Full Length Article

Sustainable and convenient: Bi-modal public transit systems outperforming the private car

Puneet Sharma, Knut M. Heidemann, Helge Heuer, Steffen Mühle, Stephan Herminghaus

Max Planck Institute for Dynamics and Self-Organization (MPIDS), Am Füßberg 17, Göttingen 37077, Germany

1. Introduction

Transportation accounts for about one fifth of global anthropogenic carbon emissions (Fan et al., 2018; Kontovas and Psarafitis, 2016), owing mainly to the fact that humans rely mostly on motorized individual vehicles (MIV), i.e., private cars (Douglas et al., 2011; Newman and Kenworthy, 1989). Aside from the ensuing logistic (Jang et al., 2016; Manville and Shoup, 2005; Mingardo et al., 2022; Park et al., 2012; Swenseth and Godfrey, 2002) and environmental (eea, 2020; Joireman et al., 2004) problems of traffic congestion (Arnott and Small, 1994; Barth and Boriboonsomsin, 2009; Chin, 1996; Kožlak and Wach, 2018) and air pollution (eeab, 2020; Caiazzo et al., 2013), MIV represents an enormous waste of energy. On average it amounts to moving about a ton of material (MacKenzie et al., 2014) in order to move just one person (Tachet et al., 2017). Nevertheless, it maintains a leading market position (eurostat, 2022; Fiorello et al., 2016) because it is convenient (Kent, 2013) and relatively cheap for its users.

A well-known answer to this problem is ride-pooling (Chen et al., 2017; Shaheen and Cohen, 2019; Zwick et al., 2021), i.e., combining routes of several passengers such that they can be served by one vehicle (Merlin, 2019). This is done most efficiently by line services, like trains, light rail, or the underground. A light rail car easily takes a hundred passengers or more, uses much less...
energy than the same number of MIV and contributes next to nothing to traffic congestion (Pietrzak, 2016). Therefore, many large cities (like, e.g., Tokyo) rely heavily on transportation by line services (Ferbrache and Knowles, 2017; Kato et al., 2014; Pietrzak and Pietrzak, 2019).

They come, however, with a serious downside when compared to MIV. With the latter, users can freely choose the starting time and location as well as the destination. This is not possible for line services, which must follow fixed schedules and fixed routes (Alam et al., 2015). Users thus may have to walk significant distances to and from stations, and need to know the schedules of the involved lines. Demand-responsive ride-pooling (DRRP) (Herminghaus, 2019) services try to address this problem by deploying a large number of shuttles which pick up and drop off users at the desired locations. This requires a central facility which collects travel requests, along with a powerful algorithm which combines these requests into appropriate routes of the shuttles (Alonso-Mora et al., 2017). In such systems, users necessarily experience some detour (Herminghaus, 2019; Lobel and Martin, 2020) with respect to the direct route they could have taken via MIV. This trade-off (Daganzo et al., 2020) severely limits the achievable pooling efficiency to well below ten passengers per vehicle (Zwick et al., 2021). Hence while DRRP is more attractive than line services because it provides door-to-door transport, its pooling efficiency is intrinsically much inferior.

The observations above suggest the following question: “Can we achieve both, high pooling efficiency and attractive service quality, by combining line services with on-demand door-to-door transport in a single system of transportation?” While there have been quite a number of studies on ride-pooling systems before (Chen et al., 2017; Lotze et al., 2022; Molkenthin et al., 2020; Salonen, Toivonen, 2013; Santi et al., 2014; Sorge et al., 2015; Storch et al., 2021; Sundt et al., 2021; Tachet et al., 2017; Vazifeh et al., 2018; Wolf et al., 2022), the combination of different transport modes has so far been only scarcely addressed and remains to be explored. Thus, in this paper we fill this gap and investigate a bi-modal public transit system, which consists of a combination of both transport modes. A line service, with fixed routes and schedule, shall coexist with a fleet of shuttles which pick-up users and bring them either to and from line service stations, or serve shorter-distance requests directly. This provides both door-to-door transport by virtue of the shuttles and a large average pooling efficiency due to the involvement of line service vehicles.

In Section 2, we first introduce the bi-modal model system with a square-grid geometry together with a mean-field description. We identify the relevant parameters in Section 3, which need to be controlled in such a combined system of transportation to optimize the objectives of operation (see Section 4). By computing Pareto fronts, we explore in Section 5 what performance may be achieved in terms of energy consumption, quality of service, and traffic volume. We find that bi-modal transit has the potential to provide on-demand door-to-door service with a quality superior to customary public transportation, while at the same time consuming only a fraction of the energy a corresponding fleet of MIV would require, and with a road traffic volume reduced by an order of magnitude.

2. Definition of the system

2.1. User environment

For the sake of conciseness and simplicity, we consider a planar area uniformly populated at density € with potential users of the public transit system under study. Users are assumed to place transit requests in an uncorrelated fashion, each consisting of a desired pick-up (P) and drop-off (D) location, at an average rate ν per passenger. Requested travel distances d = PD are assumed to follow a certain distribution, p(d), with mean D (Herminghaus, 2019).

For a transparent discussion, it is useful to introduce dimensionless parameters characterizing the system under study. By combining the intrinsic length scale D with a characteristic road vehicle velocity, v0, we obtain an intrinsic time scale, t0 = D/v0. This is the average time a travel request would need to be completed by MIV. The demand of transport within the system can then be characterized by the dimensionless parameter Λ = EνD2/v0, which can reach well beyond 10^4 in a densely populated area. Tab. 1 provides a few typical parameters encountered in real systems for reference. Note that D = √m/D, where m is the average area enclosed by surrounding rail (line) services, and D is the spacing of line service routes (see below, Fig. 2.2).

Tab. 1

<table>
<thead>
<tr>
<th>city/district</th>
<th>type</th>
<th>E</th>
<th>D</th>
<th>c0v</th>
<th>m</th>
<th>ℓ &gt;</th>
<th>Λ</th>
<th>Q</th>
<th>ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York City</td>
<td>dense urban</td>
<td>1.1·10^4</td>
<td>4.99</td>
<td>11.3</td>
<td>2.0</td>
<td>0.28</td>
<td>1.5·10^4</td>
<td>0.33</td>
<td>Herminghaus, 2019; NYCDOT, 2018; TLC, 2022; USCB, 2020</td>
</tr>
<tr>
<td>Berlin</td>
<td>urban</td>
<td>4.1·10^3</td>
<td>5.90</td>
<td>19.8</td>
<td>3.6</td>
<td>0.32</td>
<td>5.0·10^3</td>
<td>0.64</td>
<td>AIS, 2021; Gerike et al., 2018</td>
</tr>
<tr>
<td>Ruhr (north)</td>
<td>urban</td>
<td>3.6·10^3</td>
<td>15.7</td>
<td>44.9</td>
<td>94</td>
<td>0.62</td>
<td>3.6·10^4</td>
<td>0.34</td>
<td>Haller and Dauth, 2018</td>
</tr>
<tr>
<td>Emsland</td>
<td>rural</td>
<td>1.1·10^2</td>
<td>16.7</td>
<td>58.7</td>
<td>1200</td>
<td>2.1</td>
<td>1.0·10^5</td>
<td>0.35</td>
<td>IAB, 2018; SIBA, 2020</td>
</tr>
</tbody>
</table>

1 Average number of incoming requests in an area D^2 in time t0.
2.2. Model system geometry

For an overarching systematic study, it is useful to consider an idealized model geometry (see Fig. 2.2). We assume that transport occurs via DRRP shuttle service, combined with a square grid of railways on which transport occurs via trains. The connection points (train stations) between the two subsystems lie at all railway intersections and are spaced with a lattice constant $\ell$ (see Fig. 2.2). The transit system is further characterized by a shuttle density $B$ in the plane and a train frequency $\mu$ at all train stations, with trains having a seating capacity $k$. Shuttles and trains move with velocities $v_0$ and $v_{\text{train}}$, respectively. They require energy $e_{\text{shuttle}}$ and $e_{\text{train}}$, respectively, per unit distance of travel.

The main goal of the bi-modal system under consideration is to provide high quality (i.e., rapid) door-to-door transportation service at minimal energy consumption, thereby minimal carbon emission. To reach this goal, the provider of bi-modal transit may vary certain parameters of operation. We will first introduce these control parameters in Section 3 and then derive expressions for the system’s service quality and energy consumption as functions of these parameters in Section 4.

3. Parameters of operation

3.1. Choosing the type of transport service

A single user in the model system may either be transported by uni-modal service, i.e., by shuttle (DRRP) only, or by bi-modal service, i.e., be brought from $P = (x_P, y_P)$ to the nearest train station by means of a shuttle, followed by a train journey, which is again followed by a shuttle journey to $D = (x_D, y_D)$ (see Fig. 2.2). Aside from assembling the routes of the shuttles such as to optimize pooling efficiency, one central task of the dispatcher system will be to decide, for each individual request $(P, D)$, whether the desired door-to-door service should be completed uni-modally or bi-modally.

If only user convenience were considered relevant, one would just need to calculate which type of transport service (uni-modal or bi-modal) requires less time for completing the transit, and then to choose that one. This requires knowledge of the parameters $v_0$, $v_{\text{train}}$, and the frequency of line service, $\mu$. The latter can be assumed to be just sufficient to carry the bi-modal passenger load. Its derivation will be discussed further below (see Subsec. 3.2). By sampling random transport requests in the plane, and distances from the probability distribution $p(d)$ of travel distances, we can compile a histogram of relative travel times, $t_{\text{sh}}$ and $t_{\text{uni}}$, as shown in Fig. 2a. It displays the resulting ratio $t_{\text{sh}}/t_{\text{uni}}$ as a scatter heat map in the plane spanned by the vector $\vec{PD}$. Requests corresponding to the area within the black curve (contour curve of $t_{\text{sh}}/t_{\text{uni}} = 1$) would then be served uni-modally by a single shuttle services, while for all others, the dispatcher would offer bi-modal transport service.

In order to choose, for an incoming request $(P, D)$, the type of transport service which consumes the least incremental amount of energy, we have to compute the energy increment $(\Delta E)_{\text{bi}}$ needed for bi-modal transport, and compare it to the energy increment

\[2 \text{ Subsec. 4.1 displays the mathematical expressions of } t_{\text{sh}}, t_{\text{uni}}.\]
Fig. 2. Choosing the type of transport service. Relative characteristics of either bi-modal (shuttle-train(s)-shuttle) or uni-modal (just shuttle) service, in the plane spanned by the individual trip vector from pick-up $P = (x_P, y_P)$ to drop-off $D = (x_D, y_D)$. (a) Bi-modal travel time, $t_{\text{bi}}$, divided by uni-modal travel time, $t_{\text{uni}}$. The black curve represents the contour line where both are equal. Requests outside this region are served faster with bi-modal transportation. (b) Increment in total energy consumption if a new user is served by bi-modal transportation, $(\Delta E)_\text{bi}$, divided by the increment in total energy consumption if the same user is served by uni-modal transportation, $(\Delta E)_\text{uni}$. The black curve represents the contour line where both are equal, i.e., from the perspective of energy consumption both types of transport service are equivalent. Requests outside the white region lead to lower energy consumption if served by bi-modal transportation. See Supporting Information for details.

$(\Delta E)_\text{uni}$ assuming direct transport via a single shuttle. This ratio of energy increments is equal to the ratio of driven distances by the shuttles for each type of service (see also Fig. 2.2). We assume that a single request does not alter the line service frequency, i.e., the energy consumption of the line service does not change. In analogy to the travel times shown in Fig. 2a, the ratio of the energy consumption increments (i.e., traveled distances by shuttles) is depicted in Fig. 2b. Again, we see that while for small requested distances uni-modal service is advisable, bi-modal service should be preferred for larger distances, corresponding to the area outside the black contour curve.

Comparison of Fig. 2a and 2b reveals that there is a significant range of distances which lie outside the solid curve in Fig. 2b, but still well inside the curve depicted in Fig. 2a. This shows that we may have to deal with conflicting objectives for quite a number of incoming transport requests. The notion of optimality then depends upon the relative valuation of energy consumption and service quality. As a generally accepted way of dealing with conflicting objectives, we will tackle this problem by means of Pareto fronts (Debreu, 1959; Greenwald and Stiglitz, 1959; Magill and Quinzii, 2002) further below (see Subsec. 5.1).

While the plot in Fig. 2b represents a rather isotropic structure, we encounter a shamrock-like shape in Fig. 2a. This reflects the orthogonal geometry of our model line service system (Fig. 2.2). In a real situation, the geometry will in general not be this simple. Instead, the directions at which rails are installed will vary from one station to another. We thus expect a structure like the ‘shamrock’ to be less pronounced in reality, if discernible at all. Hence although a perfectly isotropic structure may not be expected, the anisotropy will certainly be less pronounced. We assume that it will be a reasonable approximation to consider the contour line of service times as ‘sufficiently’ circular. Therefore we consider henceforth only the requested travel distance, $d = |\vec{PD}|$, as the discriminating parameter for the choice of type of transport service, irrespective of its direction. The task of the dispatcher will then be to determine a proper cutoff distance, $d_c$, such that for $d > d_c$, bi-modal service is offered, while for $d \leq d_c$, the system will provide uni-modal service, by shuttle only.

Note that the above approximation provides a lower bound of the performance achievable. In a real system, the type of transport service may be decided upon the true expected travel times and energy consumption, for which data will be available with ever improving quality over time.

3.2. Choice of line service frequency

It is clear that the capacity $k$ and frequency $\mu$ of the line service must be sufficient to carry the flux of shuttle passengers towards and from the train stations. The total number of requests emanating in unit time in an area of $\epsilon^2$ around a train station is $vE\epsilon^2$. Out of these, only a fraction $F = F(d) = \int_{d_c}^{\infty} p(d') \, dd'$ is served by bi-modal transportation. However, trains are also occupied by passengers from previous stations. If $D_{\text{train}}$ is the average distance that users travel on trains, then a user travels $D_{\text{train}}/\epsilon$ stations on train on average. Therefore, the total number of users to be transported at this station per unit time is

$$J = vE\epsilon^2 F \frac{D_{\text{train}}}{\epsilon}.$$  

We find that $D_{\text{train}} = \frac{4}{3} \langle d \rangle_{d > d_c}$, and $\langle d \rangle_{d > d_c}$ the mean of requested distances larger than $d_c$.

3 Averaging the 1-norm $||\vec{PD}||_1$ over distances and orientations.
A similar relation holds for the number of users per unit time that can be transported by trains arriving at one train station (with frequency $\mu_0$ and going into four directions), namely
\[ J_{\text{out}} = 4 \cdot \mu_0 \cdot k. \]  
(2)

Balancing $J_{\text{in}}$ with $J_{\text{out}}$, we obtain
\[ \tilde{\mu}_0 = \frac{\Lambda F}{\Delta k} (d)_{d > d_k} F \]
(3)
for the minimum frequency required to carry all passengers conveyed by the shuttles. The $\tilde{}$ indicates quantities non-dimensionalized via division by the respective unit, i.e., $D$ or $t_0$. We refer to Eq. 3 as passenger flux balance. 

If we allow trains to operate at a frequency $\tilde{\mu}$ larger than the minimum required frequency $\tilde{\mu}_0$, the train occupancy is given by $\alpha = \tilde{\mu}_0 / \tilde{\mu} \in [0, 1]$. As this can be adjusted within some range when operating the line service, $\alpha$ provides an additional free variable in system operation.

4. Objectives of operation

4.1. Service quality

We define the service quality as the ratio between average travel time by MIV and by the bi-modal system
\[ Q = \frac{t_0}{(1 - F) \cdot t_{\text{uni}} + F \cdot t_{\text{bi}}}. \]
(4)
For assessing the overall quality of service, suitable averaging has to be applied. Transportation by shuttles is always assumed to be delayed with respect to MIV by a waiting time, which we assume (on average) to be of order one half the direct travel time, $t_0/2$ (see Supporting Information for motivation). The average time taken to serve a request in a bi-modal system (i.e., the denominator of $Q$ in Eq. 4) is then
\[ t_0 Q^{-1} = (1 - F) \cdot \left( \frac{t_0}{2} + \frac{\delta (d)_{d < d_k}}{v_0} \right) + F \cdot \left( t_0 + \frac{2 \beta \epsilon \delta}{v_0} + \frac{1}{\mu} + \frac{4}{\pi} \frac{(d)_{d > d_k}}{v_{\text{train}}} \right), \]
(5)
where $\langle d \rangle_{d < d_k}$ represents the mean of all requested distances less than $d_k$ (i.e., served uni-modally) and $\delta$ is the average detour incurred by the shuttles due to the necessity of pooling several different transport requests into one vehicle route. For the expected detour with a shuttle we set $\delta = 1.5$ (see Supporting Information for details). For the second term ($t_0/2$), $t_0$ is the total average waiting time for two shuttle trips (to and from the station), $1/\mu$ is the average waiting time for two train rides (usually there is a change involved), $\beta \epsilon$ is the average distance of a randomly chosen point from the next train station, with a geometrical constant $\beta \approx 0.383^5$, and $4\pi^{-1}(d)_{d > d_k}$ is the average distance traveled on trains. The effective train velocity, $v_{\text{train}}$, depends on the inter-station distance $\ell^\prime$ and is modeled based on train vehicle data (see Supporting Information for details).

If we use $D$, $t_0$, and $v_0$ as units for length, time, and velocity, respectively, we can write:
\[ Q^{-1} = (1 - F) \cdot \left( \frac{1}{2} + \delta (d)_{d < d_k} \right) + F \cdot \left( 1 + 2 \beta \epsilon \delta + \frac{1}{\mu} + \frac{4}{\pi} \frac{(d)_{d > d_k}}{v_{\text{train}}} \right). \]
(6)

4.2. Energy consumption

In order to assess the efficiency of a transit system in terms of energy consumption, it is essential to consider the total distances over which passengers are being transported in the different vehicles involved (see Eq. 7). The bi-modal energy consumption can be written as
\[ \mathcal{E} = \frac{\Delta_{\text{shuttle}} \cdot e_{\text{shuttle}} + \Delta_{\text{train}} \cdot e_{\text{train}}}{\Delta_{\text{MIV}} \cdot e_{\text{MIV}}}. \]
(7)
where $\Delta$ denotes the (mode-specific) total distance traveled in a unit cell of size $\ell^2$ per unit time, and $e_{\text{shuttle/train}}$ is the vehicle-specific energy consumption per unit distance. Note that this expression is already normalized with respect to the MIV energy consumption (denominator), as this is the door-to-door transportation system we intend to compare with. For $\mathcal{E} > 1$ ($< 1$) energy requirement for bi-modal transportation is more (less) than for private cars serving the same requests.

For the simulations we will describe below, we use numbers found for frequently used transport vehicles. Specifically, we consider electric light rails with a maximum seating-capacity $k = 100$ and $e_{\text{train}} = 9.72$ kN (Knörr et al. (2016)) for the line service. For MIV we consider Diesel cars with $e_{\text{MIV}} = 2.47$ kN (BMDV, 2022). For the shuttle we choose Mercedes Sprinter (8.8 liters of Diesel per 100 km (mbenz, 2022), thus $e_{\text{shuttle}} = 3.28$ kN.

\[ ^4 \text{ driven distance / direct distance} \]
\[ ^5 \text{ A simple calculation shows that } \beta = \frac{1}{\pi} (\sqrt{2} + \log(1 + \sqrt{2})) = 0.383. \]
Shuttles Both uni-modal (shuttle only) and bi-modal trips contribute to the total distance driven by shuttles per unit time due to requests from a unit cell of size $\epsilon^2$, hence

$$\Delta_{\text{shuttle}} = \frac{\nu E \epsilon^2}{\eta} \left( (d)_{d<d_{\text{shuttle}}} (1 - F) + \frac{2 \beta F}{\text{twoshuttletrips}} \right),$$

(8)

where $\eta$ is the DRRP pooling efficiency, which is the ratio of requested direct distance by the users and the driven distance by the shuttles (for MIV, $\eta = 1$).

In simulations of the uni-modal system (shuttles only), it has been observed that $\eta$ scales with demand $\Lambda$ roughly in an algebraic manner, $\eta(\Lambda) \propto \Lambda^{\gamma}$, with $\gamma \approx 0.12$ (Mühle (2022)). In a bi-modal system, however, some of the demand $\Lambda$ is directed towards trains, therefore, we need to compute an adjusted demand, $\Lambda_{\text{adjusted}} \equiv (E_{\text{shuttle}} D_{\text{shuttle}})/\nu_0$, considering shuttle trips only; $\nu_0$ is the effective request frequency for shuttle trips and $D_{\text{shuttle}}$ is the average distance of a shuttle trip. Bi-modal trips consist of two shuttle trips (from and to the station), therefore

$$\nu_{\text{shuttle}} = \frac{2 \nu F}{\text{twoshuttletrips}} + \nu (1 - F) = \nu (1 + F).$$

(9)

Similarly, the average requested distance for shuttle-borne trips involved in bi-modal transport is

$$D_{\text{shuttle}} = \left( \frac{2 \beta F}{\text{twoshuttletrips}} + \frac{(d)_{d<d_{\text{shuttle}}} (1 - F)}{\text{shuttleonly}} / (1 + F) \right).$$

(10)

where $(1 + F)$ is due to normalization. The bi-modal demand for shuttles is thus given by:

$$\Lambda_{\text{shuttle}} = (E_{\text{shuttle}} D_{\text{shuttle}}) / \nu_0 = \Lambda (1 + F)^{-2} ((1 - F)(d)_{d<d_{\text{shuttle}}} + 2 \beta F)^2.$$

(11)

In simulations we observe a higher efficiency than suggested by $\eta \propto \Lambda^{0.12}$ (see Supporting Information for simulation data). We call this the common stop effect, meaning that pooling gets more efficient because bi-modal requests are spatially correlated due to shared pick-up and drop-off locations, i.e., the train stations. We account for this effect via an empirical function $h(F)$ ($1 \leq h \leq 1.35$, see Supporting Information for details). In particular, we set

$$\eta \equiv \Lambda^{0.12} \cdot h(F).$$

(12)

Line Service and MIV Trains are recurrent every $1/\mu$ time units. Therefore, the cumulative distance driven by all trains in a unit cell of side length $\epsilon$ per unit time is

$$\Delta_{\text{train}} = 4 \cdot \mu \cdot \epsilon^2.$$

(13)

There is a multiplicative factor of 4 because trains go in four directions at every train station. The total distance driven via MIV for requests from the unit cell amounts to

$$\Delta_{\text{MIV}} = \nu E \epsilon^2 D.$$  

(14)

Replacing $\Delta_{\text{shuttle}}, \Delta_{\text{train}},$ and $\Delta_{\text{MIV}}$ in Eq. 7 from Eq. 8, 13, and 14 we obtain the final expression for the energy consumption of bi-modal transit normalized with respect to MIV

$$E = \frac{\eta^{-1} ((d)_{d<d_{\text{shuttle}}} (1 - F) + 2 \beta F)}{\text{shuttle}} \cdot \frac{e_{\text{shuttle}}}{e_{\text{MIV}}} + \frac{4 \mu}{\Lambda^2} \cdot \frac{e_{\text{train}}}{e_{\text{MIV}}}.$$

(15)

5. Results

We now analyze how the objectives, i.e., energy consumption (Eqs. 7, 15) and quality (Eqs. 4, 6), can be optimized by choice of parameters of operation, i.e., cutoff distance $d_c$ and train occupancy $\alpha$, under different ‘external’ conditions, $\Lambda$ and $\epsilon$. Notice that the two control parameters, $\alpha$ and $d_c$, enter the objectives, $Q$ and $E$, via $(d)_{d<d_{\text{shuttle}}}, F(d_c)$, and $\mu = \bar{\mu}(d_c)/\alpha$ (Eq. 3).

In Fig. 3, energy consumption, $E$, and service quality, $Q$, for the combined system are shown as a function of the share of bi-modal transport $F(d_c)$ at $\bar{\epsilon} = 0.8$ for three different values of dimensionless demand, $\Lambda$. Trains are operated at full occupancy, i.e., $\alpha = 1$. The general trend of reduction of energy consumption with increasing demand and involvement of line services is obvious from the data for $E$. We encounter a minimum of energy consumption at around $F \approx 0.6$ for all values of $\Lambda$ investigated. Energy consumption can be less than 30% of MIV for sufficiently large (but realistic, see Table 1) demand. For service quality, we find that typical values are around one half the service quality of MIV, i.e., about twice the travel time. This is customary for public transport systems (Liao et al., 2020; Salonen and Toivonen, 2013b) and generally well accepted by users. Note that our data for service quality represent a safe lower bound, as the (sometimes quite substantial) time required for parking spot search (Chaniotakis and Pel, 2015; Fulman and Benenson, 2021) is neglected here in $\nu_0$. While for low $\Lambda$ the involvement of line service seems to generally increase travel times (i.e., reduce service quality), we find an optimum at $F \approx 0.25$ for large $\Lambda$. The primary message from Fig. 3, however, is that minimizing energy consumption and maximizing service quality cannot be simultaneously achieved.
Fig. 3. Bi-modal performance characteristics. Energy consumption $\mathcal{E}$ and service quality $Q$ for bi-modal transport, normalized with respect to MIV, as a function of the bi-modal fraction $F(d_c)$, for three different values of demand $\Lambda = \{10^2, 10^3, 10^4\}$. All data for $\tilde{\ell} = 0.8$ and fully-occupied trains, $\alpha = 1$.

Fig. 4. Emergence of Pareto fronts and effects of train occupancy. (a) Grey circles: admissible data for full variation of $\alpha$ and $d_c$, at $\tilde{\ell} = 0.2$ and $\Lambda = 10^3$. Black curve: Pareto front, i.e., the boundary of the full data set towards optimality (low $\mathcal{E}$ and high $Q$). The slope of the dashed tangent to the Pareto front represents the ratio of valuations (see text). (b). Black curves: Pareto fronts for variable train occupancy $\alpha < 1$ at $\Lambda = \{10^2, 10^3, 10^4\}$ and $\tilde{\ell} = 0.2$. Grey dashed curves: degenerate Pareto fronts obtained at full train occupancy ($\alpha = 1$) at corresponding values of $\Lambda$. Grey circles mark the ends of these fronts which are determined by minimum achievable energy consumption and maximum achievable service quality, specific to $\Lambda$ and $\tilde{\ell}$. (c) Same as (b) but for $\tilde{\ell} = 0.8$.

5.1. Pareto fronts in energy consumption and service quality

A tuple of parameter values, in our case $(\mathcal{E}, Q)$, is called Pareto optimal if none of the parameters (or objectives) can be further optimized without compromising on at least one of the others. The set of all such tuples of parameters is called the Pareto front. To illustrate this concept, in Fig. 4a, we show all values $(\mathcal{E}(d_c, \alpha), Q(d_c, \alpha))$ obtained for different values of $d_c$ and $\alpha$ as grey dots. The solid black line represents the Pareto-optimal subset, i.e., the Pareto front.

In order to choose the truly optimal point on the Pareto front, one needs to define the relative valuation of the objectives, $\mathcal{E}$ and $Q$. In other words, the authorities operating the system have to decide how much reduction in service quality they (i.e., the users) would be willing to accept for how much savings in energy. The ratio of these valuations is then expressed in the slope of the dashed line in Fig. 4a, which is a tangent to the Pareto front. The point where it touches the front (black dot) represents the optimal set of parameters, under the given valuation.

For our analysis, we fix the train capacity to $k = 100$, and choose a representative set of values for line service mesh size $\tilde{\ell} = \{0.2, 0.4, 0.8\}$ and demand $\Lambda = \{10^2, 10^3, 10^4\}$, corresponding to a typical parameter range encountered in real systems (see Table 1). Note that fixing $k$ does not reduce the generality of our study, because a different $k$ can be compensated for by properly readjusting $\mu_0$ (see Eq. 3).
Fig. 5. Bi-modal performance with fully-occupied trains. (a) Black curves show degenerate Pareto fronts for fully-occupied trains (α = 1) for varying demands \( \Lambda = \{10^2, 10^3, 10^4\} \) shown as annotations and \( \bar{\ell} = 0.8 \). Black circles mark the end points of the Pareto fronts which are determined by the minimum achievable energy consumption and the maximum achievable service quality. Grey curves show the entire data, i.e., not only the Pareto-optimal set, but all admissible values with \( d_e \) as the control parameter. (b) Data as in (a), but normalized with respect to the performance, \( (Q_0, \xi_0) \), of the uni-modal system (shuttles only). (c) Degenerate Pareto fronts as in (a), but for \( \bar{\ell} = \{0.2, 0.4, 0.8\} \) in blue, orange, and black, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 b demonstrates the effect of \( \Lambda \) on the overall performance of the system. Pareto fronts are shown in black. We see that for typical values of \( \Lambda \), as listed in Table 1, energy consumption reaches down to well below 40% (even below 20%) of MIV for large values of \( \Lambda \). At the same time, quality levels are comparable to, mostly even in excess of, what is found in existing public transport in terms of transit time (see Table 1). Note, however, that our system even provides on-demand door-to-door service, comparable to MIV.

The dashed grey curves indicate the subset of data for train occupancy \( \alpha = 1 \). We will henceforth call them degenerate Pareto fronts, as they correspond to the variation of only one parameter. They are lying, slightly but consistently, above the Pareto fronts. This indicates that by reducing train occupancy below its maximum (\( \alpha < 1 \)), i.e., operating trains at higher frequency than necessary, one can enhance the overall performance of the system. Service quality increases because the time spent waiting for trains, which is proportional to \( 1/\mu \), is smaller when trains are operated more frequently. The increase in service quality is found to overcompensate the increase in energy consumption due to higher operation frequency. Since waiting time is inversely proportional to both \( \Lambda \) and \( \bar{\ell} \) (see Eq. 3), this effect is more pronounced for small \( \Lambda \) and \( \bar{\ell} \), which is qualitatively corroborated in Fig. 4c which shows corresponding data for large mesh size (\( \bar{\ell} = 0.8 \)). At the resolution of the figure, the Pareto fronts (black) and their degenerate partner (dashed grey) are distinguishable only for smallest values of \( \Lambda \). Hence for typical values of \( \Lambda \) and \( \bar{\ell} \) it appears sufficient to discuss the degenerate Pareto fronts, which only need one parameter (\( d_e \)) to be varied. We keep in mind that the true Pareto optimum will still be superior.

Degenerate Pareto fronts are shown in Fig. 5 in various presentations. Fig. 5a has basically the same information as Fig. 4c, but here we show the full data set (grey), where the black curves are the Pareto-optimal subset, i.e., the degenerate Pareto fronts. They terminate wherever the tangent becomes either vertical or horizontal (black dots), thus offsetting any tradeoff between the objectives.

Fig. 5 b shows these data normalized with respect to the performance of a uni-modal (shuttles only) DRRP system, represented by \( \xi_0 \) and \( Q_0 \). Clearly in relevant parameter regions the combined bi-modal system outperforms the uni-modal system in both energy consumption (\( \xi < \xi_0 \)) and service quality (\( Q > Q_0 \)).

The effect of inter-station distance (mesh size) on the (degenerate) Pareto fronts is explored in Fig. 5c. We see that a dense network of rails (small \( \bar{\ell} \), blue fronts) achieves the best results concerning energy consumption, reaching down to below 20% of MIV, but compromises on achievable service quality. For admissible quality \( Q \approx 0.5 \), sparse train networks (black fronts) are Pareto-optimal for low demands. For larger demands, denser train networks (orange and blue fronts) are advantageous.

However, it is remarkable that the overall position of the Pareto fronts in the plane spanned by \( Q \) and \( \xi \) does not vary dramatically with mesh size, as the position on the front at which the system is operated is largely at the discretion of the operator. This suggests that the density of currently installed rail track systems might already be well suited for deploying a bi-modal on-demand transport systems of the kind we have studied.

5.2 Traffic volume

Energy consumption and service quality are not the only possible objectives for optimization of public transport. Road traffic, for example, is a source of noise and local air pollution and occupies significant shares of urban space. Bi-modal ride-pooling reduces traffic by use of shared shuttles, and by directing certain trips towards trains. We quantify this reduction by introducing the relative bi-modal traffic \( \Gamma \) as the ratio of the number of on-road vehicles necessary for bi-modal transportation (i.e., shuttles) to the number
of MIV (i.e., cars) needed to serve the same demand. We have (see Supporting Information for details)
\[
\hat{\Gamma} = \eta^{-1} (1 + F) \hat{D}_{\text{shuttle}}. 
\]  
(16)

In Fig. 6, \(\hat{\Gamma}\) is plotted against service quality \(Q\) as determined along the degenerate Pareto fronts shown in Fig. 5c. For low demand (\(\Lambda = 10^2\)), uni-modal (shuttles only) transportation allows for about 50% reduction in traffic as compared to MIV (triangle). Bi-modal transportation allows for further reduction in traffic at the cost of service quality. For intermediate and high demand (\(\Lambda = (10^3, 10^4)\)), uni-modal transportation allows for about 70% to 80% reduction in traffic as compared to MIV (square, circle). In these cases, bi-modal transportation allows for truly dramatic reductions in traffic (> 90%), at equal or even higher service quality than for uni-modal transport. Combining this finding with typical parameter values in Table 1, we recognize that even in rural settings, traffic volume is expected to decrease by an order of magnitude, relative to MIV. In urban environments, traffic volume may even be reduced by more than a factor of ten.

6. Discussion

The goal of this study was to find whether and under what circumstances bi-modal transport, i.e., on-demand ride-pooling with shared shuttles combined with fixed schedule line services (railway), can be a viable alternative to customary public transportation (line services or DRRP alone) or MIV. For that purpose we introduced a simple model system for bi-modal transport, combined with a mean-field approach, which allowed us to parameterize the user environment (dimensionless demand) as well as the bi-modal service operations (cutoff distance, train occupancy) with few variables, and to write down analytic expressions for key performance characteristics, namely energy consumption and service quality (i.e., transit times), as well as road traffic volume. Our results, in form of Pareto fronts, indicate that bi-modal public transportation systems have the potential to provide on-demand door-to-door service with a quality superior to customary public transportation, while at the same time consuming only a fraction of the energy a corresponding fleet of MIV would require, and with a road traffic volume reduced by an order of magnitude.

In our model, we assume customer demand to be given, more precisely, we assume a spatially uniform demand with constant average request frequency. However, in realistic scenarios, demand will be heterogeneous in space, due to non-homogeneous population density and individual mobility patterns, and fluctuating in time due to phenomena like rush hours or workdays versus weekends. Such spatio-temporal demand patterns can be taken into account by structured railway networks (e.g., more dense in highly-populated areas), spatially varying shuttle densities, and variable service frequencies. In future case studies or real-world applications, these context-specific adjustments can be implemented to provide tailored solutions.

Although demand patterns can be estimated from historical data, there will be fundamental uncertainties in predicting future demand. Such uncertainties will alter the performance of a bi-modal system. Nevertheless, our considerations based on known (average) transportation demand give a valuable estimate of the performance potential of bi-modal transport under various external conditions.

Moreover, in reality, demand will couple to service quality and other parameters concerning user satisfaction; a well-functioning system of transportation attracts demand. Since service quality depends on demand, too—as seen in this study—modeling user
behavior introduces a feedback loop. In a forthcoming study we will integrate this feedback loop into our modeling framework, a step towards resolving user adoption dynamically. Broader analysis of user adoption of bi-modal transportation based on incentives, customer convenience, and customer preferences remains to be explored. Recent results from experimental social science, however, point into a very favorable direction (Avermann and Schlüter, 2019; Nyga et al., 2020; Sörensen et al., 2021).

It is important to note that our results, represent only a lower bound on bi-modal performance, in particular as far as service quality is concerned. First, we have based the decision on the type of transport service (uni-modal or bi-modal) on a single scalar parameter, \(d\), which amounts to representing the decision process by a binary-valued function of a single scalar variable, \(d \rightarrow \{0, 1\}\). The true structure would be a binary-valued field \(\Phi\) on the four-dimensional space of the pick-up and drop-off coordinates, \(\Phi : (x, y, x', y') \rightarrow \{0, 1\}\). This would be extremely cumbersome to study in a statistical manner. However, in a real system, data on \(\Phi(x, y, x', y')\) are being collected on a daily basis, such that over time the system can be ever improving its performance over the data we have presented here.

Second, bi-modal service of the kind studied here provides door-to-door service, while MIV involves the search for parking space, which was disregarded in our study due to lack of reliable data. This can be quite significant, and fully adds to the MIV transit time, thus further improving on the relative service quality, \(Q\), of bi-modal service.

Moreover, it should be mentioned that riding a bi-modal transport system neither involves having to drive nor taking care of vehicle maintenance.

In summary, bi-modal public transport systems have the potential of outperforming customary public transportation and the private car in a number of ways. With this study, we provided broad fundamental evidence and laid the foundation for taking the idea of bi-modal public transit systems to real-world applications.

### Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

The authors would like to appreciate fruitful discussions with Tariq Baig-Meininghaus, Michael Patschke and Prakhar Godar

### Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.multra.2023.100083

### References


