High-Throughput and Low-Latency Hyperloop

Manuel Eichelberger, David T. Geiter, Roland Schmid and Roger Wattenhofer
{manuelei,dgeiter,roschmi,wattenhofer}@ethz.ch
ETH Zürich, Switzerland

Abstract — Hyperloop pods are expected to travel faster than 1,000 km/h. Apart from high speed, high throughput and low latency are crucial to hyperloop's success. We show that hyperloop networks could transport as many passengers as train or plane networks. Our on-demand pod scheduling method provides passenger waiting times of only a few minutes, even at peak times. That minimizes the overall trip latencies. Further, our scheduling results in low resource usage in terms of consumed energy and required number of pods in the system.

With on-demand scheduling, passengers need not look up schedules and cannot miss connections. Rather, the schedule follows passengers' itineraries. In addition, the hyperloop concept can enable many direct connections due to small pod capacities.

We conclude that hyperloop systems may become the preferred mode of transportation by being fast, reducing waiting times and keeping up with high demand — all while offering more convenience than current public transportation.

Index Terms — feasibility, modeling, on-demand, scheduling, waiting time, transportation

I. INTRODUCTION

Commercial planes carry more than four billion passengers each year and cruise at speeds of about 900 km/h [1]. While this is fast, planes do have several drawbacks. For instance, (i) planes take about half an hour to climb to a cruise height of 10 km and slow down when descending for landing, resulting in a low average speed for short flights [2]; (ii) planes have fixed schedules, for example leaving 4-8 times per day, which may not fit some passengers' itineraries and may thus result in long waiting times; (iii) tickets are normally booked in advance, making trips inflexible and necessitating some buffer time on the way to and through the airport in order to make sure not to miss the booked connection; (iv) planes are large, requiring many passengers to make the same trip to be economical, which limits the number of direct connections; (v) boarding many passengers through few doors is slow; and (vi) planes consume a lot of energy, which negatively impacts the environment.

While trains solve the Problems (iii), (v) and (vi), they share the Problems (i), (ii) and (iv).

One proposal to overcome all those drawbacks is to build a hyperloop system [3]. Hyperloop is a high-speed rail system using evacuated tubes to minimize air drag. The hyperloop vehicles, so-called pods, are expected to carry a few tens of passengers at speeds over 1,000 km/h [3]. That speed beats any other low-cost transportation, including cars, trains and planes: A 100 km long hyperloop trip will only take a few minutes. Further hyperloop advantages are: With light pods and low air drag, the maximum speed is quickly reached (i). Small pods enable on-demand scheduling and many direct connections because pods can quickly be filled with passengers for the same destination (ii)–(v). Direct connections also eliminate transfer times between vehicles. Small pods minimize boarding times (v). Due to traveling on or near the surface, no time or energy is wasted ascending to a cruise height (vi). Hyperloop systems also have drawbacks. For instance, the operation principles and economics are still unknown. Also, current technology does not reasonably allow building tunnels for a hyperloop track crossing oceans. However, some ideas exist for underwater tubes [4].

A. Hyperloop Viability

Due to the low pod capacity, a central question concerning the usefulness and economic viability of hyperloop systems is whether a hyperloop network can sustain high passenger volumes. Even some train operators struggle to fit more vehicles on their tracks while maintaining sufficient safety distances between vehicles. Since pods have a lower capacity than trains, the same number of vehicles carries fewer passengers. Therefore, to reach or surpass the rail passenger throughput, many pods need to be accommodated in a hyperloop network. As building costs are a crucial factor for hyperloop’s success, we consider using additional tracks not a viable option.

In this paper, we present a pod scheduling method for hyperloop networks which easily matches the passenger throughput of air and even railroad transportation, but with minimal departure delays. Even at peak times, queues do not exceed a few minutes. To show that, we model the passenger demand throughout a typical day, relying on data from rail and air transportation. Our scheduling is on-demand, that is, passengers do not need to book or look up schedules in advance. For that, we solve the following challenges.

B. Challenges

Online Scheduling: Passenger arrival times and destinations are unknown before passengers start waiting to board a pod. That random nature of the trip requests makes it hard to find a good trade-off between passenger waiting times and resource consumption: In order to limit the waiting time, some pods need to leave partially filled, which increases the required energy and number of pods. Also, due to the high pod speed compared to existing transportation such as cars or trains, a fixed waiting time constitutes a larger fraction of the total trip time in hyperloop travel.

*The authors of this paper are alphabetically ordered.
Pod Balancing: Scheduling should ensure continuous pod availability at all stations. Without a pod balancing mechanism, an unbalanced trip demand, for instance caused by commuters travelling from suburbs to the city in the morning and returning in the evening, causes problems. First, stations need numerous spare pods in case many passengers suddenly arrive. Second, popular destinations need a huge space to store incoming pods until they might eventually return.

C. Contributions

We introduce a pod scheduling mechanism ensuring both low waiting times and perpetual pod availability at all stations while requiring few pods. Our Balanced Departure Bin (BDB) scheduling limits the accumulated waiting time of all passengers waiting for a specific trip. To ensure pod availability, we argue that sending pods symmetrically is efficient and allows for on-demand scheduling.

We compare the BDB mechanism to two baselines concerning the waiting times and the consumed energy for moving the pods (Sec. IV). Our evaluation considers hyperloop networks at different scales, one with sections of a few hundred kilometers each and one with sections of several thousand kilometers. We model passenger demands based on real-world rail and air travel statistics. Our results show that BDB scheduling achieves high throughput, serving the same demand as train and plane networks combined, while passengers can simply walk into a station and expect waiting times of a few minutes at most. BDB scheduling uses just slightly more energy and pods compared to a scheduling mechanism that optimizes for one of those resources only.

II. RELATED WORK

A. Vacuum Trains

The fundamental problem addressed by hyperloop is that the main speed limitation for vehicles close to the ground is air drag. Already in 1904, Robert Goddard proposed the 
\textit{vactrain}, consisting of cars in a steel vacuum tube, floating through electromagnetic levitation and being accelerated as much as bearable for passengers [5]. So, why do hyperloop systems not yet exist? Reasons may be their high initial costs and uncertainties regarding their passenger capacities [6], [7], [8]. For instance, in 2013, Elon Musk, who coined the term \textit{hyperloop}, proposed a system with a throughput of only 840 passengers per hour [3]. In this paper, we show that high throughput and low waiting times can be achieved while filling up most pods, thus keeping running costs low.

B. Scheduling

To the best of our knowledge, the achievable throughput and latency (i.e. factors impacting waiting times) of hyperloop networks have not been studied specifically. While several works estimate passenger capacities ranging from 840 to 3,360 passengers per hour, it is commonly argued that the throughput could be increased by enlarging the pod size [3], [9], [10]. Also, the need to keep a safety distance between consecutive pods that allows to come to a full stop without crashing has been considered [6]. Apart from that, scheduling is generally overshadowed by the technical challenges that need to be resolved in order to build an actual hyperloop system. However, both throughput and latency impact the viability of a hyperloop system. In this paper, we show that a hyperloop system can cope with high passenger demands without needing large pod sizes. This allows preserving flexibility in the system.

In contrast, there is an extensive body of work on the general topic of scheduling. For instance, airlines schedule their aircraft, crews, take-off and landing slots at airports and consider popular flight connections (i.e. the customer demand); each imposing various constraints on the resulting flight plan [11], [12], [13]. However, since the hyperloop system, which we study, is autonomous and on-demand, it has different constraints than airline scheduling.

We find closer relations to the problem of train scheduling; both train and hyperloop systems operate on an inflexible network structure and will experience similar demand distributions throughout the day, including commuters and leisure travelers. While trains commonly operate on fixed schedules, we suggest that a hyperloop system should be offered as an on-demand service that allows customers to walk into a hyperloop station and wait for a few minutes in the worst case. What seems like a bold promise at first sight actually bears the opportunity to make the scheduling simpler and more efficient. Train schedules aim to be periodic to make them easier to remember for the customers and often include time buffers to increase the stability of the schedule [14], [15]. Both measures complicate the computation of an optimized schedule. Furthermore, the problem of train re-scheduling in response to local disturbances has been investigated largely [16], [17], [18], [19], but this is not an immanent problem to an on-demand hyperloop system which needs not converge back to a fixed schedule.

Other on-demand transportation has also been studied, for instance bus scheduling [20], [21] and autonomous vehicle services [22], [23]. But due to more rerouting options in (usually dense) road networks, the relation to hyperloop scheduling is only distant. Since such systems exist already, they can be analyzed both empirically and by simulation [24], [25], while hyperloop systems have yet to be built. We thus evaluate our hyperloop scheduling through simulations.

As a theoretic foundation of our work, we study algorithmic models for on-demand services, namely \textit{online scheduling} algorithms. We draw inspiration from the \textit{preemptive service algorithm} (PSA) presented in the context of the \textit{(generalized) online service with delay (OSD)} problem [26]. In the OSD problem, requests are received in an \textit{online} manner, that is, without prior knowledge about when and where they will appear. Special about the OSD problem is the option to postpone serving a request, which causes a delay penalty. Nonetheless, the PSA schedules a set of \textit{k} servers such that all requests are served quickly. The idea of the PSA is to accumulate the delay penalty cost until a threshold is reached and an action is triggered. We employ this technique in our hyperloop scheduling strategy. Finally, the scheduling of pods is similar to the \textit{k}-taxi problem [27]
Fig. 1: Air lock for a platoon of up to four pods. Doors minimize the amount of pumped air, allowing for fast cycles.

Fig. 2: Pods can switch tubes at stations without passing through an air lock.

with more than one passenger per vehicle and the option to delay pod departures.

III. SCHEDULING

A. System Model

Pod Separation: The minimum distance between pods is determined by (1) a potential safety distance to limit the impact of emergencies such as pods getting stuck, and by (2) the air lock capacities to insert and remove pods into and from tubes. Whether safety distances are required is debatable: From an engineering perspective, sensor data monitoring the tube and the pods can travel at the speed of light, so pods can react quickly. The only limiting factor is the maximum deceleration bearable for passengers. Previous work assumes a 30s distance between pods at a maximum deceleration of 1 g [6]. However, it is possible to form platoons consisting of multiple pods attached to each other, like trains. Platoons allow overcoming the throughput limit imposed by safety distances.

Air Locks: To form platoons, multiple pods may go through air locks simultaneously. One option is to use parallel air locks and connect pods in the evacuated space. Another possibility is to use long air locks, which can fit a platoon. Since platoon lengths will vary depending on the passenger demand, the length of such an air lock should be adaptive, for example, by means of automatic doors as depicted in Figure [1].

Connections: Having direct tubes between any pair of stations in the hyperloop network is unreasonable due to high construction and evacuation costs. But since hyperloop travel will be fast, changing pods at stations would substantially prolong trips, mostly due to time required for passing through air locks and for changing platforms. Instead, since pods are small, pods can be filled with passengers for the same destination and travel multiple hops without passing through an air lock. To this, pods are routed through an evacuated section around a station, as sketched in Figure [2]. A platoon is then formed with other pods going in the same tube.

Pod Capacity: In this paper, we assume that each pod has a capacity of 28 passengers. While our scheduling method is independent of the pods’ capacity, it is designed for a small pod size compared to the passenger demand.

Pod Speed: For good passenger comfort, we assume that pods accelerate and decelerate at a maximum rate of 0.5 g. This is in line with previous work which allows up to 1 g [6], [3]. We assume a pod top speed of 1,220 km/h.

B. Scheduling Mechanism

Pod scheduling is the task of picking pod departure times. Scheduling should provide low waiting times, especially for short distance trips, while using little energy and few pods. Energy can be minimized by filling up pods with passengers, while the number of required pods can be reduced through redistribution of the available pods in the system. However, redistribution requires that non-full pods must be sent at times. Further, to deal with unknown passenger arrival times, pods always need to be available at each station.

We present the balanced departure bin (BDB) scheduling mechanism, limiting the waiting time per customer, and keeping the pod distribution balanced. We optimize along two dimensions. First, we select a mechanism triggering the next pod departure depending on the passengers’ number and waiting time. Second, we may employ pod redistribution.

Departure Trigger: Our departure bin (DB) trigger criterion is adaptive, considering the accumulated waiting time of the passengers assigned to a pod. Naturally, a pod departs at the latest when it is full. Given the online setup in which we do not know when further passengers for the same destination will arrive, we set a waiting time threshold for the departure to avoid passengers getting stuck and maintain customer satisfaction.

Inspired by the preemptive service algorithm [26], we assign a virtual departure bin to each pod. Furthermore, let \( \delta(t) \) be a non-negative, non-decreasing delay penalty function. At every time step (e.g. every 20s), we evaluate \( \delta(t - t_i) \) for each passenger \( i \in p(t) \), where \( t_i \) is the arrival time of passenger \( i \) at the pod and \( p(t) \) is the set of waiting customers in the pod at time \( t \), and accumulate the values \( \delta(t - t_i) \) in the bin:

\[
\text{bin}(t) = \sum_{\tau=0}^{t} \sum_{i \in p(\tau)} \delta(\tau - t_i).
\]

Departure is triggered as soon as the departure bin is full, that is, when it exceeds some threshold. Hence, the combination of bin size and delay penalty function \( \delta \) allow setting the maximum waiting time. For instance, with \( \delta(t) = 1 \) and a bin size of 30, the passengers in a pod wait at most an accumulated total of 10min until departure (see Fig. [3]). The delay penalty function also allows different weighting of longer versus shorter waiting times (e.g. waiting 5min can have 0 penalty, as it might not bother customers at all).

Rearranging the formula and making it continuous yields

\[
\text{Bin}(t) = \sum_{i \in p(t)} \int_{\tau=0}^{t-t_i} \delta(\tau) \, d\tau.
\]

accumulated delay penalty for passenger \( i \)
So, the waiting time $t_{\text{max}}$ for a single passenger may be limited by setting $\delta(t)$ so that $\int_0^{t_{\text{max}}} \delta(\tau) d\tau$ equals the bin size.

**Pod Redistribution:** We present a simple, yet powerful pod redistribution mechanism, which we call symmetric redistribution: Between two stations $A$ and $B$, whenever a pod departs from $A \rightarrow B$, we send another pod from $B \rightarrow A$. While this mechanism may seem trivial and wasteful (w.r.t. consumed energy), it is well-suited for a hyperloop system: Energy costs to move the pods are low and symmetric redistribution fits the online scenario as it makes no assumptions on future passenger demands, but maintains a perfectly balanced pod distribution.

While an online algorithm is usually analyzed with respect to a worst-case scenario, that could mean that all passengers wanted to depart from a single station with a small pod stock. Symmetric rebalancing qualifies as a 2-competitive algorithm regarding the energy cost, i.e. uses at most twice as much energy as needed by an optimal scheduling. But the waiting times could inflate due to an exhaustion of the available transport capacity. In practice, the overall demand for a hyperloop system can be estimated from train and air travel statistics. Furthermore, demand depends on the population density of a connection’s endpoints. Finally, for any track segment, we do not need more pods in stock than the air locks can introduce into the system before the first return pod arrives from the destination. In sum, it should be possible to assign pods to stations based on historical data, possibly adapting it regularly as the demand is observed. In this paper, we assume a fixed demand model (see Sec. IV-B) and evaluate the scheduling methods with regard to how possible it is to assign pods to stations based on historical data, possibly adapting it regularly as the demand is observed. In this case, both pods depart at stations $A$ and $B$ for the respective other station as

$$Bin_{A\rightarrow B}(t) = \max\{Bin_A(t), Bin_B(t)\}$$  \hspace{1cm} (1)

with the same bin size and delay penalty $\delta(t)$ as before. So, as soon as one of the bins (or pods) is full, both pods depart.

Eq. (1) is not the only choice for combining the two separate departure mechanisms. For instance, the triggers could be added, i.e. $Bin_{A\rightarrow B}(t) = Bin_A(t) + Bin_B(t)$. In that way, the mechanism would directly relate waiting time to energy costs: For each empty seat in the two pods, whose departure is triggered together, a passenger actually has to wait longer. However, we choose to maximize customer satisfaction by setting the waiting time limit for a single pod, so passengers do not have to wait because of empty seats in the pod coming from their destination.

**IV. EVALUATION**

To evaluate our scheduling algorithm, we simulate traffic in two hyperloop networks, shown in Figure 4. The larger one connects all European cities with more than one millions inhabitants and the smaller one connects all Swiss cities with more than 50,000 inhabitants. The thresholds are set on the city agglomeration population.

**A. Simulation Parameters**

The maximum number of pods per platoon is 2 and the departure rate is limited by the throughput of the vacuum chambers, which we set to one platoon per minute. That is, a platoon can depart every minute in each direction from every station. The simulation is discretized in 20 second time steps.

**Energy:** We only discuss the energy used to move pods between stations. Scheduling cannot optimize the energy costs to keep the tubes evacuated, which is fixed for a given target pressure. In line with previous work, we assume a tube air pressure of 100 Pa (= 1 mbar) [3]. When computing the energy to move pods, we account for the remaining air drag, assume an energy recuperation fraction of one third when pods decelerate and employ a motor efficiency of 0.8. Those numbers assume that the hyperloop system will be implemented with Maglev technology using linear synchronous motors [28], [6].

**B. Demand Model**

The passenger demand is defined by the requested trips, consisting of an origin, a destination and a departure time. Our simulation spans a day during which the trips are distributed over different routes and over time as follows.

**Routes:** The demand for a route is a function depending on the travel time for that route and the populations of the origin and destination. First, we set the number of potential travelers in each city as the sum of 8.1 % short range travelers according to some Swiss rail statistics and 7,233 long range travelers according to some average statistics of daily plane passengers per airport in Europe [31], [30]. Those potential travelers of each city are then distributed to the demand to each destination from that city. The distribution is weighted based on the travel time to the destinations, as shown in Figure 5. The travel time calculation for a route considers acceleration and deceleration phases and includes one minute for each station at which the pods have to change the tube. Last, to incorporate that more people travel from smaller to larger cities, e.g. for work, those demands are adjusted as follows. First, for each city pair, the larger demand of the two directions is reduced to the smaller one. Then, the demand from the larger city to the smaller one is multiplied by $\sqrt{\frac{\text{population}_{\text{large}}}{\text{population}_{\text{small}}}}$. As
Fig. 4: We use a small and a large hyperloop network for our evaluation.

(a) Small network for Switzerland.

(b) Large network for Europe.

Fig. 5: Passenger demand distribution according to city distance (time). The demands are modeled as negative binomial distributions approximating German commuter and European air passenger travel data and their sum [29], [30].

Fig. 6: Demand distributions over a day. Morning (evening) demand is the probability for one passenger to arrive in the morning (evening) in the corresponding 30min interval.

explained next, those trips are performed during the morning and each such "passenger" travels back in the evening.

Departure Time: In our simulation spanning a day, each passenger commutes. That is, the number of trips in each direction of a city pair equals out over the whole day, leaving the city populations unmodified. We arrange the demand in two “rush hours”, modeled as normal distributions with peaks at 8:30 and 17:30. The evening (return) demand is more widely spread than the morning demand, with $\sigma = 1.35 \text{ h}$ and $\sigma = 2.5 \text{ h}$, respectively. The two curves, multiplied with the number of travelers in a city, model the distribution of the total demand, shown in Figure 6. The passenger arrival times are sampled from the resulting combined distribution.

C. Baselines

To compare scheduling methods, we evaluate three metrics: kinetic energy for moving pods, passenger waiting time, and the required number of pods. First, we give some simple baselines, whose results are given in Table II.

<table>
<thead>
<tr>
<th>Departure trigger</th>
<th>Energy Baseline</th>
<th>Waiting Time Baseline</th>
<th>BDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric redistribution</td>
<td>full pod</td>
<td>no</td>
<td>waited 1min</td>
</tr>
</tbody>
</table>

1) Kinetic Energy: Neglecting the passengers’ weight, the system’s required kinetic energy is proportional to the number of pod trips. We can give a lower bound regarding the required kinetic energy to transport all passengers, by sending only full pods (except for the last one per route). For this baseline, we assume unlimited pod availability and unrestricted waiting times. Further, platoon departure rates are not limited and air drag is omitted. The latter models the ideal case that all pods travel in a single long platoon, thus saving the air drag of all but one pod. Table II shows that only minimizing energy results in a huge maximum waiting time of 65 minutes.

2) Waiting Time: For a waiting time baseline, we simulate a “best case” scenario for passengers by limiting the waiting time to one minute. Since our air locks are restricted to one departure every minute, this means that every minute, all waiting passengers depart immediately. This means that one pod per destination can be partially filled. Also this baseline assumes unlimited pod availability.

D. Dynamic Scheduling

The baselines above illustrate the trade-offs involved in the pod scheduling. Ideally, our proposed balanced departure bin (BDB) scheduling method should get close to the baselines for both energy and waiting time, while using few pods.

The priority of our scheduling are low waiting times. Our parameters for the DB trigger are $\delta(t) = 0.005/s \cdot t [t \text{ in s}]$, and the bin sizes are 700 and 7,500 for the Swiss and European networks, respectively. Note that the bin is only filled every 20s, as that is our simulation discretization. The maximum waiting time for a passenger is 15min. The second priority is a low number of required pods, which we reduce at some additional energy usage for moving non-full pods.

As Table II shows, the BDB scheduling requires only about twice the energy as the theoretical energy baseline;
while reducing the maximum waiting time from 65 to 15 minutes. The average waiting time is a few minutes. This is similar to the most frequented transports like subways or urban trains, while easily beating long-range trains that seldom depart more often than every 30min.

Symmetric redistribution prevents any pod imbalance, limiting the number of required pods at stations with high pod outflow and low pod inflow during some time, e.g. the morning rush hour. Table II confirms that the number of pods is indeed lower than for the baselines.

V. CONCLUSION

Hyperloop systems with small pods can achieve a high throughput and low waiting times. Our scheduling scales to any passenger throughput, up to the tube capacities, with a proportionally higher number of pods and air lock chambers.

Small pod capacities allow for on-demand operation and enable many direct connections, saving some of the time overhead of other public transportation such as trains or planes.

REFERENCES


TABLE II: Simulation statistics for a day (6:00-24:00) with the different scheduling configurations.