerg. Technol. 94, 288–306. https://doi.org/10.1016/j.trc.2017.08.008
Hadjar, A., Marcotte, O., Soumis, F., 2006. A branch-and-cut algorithm for the multiple depot vehicle scheduling problem. Oper. Res. 54 (1), 130–149.
Han, Y., Ma, X., Yu, B., Li, Q., Zhang, R., Li, X., 2024. Planning two-dimensional trajectories for modular bus enroute docking. Transp. Res. E Logist. Transp. Rev. 192, 103769.
Hatzenbühler, J., Jenelius, E., Gidófalvi, G., Cats, O., 2023. Modular vehicle routing for combined passenger and freight transport. Transp. Res. A Policy Pract. 173, 103688.
Huang, D., Hu, Z., Tian, J., Tu, R., 2025. Improving conventional transit services with modular autonomous vehicles: a bi-level programming approach. Travel Behav. Soc. 39, 100939.
Jara-Diaz, S., Monzon, A., Cascajo, R., Garcia-Martinez, A., 2022. An international time equivalency of the pure transfer penalty in urban transit trips: closing the gap. Transp. Policy 125, 48–55.
Jia, Z., An, K., 2025. Multi-depot battery electric bus scheduling and charging coordination under resource limitations. Appl. Energy 398, 126444.
Khan, Z.S., He, W., Menéndez, M., 2023. Application of modular vehicle technology to mitigate bus bunching. Transp. Res. C Emerg. Technol. 146, 103953.
Khan, Z.S., Menendez, M., 2022. No time for stopping: a stop-less autonomous modular (Slam) bus service. Available SSRN 4225419 .
Khan, Z.S., Menéndez, M., 2023. Bus splitting and bus holding: a new strategy using autonomous modular buses for preventing bus bunching. Transp. Res. A Policy Pract. 177, 103825.
Khan, Z.S., Menéndez, M., 2024. A seamless bus network without external transfers using autonomous modular vehicles. Transp. Res. C Emerg. Technol. 158, 104475. https://doi.org/10.1016/j.trc.2024.104475
Kliewer, N., Mellouli, T., Suhl, L., 2006. A time–space network based exact optimization model for multi-depot bus scheduling. Eur. J. Oper. Res. 175 (3), 1616–1627.
Kulkarni, S., Krishnamoorthy, M., Ranade, A., Ernst, A.T., Patil, R., 2018. A new formulation and a column generation-based heuristic for the multiple depot vehicle scheduling problem. Transp. Res. B Methodol. 118, 457–487.
Laurent, B., Hao, J.-K., 2009. Iterated local search for the multiple depot vehicle scheduling problem. Comput. Ind. Eng. 57 (1), 277–286.
Li, Q., Li, X., 2022. Trajectory planning for autonomous modular vehicle docking and autonomous vehicle platooning operations. Transp. Res. E Logist. Transp. Rev. 166, 102886.
Lin, J., Nie, Y.M., Kawamura, K., 2022. An autonomous modular mobility paradigm. IEEE Intell. Transp. Syst. Mag. 15 (1), 378–386.
Lin, Y., Zhang, L., Wang, J., 2024. Bus bunching control with modular autonomous vehicles and in-motion passenger transfers. Transp. Res. C Emerg. Technol. 162, 104578. https://doi.org/10.1016/j.trc.2024.104578
Liu, T., Ceder, A., Rau, A., 2018. Deployment planning of single-line modular-vehicle semi-rapid transit system. In: Proceedings of the 14th International Conference on Advanced Systems in Public Transport (CASPT 2018). Brisbane, Australia.
Liu, Y., Chen, Z., Wang, X., 2024. Alleviating bus bunching via modular vehicles. Transp. Res. B Methodol. 189, 103051.
Liu, Z., de Almeida Correia, G.H., Ma, Z., Li, S., Ma, X., 2023. Integrated optimization of timetable, bus formation, and vehicle scheduling in autonomous modular public transport systems. Transp. Res. C Emerg. Technol. 155, 104306.
Löbel, A., 1999. Solving large-scale multiple-depot vehicle scheduling problems. In: Computer-Aided Transit Scheduling: Proceedings, Cambridge, MA, USA, August 1997. Springer, pp. 193–220.
Luo, X., Fan, W., Xu, M., Yan, X., 2025. Optimal design of on-demand modular feeder transit services. Transp. Res. C Emerg. Technol. 178, 105245.
Luo, X., Fan, W., Zhang, Y., D'Ariano, A., Bosi, T., Liu, Y., Wu, W., Li, X., 2026. A continuum approximation approach for designing corridor-based heterogeneous transit service using modular autonomous vehicles. Transp. Res. E Logist. Transp. Rev. 205, 104512.
Ma, C., Zhou, H., Zhang, P., Ma, K., Shi, H., Li, X., 2025. Safety assurance adaptive control for modular autonomous vehicles. Commun. Transp. Res. 5, 100204.
Masoud, N., Jayakrishnan, R., 2017. A decomposition algorithm to solve the multi-hop peer-to-peer ride-matching problem. Transp. Res. B Methodol. 99, 1–29.
Nassir, N., Hickman, M., Zheng, H., Chiu, Y.-C., 2014a. Network flow solution method for optimal evacuation traffic routing and signal control with nonuniform threat. Transp. Res. Rec. 2459 (1), 54–62.
Nassir, N., Zheng, H., Hickman, M., 2014b. Efficient negative cycle–canceling algorithm for finding the optimal traffic routing for network evacuation with nonuniform threats. Transp. Res. Rec. 2459 (1), 81–90.
NEXT Future Transportation Inc., 2018. Next future transportation. http://www.next-future-mobility.com. Accessed: 2020-10-03.
Pepin, A.-S., Desaulniers, G., Hertz, A., Huisman, D., 2009. A comparison of five heuristics for the multiple depot vehicle scheduling problem. J. Sched. 12 (1), 17–30.
Rau, A., Tian, L., Jain, M., Xie, M., Liu, T., Zhou, Y., 2019. Dynamic autonomous road transit (DART) for use-case capacity more than bus. Transp. Res. Procedia 41, 812–823. https://doi.org/10.1016/j.trpro.2019.09.131
Ricard, L., Desaulniers, G., Lodi, A., Rousseau, L.-M., 2024. Increasing schedule reliability in the multiple depot vehicle scheduling problem with stochastic travel time. Omega 127, 103100.
Romea, G., Estrada, M., 2021. Analysis of an autonomous driving modular bus system. Transp. Res. Procedia 58, 181–188.
Shafiee, A., Moghaddam, H.R., Lin, J., 2024. Using autonomous modular vehicle technology as an alternative for last-mile delivery. In: 2024 Forum for Innovative Sustainable Transportation Systems (FISTS). IEEE, pp. 1–6.
Shi, X., Chen, Z., Li, X., Qu, X., 2024. Modular vehicles can reduce greenhouse gas emissions for departure flight baggage transportation. J. Air Transp. Manage. 119, 102633.
Shi, X., Li, X., 2021. Operations design of modular vehicles on an oversaturated corridor with first-in, first-out passenger queueing. Transp. Sci. 55 (5), 1187–1205.
Sun, S., Wong, Y.D., Rau, A., 2020. Economic assessment of a dynamic autonomous road transit system for Singapore. Res. Transp. Econ. 83, 100843.
Tang, C., Liu, J., Ceder, A., Jiang, Y., 2024. Optimisation of a new hybrid transit service with modular autonomous vehicles. Transp. A Transp. Sci. 20 (2), 2165424.
Tian, Q., Lin, Y.H., Wang, D. Z.W., 2023. Joint scheduling and formation design for modular-vehicle transit service with time-dependent demand. Transp. Res. C Emerg. Technol. 147, 103986.
Tian, Q., Lin, Y.H., Wang, D. Z.W., Liu, Y., 2022. Planning for modular-vehicle transit service system: model formulation and solution methods. Transp. Res. C Emerg. Technol. 138, 103627.
Tian, Q., Lin, Y.H., Wang, D. Z.W., Yang, K., 2025. Toward real-time operations of modular-vehicle transit services: from rolling horizon control to learning-based approach. Transp. Res. C Emerg. Technol. 170, 104938.

Torres, T., Estrada, M., Badia, H., 2025. Design of urban transit services using autonomous modular buses. Transp. B Transp. Dyn. 13 (1), 2540830.
Ulrich, C., Friedrich, H.E., Weimer, J., Hahn, R., Kopp, G., Münster, M., 2019. Technologies for a modular vehicle concept used in passenger and goods transport. In: 19. Internationales Stuttgarter Symposium: Automobil-und Motorentechnik. Springer, pp. 587–598.
van Lieshout, R.N., van der Schaft, T., 2025. Dynamic discretization discovery for the multidepot vehicle scheduling problem with trip shifting. INFORMS J. Comput. 38 (3), 693–1031.
Wang, Y., Ceder, A., Cao, Z., Zhang, S., 2025a. Optimal public transport timetabling with autonomous-vehicle units using coupling and decoupling tactics. Transp. A Transp. Sci. 21 (1), 2220423.
Wang, Z., An, K., de Almeida Correia, G.H., 2025b. Optimizing a modular autonomous vehicle hub-and-spoke public transportation system: routing, scheduling, and repositioning. Transp. Res. E Logist. Transp. Rev. 201, 104251.
Wu, J., Kulcsár, B., Qu, X., et al., 2021. A modular, adaptive, and autonomous transit system (MAATS): an in-motion transfer strategy and performance evaluation in urban grid transit networks. Transp. Res. A Policy Pract. 151, 81–98.
Yao, Z., Zhang, Q., Fu, C., Wu, Y., Jiang, Y., 2025. Collaborative optimization of vehicle formation and timetable for modular autonomous bus considering demand uncertainty: a data-driven continuous approximate method. Transp. Res. E Logist. Transp. Rev. 200, 104176.
Ye, L., Yamamoto, T., 2024. Modular autonomous vehicle in heterogeneous traffic flow: modeling, simulation, and implication. arXiv preprint arXiv:2409.17945.
Yi, H., Liu, Y., Yang, H., Menéndez, M., 2025. Charging-on-the-move public transit systems constituted by modular trolleybuses and modular buses. IEEE Trans. Transp. Electrif. 11 (1), 3491–3501.
Zermasli, D., Iliopoulou, C., Laskaris, G., Kepaptsoglou, K., 2023. Feeder bus network design with modular transit vehicles. J. Public Transp. 25, 100078.
Zhang, J., Ge, Y.-E., Tang, C., Zhong, M., 2024. Optimising modular-autonomous-vehicle transit service employing coupling–decoupling operations plus skip-stop strategy. Transp. Res. E Logist. Transp. Rev. 184, 103450.
Zhang, Y., Chen, X., Zhao, Y., Ma, J., Yu, L., 2022. A methodology for measuring the environmental effect of autonomous bus considering platooning. Transp. Res. D Transp. Environ. 107, 103300.
Zhang, Z., Tafreshian, A., Masoud, N., 2020. Modular transit: using autonomy and modularity to improve performance in public transportation. Transp. Res. E Logist. Transp. Rev. 141, 102033.
Zhao, Y., Liu, T., Meng, W., Dai, Z., Fan, W., 2026. Optimizing timetabling and platoon formation of shuttle transit with modular autonomous vehicles considering time-dependent passenger demand. Transp. Lett. 18 (4), 823–840.