Transportation Research Part B 181 (2024) 102892

Contents lists available at ScienceDirect

Transportation Research Part B

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

A bathtub model of transit congestion

Lewis J. Lehe *, Ayush Pandey

Urban Traffic and Economics Lab, Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, United States

ARTICLE INFO

Keywords:
Congestion
Stability
Economics
Bathtub model
Transit
MFD
Equilibrium

ABSTRACT

Studies of transit dwell times suggest that the delay caused by passengers boarding and alighting rises with the number of passengers on each vehicle. This paper incorporates such a "friction effect" into an isotropic model of a transit route with elastic demand. We derive a strongly unimodal "Network Alighting Function" giving the steady-state rate of passenger flows in terms of the accumulation of passengers on vehicles. Like the Network Exit Function developed for isotropic models of vehicle traffic, the system may exhibit hypercongestion. Since ridership depends on travel times, wait times and the level of crowding, the physical model is used to solve for (possibly multiple) equilibria as well as the social optimum. Using replicator dynamics to describe the evolution of demand, we also investigate the asymptotic local stability of different kinds of equilibria.

1. Introduction

1.1. Background

Daganzo (2007) discusses a class of systems which may be treated as "reservoirs...where specific items flow into the system, spend some time in the system and then flow out". The items' locations within the reservoir are not of first importance, but rather their status "in" or "out" of it. In turn, Daganzo characterizes reservoir systems as being of three types, which we can formalize by letting n stand for the mass of items and $G(n)$ for the rate they flow out. The three types are distinguished by how $G'(n)$ behaves:

(i) $G'(n) = 0$ (i.e., $G(n)$ constant) for all n;
(ii) $G'(n) \geq 0$ for all n;
(iii) $G'(n) < 0$ for large-enough n.

Daganzo gives supermarket checkouts as an example of type (i), actual water reservoirs of type (ii) and, of course, zones of city streets as examples of type (iii). Fig. 1 portrays an example $G(n)$ for a type (iii) system. In literature using reservoir models, states on the left side are called, variously, "uncongested", "under-saturated", or "free flow" and those on the right are called "congested", "oversaturated" or "hypercongested" (Small and Chu, 2003). Below, we use *uncongested* and *hypercongested* and call such type (iii) systems *hypercongestible*. What distinguishes hypercongestible systems is that "items must complete a task before leaving, and crowded conditions reduce the efficiency with which the tasks can be completed" (Daganzo, 2005).

The general idea of treating a city zone as a macroscopic entity is about fifty years old (see Johari et al., 2021 for a review) but since Geroliminis and Daganzo (2008) provided evidence that traffic in urban zones does in fact behave as a hypercongestible reservoir (affirmed in later by many studies such as Buisson and Ladier, 2009; Tsekeris and Geroliminis, 2013 and Mariotte et al., 2020), a large literature has sprung up which takes this approach to study aspects of traffic control. Applications include dynamic

* Corresponding author.
E-mail address: lehe@illinois.edu (L.J. Lehe).

https://doi.org/10.1016/j.trb.2024.102892
Received 21 September 2023; Received in revised form 4 January 2024; Accepted 26 January 2024
Available online 29 January 2024
0191-2615/

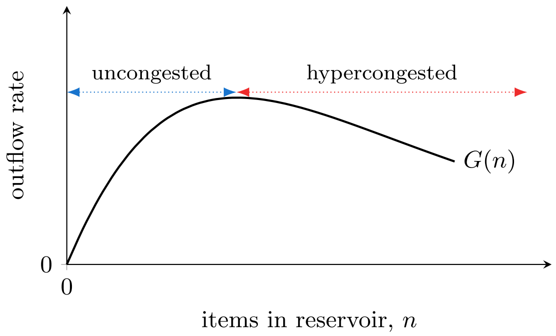


Fig. 1. $G(n)$ for a hypercongestible system.

tolling (Geroliminis and Levinson, 2009; Arnott, 2013), parking (Geroliminis, 2015), perimeter control (Haddad and Geroliminis, 2012), traffic signal timing (Zhang et al., 2013; Daganzo and Lehe, 2016) and many others. Other work has made strides in the theory (Laval et al., 2018; Xu and Gayah, 2023) and simulation (Mariotte et al., 2017) of such models.

Despite being the focus of most research, city streets are obviously not the only system for which "crowded conditions reduce the efficiency with which tasks can be completed". Daganzo (2007) and Small and Chu (2003) cite a messy desktop as an example, with the latter explaining "the bigger the backlog of work, the harder it is to find papers and consequently processing slows down" (p. 332). Earlier studies by Daganzo show the potential for gridlock in airport baggage carousels (Ghobrial et al., 1982) and certain highway networks (Daganzo, 1998). Arnott (1996) modeled dispatch taxi service as a hypercongestible reservoir, using a two-branched "supply curve" which implies a unimodal $G(n)$. While Arnott focused on uncongested states, recent studies extended his insight to analysis of remedies for hypercongested states in ridehailing (Xu et al., 2020; Castillo, 2022; Castillo et al., 2023). Safadi et al. (2023) even extends the framework to low-altitude aircraft.

1.2. Aims of the paper

Models of traffic using a form of unimodal $G(n)$ are sometimes called *bathtub models*[1] after a metaphor coined some time ago by William Vickrey (published later as Vickrey, 2020). This paper develops a bathtub model of a transit route with elastic demand. Passengers are "items". The transit fleet makes up a "reservoir" because the route is *isotropic*: it is the same everywhere and characterized by continuously-varying statistics. Isotropy is a common assumption in the theoretical studies[2] of transit operations.

In turn, we incorporate the physical model into a static economic model with elastic demand. Zhang et al. (2020) does the same and derives—for a given choice of operational variables—a unique equilibrium via feedback between demand and passenger congestion, such as crowding costs and delays imposed by boarding/alighting. The chief difference is that Zhang et al. (2020) uses a *flow-based* congestion technology: travel times and crowding are increasing with the flow of passengers. These physics are akin to volume-delay functions in traffic flow. Our model's physics, on the other hand, are written in terms of the accumulation of passengers and our route may exhibit hypercongestion, such that travel times and crowding may rise or fall with passenger flows. Consequently, equilibrium is not unique.

What makes the route hypercongestible is a stylized fact gleaned from empirical studies of transit operations: when the number of passengers on a transit vehicle (the *load*) becomes high enough, standing passengers get in the way of boarding/alighting passengers and raise vehicle dwell times at stops. See Sec. 2.1 of Tirachini et al. (2013) for a review of various studies. For example, Milkovits (2008) finds bus dwell time rises with the product of (i) the square of the number of standing passengers; (ii) the number of passengers boarding/alighting per stop. From this fact, we derive a strongly unimodal function giving steady-state passenger flows (and other variables, such as headways) in terms of the fleet size and the total accumulation of passengers on all vehicles. We call this the "Network Alighting Function" to highlight the correspondence to the "Network Exit Function" (sometimes called a "Production MFD") used in bathtub traffic models (Daganzo et al., 2012).

The main goals of the paper are threefold: First, to develop a hypercongestive reservoir model of a transit system. While other studies have incorporated buses into bathtub models of city traffic (Chiabaut, 2015; Ampountolas et al., 2017; Paipuri and Leclercq, 2020), here the transit system itself is a reservoir where the stock of passengers interferes with their delivery. The second goal is to show how concepts refined in the study of traffic may be modified for other systems. The third goal is to derive a new way that transit systems might exhibit multiple equilibria.

There are many equivalences between our model and bathtub traffic models, and we highlight them throughout. The reader familiar with the traffic literature may benefit from knowing two essential differences in advance:

[1] For example, Arnott (2013), Fosgerau (2015), Arnott et al. (2016), Jin (2020), Bao et al. (2021).

[2] Jara-Diaz and Gschwender (2003) sketches a general portrait of the family of isotropic transit models, and Hörcher and Tirachini (2021) reviews many such studies alongside empirical results

Table 1
Notation reference.

(a) Variables

Variable	Unit	Description
V	veh	Fleet size
A	pax/tu	Alighting rate
a	pax/veh-tu	Alighting rate per vehicle
B	pax/tu	Boarding rate
b	pax/veh-tu	Boarding rate per vehicle
Q	pax/tu	Flow of either type
R	du	Route length
u	tu/du	Unit travel time
u^0	tu/du	Free-flow unit travel time
k	pax/veh	Load
k^s	pax/veh	Friction threshold
n	pax	Accumulation
n^c	pax	Critical accumulation
ℓ	du	Average trip length
c	$/veh-tu	Vehicle operating cost
t	util	Travel time index
ϕ	$/util-tu	Welfare cost of change in t
S	$/tu	(Flow of) social welfare
d^a	veh-tu/pax	Alighting delay
d^b	veh-tu/pax	Boarding delay
p	$/pax	Fare
ω	util/tu	Wait time weight

(b) Functions. Some of these take modified versions with different arguments, indicated by hats, during the stability analysis

Function	Unit	Description
$B = D(p, t)$	pax/tu	Demand
$t = T(B; p)$	util	Inverse demand
$t = \tau(n, L)$	util	Travel cost
$A = \alpha(n, V)$	pax/tu	Network alighting function
$u = \mu(k)$	tu/du	Unit travel time
$\delta^a(k)$	veh-tu/pax	Alighting delay
$\delta^b(k)$	veh-tu/pax	Boarding delay
$\delta(k)$	veh-tu/pax	Total passenger delay

- For transit, the direct causes of delays are that people board and alight (i.e., flows of passengers). Loads cause delays indirectly insofar as they extend boarding/alighting times. Our steady-state, macroscopic relations may be written in terms of loads only because each load implies certain flows. For traffic models, on the other hand, vehicle density (the equivalent of load) is more directly causative. Vehicle density implies shorter spacings and lower speeds, and at high levels it may imply spillbacks. The flows of vehicles entering and leaving a zone (equivalent to boarding and alighting) are not usually treated as directly causative.
- Drivers are not usually treated as having preferences over traffic density *per se* (apart from the speeds they imply), whereas passengers care about crowding.

1.3. Plan of paper

Section 2 develops the model's setting and physics. Section 3 outlines passenger demands and characterizes equilibrium. Equilibrium is not unique; multiple may arise including hypercongested ones. Equilibria vary in their comparative statics. Section 4 derives properties of the social optimum. Section 5 analyzes the stability of equilibria by posing a continuous-time autonomous system for the evolution of demand (boarding rates) and accumulation. Section 6 considers "metered states", which may arise either when the system becomes too crowded for passengers to board or when the operator limits dwell times to control loads. Section 7 concludes with ideas for future work.

2. Physics

This section develops macroscopic relationships among the steady-state statistics that characterize a transit route. Discussion of dynamics is left for Section 5. For reference, Table 1 lists the paper's variables, and Table 1 the functions. We use the following units: *du*'s for distance units, *tu*'s for time units, *veh* for masses of transit vehicles and *pax* for passengers. The model is a continuum approximation model in which vehicles and passengers can take any positive value. The route is isotropic and characterized entirely by its aggregate statistics without reference to particular locations. Partial derivatives of multivariate functions are indicated by subscripts. The size of the transit vehicle, its design and its boarding/alighting discipline are assumed to be fixed for simplicity. This assumption simplifies the analysis to two decision variables (fare and fleet size) and reduces the number of arguments in our functions, but it should be relaxed in a model aimed at realism.

2.1. Passenger congestion

Let a and b (pax/veh-tu) be the rates that passengers, respectively, alight from and board a single vehicle. Let d^a and d^b (veh-tu/pax) be, respectively, the *alighting delay* and *boarding delay*. These are not necessarily the amount of time a person experiences whilst boarding/alighting. Rather, they are best thought of as the additional time required for the door to be open for each passenger who boards/alights.

We now derive expressions for how quickly the vehicle moves. Let u (tu/du) be the *unit travel time*: how long it takes a vehicle to travel one du. Let u^0 be the "free-flow" unit travel time: how fast the vehicle would move if no one boarded or alighted. u^0 presumably includes times required to stop, but not the time for people to move on/off the vehicle. During an interval of duration u (tu), au passengers alight and bu board. Hence,

$$u = u^0 + \underbrace{(ad^a + bd^b)u}_{\text{door open time}} \tag{1}$$

with the indicated term standing for the door-open time[3] consumed while the vehicle travels one distance unit.

Solving (1) for u yields

$$u = \frac{u^0}{1 - ad^a - bd^b} \quad \text{(tu/du)}. \tag{2}$$

Hence, only flows such that $1 > ad^a + bd^b$ are permissible. The opposite, $1 < ad^a + bd^b$, implies that the passengers who board/alight during one time unit take longer than one time unit to do so—a contradiction. For intuition, we may imagine the boarding rate rising infinitely: at a certain point, the vehicle can no longer move and hence cannot pick anyone up.

Let k (pax/veh) be the *load*. It is equivalent to traffic density: the number of vehicles per lane-km in a network. Or if each vehicle were thought of as its own reservoir (as in models of multi-reservoir traffic systems such as Mariotte and Leclercq, 2019), then k would map to the accumulation in each reservoir. Traffic density is a better equivalent, because multi-reservoir models capture flows between reservoirs, and there are no transfer trips here.

Next, suppose that due to friction between boarding/alighting passengers and passengers on the vehicle, the boarding and alighting delays are given by the functions $d^a = \delta^a(k), d^b = \delta^b(k)$ (veh-tu/pax).

Assumption A1. $\delta^a(k)$ and $\delta^b(k)$ are strictly positive and non-decreasing.

Because we are interested in steady-state conditions where $a = b$, until Section 5 we only use these functions' sum. Write it as the function $\delta(k) = \delta^a(k) + \delta^b(k)$ obeying

Assumption A2. There exists a load $k^s > 0$ (pax/veh) such that

- $\delta'(k) = 0$ for $k \leq k^s$;
- $\delta'(k) > 0, \delta''(k) \geq 0$ for $k > k^s$.

Here k^s is the load where crowding frictions set in. The inspiration for the threshold k^s are findings from the literature that friction only occurs once seats are taken, so that people are standing. Fletcher and El-Geneidy (2013) observes dwell times start rising sharply at about 60% of design capacity for Translink cars in Vancouver. Ostensibly, $\delta(k)$ might also be assumed to be strictly convex—as in Milkovits (2008) mentioned in the introduction. Similarly, in an experiment with a model bus, Fernandez (2011) finds alighting time per passenger rises exponentially with load. But our derivations do not depend on δ being strictly convex.

2.2. The network alighting function

We now derive a "Network Alighting Function" equivalent to the "Network Exit Function" (Daganzo et al., 2012). To do so, suppose that ℓ (du) is the *average trip length*. Make the following heroic assumption for simplicity:

Assumption A3. ℓ is invariant.

Given ℓ, a passenger's in-vehicle time is ℓu. Thus, in a steady-state with density k, the alighting and boarding rates are

$$a = b = k/\ell u \quad \text{(pax/veh)}. \tag{3}$$

Next, in (1), swap $k/\ell u$ for a and b and $\delta(k)$ for $d^a + d^b$. Solving for u yields a function $\mu(k)$ that gives the steady-state unit travel time:

$$u = \mu(k) = u^0 + \delta(k)k/\ell \quad \text{(tu/du)} \tag{4}$$

with

$$\mu' = \delta/\ell + k\delta'/\ell \qquad \mu'' = 2\delta'/\ell + k\delta''/\ell. \tag{5}$$

It follows from these equalities and δ's derivatives that

[3] Note that expressing the door-open time as a weighted sum of boarding and alighting flows makes the most sense if boarding and alighting happen through the same doors. See Lin and Wilson (1992), Milkovits (2008), Fletcher and El-Geneidy (2013) and Model 1 of Tirachini (2013) for dwell time relations with such weighted sums. By contrast, in Models 2 and 3 of Tirachini (2013), there are more complex boarding disciplines and so a more complex expression is needed. In the end, as with the isotropy assumption, our model compresses some realism here to obtain tractability. The question is also somewhat voided in the steady-state whereby the flows of boarding and alighting passengers are equal.

Proposition 1. *Under Assumption A1, A2 and A3,*

- *μ is monotonically increasing: i.e., $\mu' > 0$.*
- *When $k \le k^s$, μ is linear: i.e., $\mu'' = 0$.*
- *When $k > k^s$, μ is strictly convex: i.e., $\mu'' > 0$.*

Next, we scale up to the level of the route. Let V (veh) be the *fleet size*: the number of vehicles circulating on the route at once. If k is traffic[4] density, then V is equivalent to the lane-length of an urban zone. Let $n = kV$ (pax) be the *accumulation*: the mass of passengers on all vehicles. Let $A = Va$ and $B = Vb$ (pax/tu) be the respective rates of alighting and boarding for the entire route… the in-flow to and out-flow from the system. In steady-state where $a = b = k/\ell u$, we have

$$A = B = n/\ell u \quad \text{(pax/tu)}. \tag{6}$$

Swapping the steady-state relation $\mu(k)$ for u yields a function

$$A = \alpha(n, V) := \frac{n}{\ell\mu(n/V)} = \frac{1}{\ell u^0/n + \delta(n/V)/V.} \quad \text{(pax/tu)} \tag{7}$$

Call α the "Network Alighting Function", akin to the "Network Exit Function". It yields the rate passengers complete their trips, given a certain number of passengers in the system n and a certain fleet size V.

Proposition 2. *Under Assumption A1, A2 and A3,*

- *$\alpha(n;V)$ is strongly unimodal over n and rises from zero. Call n^c the location of its maximum, given V.*
- *$\alpha(n;V)$ is strictly concave for $n \le n^c$.*

The proof is in the appendix. We say α is "strongly" unimodal in the sense that it achieves its maximum at a unique accumulation n^c (pax), which is sometimes called the *critical accumulation* in bathtub traffic models. α cannot have a flat top like a trapezoid. The flow $\alpha(n^c;V)$ (pax/tu) is the route's flow capacity (if n^c is actually attainable). Later, in Section 5 we develop non-steady-state functions for u and A. Note also that Fig. 2(a) shows $\alpha(n;V)$ as a parabola. This is only an example, and nothing in our derivations requires it to be a parabola or even to go to zero at a large-enough value (though it reasonably ought to do so). As in road traffic models, on the left side (right side) of n^c we have an *uncongested* (*hypercongested*) regime.

2.3. Wait times

Let R (du) be the *route length*, so that the cycle time of a vehicle is Ru. When a vehicle finishes the route it starts over, so headway in a steady-state with load k and fleet size V is $H = R\mu(k)/V$ (tu/veh). Hence the headway rises with k and falls with V. Passengers arrive randomly at stops and, for now, board during the first headway after they arrive at the stop. Thus, the average passenger waits one half headway for one vehicle and experiences a waiting time of $Ru/2V$ (tu). In Section 6 we take up the possibility of limits to the rate of boarding, such that passengers must wait longer than one headway.

In summary, each flow can be observed at either of two states:

- *hypercongested:* slow vehicles, crowded conditions, long travel times, large boarding/alighting delays, long headways.
- *uncongested*: fast vehicles, short travel times, uncrowded conditions, small boarding/alighting delays, short headways.

3. Equilibrium

3.1. Demands

In equilibrium, the rate of boardings is given by a *demand function* $B = D(p, t)$ (pax/tu), where p (\$/pax) is the *fare* and t (util) is the *travel time index*: a weighted measure of passengers' travel time. Assume D declines monotonically with both arguments as long as it is positive. The formula for t is

$$t = \omega Ru/2V + \ell u\theta(k) \quad \text{(util)}, \tag{8}$$

where ω (util/tu) is the disutility of wait time and $\theta(k)$ (util/tu) is function of load which gives the disutility of in-vehicle time. Note that while we have stated t in terms of utils, its units do not play a role in the paper. Regarding $\theta(k)$, assume:

Assumption A4. $\theta(k)$ is strictly positive and non-decreasing.

[4] By this interpretation, the operator's choice of fleet size is similar to a road agency's choice of how many lanes to build or, more likely, how much street space to allocate to cars—as treated in a bathtub model by Gonzales et al. (2010), Zheng and Geroliminis (2013).

θ is non-decreasing (and probably increasing), due to *crowding*. The relationship between loads and crowding disutility has been a subject of substantial research. See Tirachini et al. (2017) for a review of crowding research and new results from Santiago de Chile. What is crowding? In a survey of crowding preferences among Parisian metro passengers, Haywood et al. (2017) lists among the costs of crowding the physical discomfort of standing (once all seats are taken), the mental stress of having people intrude into one's personal space, smell, noise and *wasted time* (e.g., the inability to read). There are various ways to define crowding, and Li and Hensher (2013) reviews various metrics in use. de Palma et al. (2015, Sec. 3) considers different functional forms for how the disutility of crowding rises with the density of passengers. Since we have fixed the size and design of the transit vehicle, load should map one-to-one to any of these measures.

Ostensibly, $\theta(k)$ is flat over a range from 0 to some level lower than n^c because n^c is around the level at which all seats are taken. However, for our purposes, details of θ's shape do not matter so long as it does not decrease. While capacity constraints (in a physical sense) do not enter this model, they could be incorporated by having $\theta(k)$ go asymptotically to infinity at a certain point—so that no equilibrium with such high load could arise and no social planner would try to choose such state for the social optimum.

In a steady-state, we have the function $t = \tau(n, V)$, where

$$\tau(n, V) := \mu(n/V) \cdot \left[\omega R/2V + \ell\theta(n/V)\right] \quad \text{(util)}, \tag{9}$$

Thus, τ rises with n and falls with V. Over n, it will be strictly convex above Vk^s because $\mu(k)$ is strictly convex above k^s.

Note this specification of travel cost is more general than models which presume a single type of traveler preference. Here, passengers may or may not trade off p and t at different rates: no universal value of travel time savings is assumed. The specification does not even assume linear values of time. However, this specification is less general than it could be. The index t supposes everyone trades off wait and in-vehicle time in the same ratio. Also, that ratio changes with crowding in the same way for everyone. It would be more general to use a function with fare, wait time, in-vehicle time and load as separate arguments. Zhang et al. (2020) supposes a generalized travel cost index which lets the coefficients on travel time and crowding vary among travelers. We have not done so for the sake of simplicity.

3.2. Equilibrium condition

Given p and V, equilibrium arises at an accumulation n° such that

$$\alpha(n^\circ, V) = D[p, \tau(n^\circ, V)] \quad \text{(pax/tu)}. \tag{10}$$

The RHS is monotonically falling with n°, while the LHS rises and then falls. Hence, there may be multiple equilibria. At most one can be uncongested, but there may be multiple hypercongested ones. This is contrast to Zhang et al. (2020) in which equilibrium is unique.

Let Q (pax/tu) be a variable that indicates passenger flow of either type (boarding or alighting) in a steady-state (when the two are equal). Fig. 2(a) plots two versions of $Q = D\left[p, \tau\left(n; V\right)\right]$ alongside $Q = \alpha(n; V)$ in (n, Q) space (note we have used the semicolon to indicate V is fixed). Intersections of $D(\cdot)$ and $\alpha(\cdot)$ represent equilibria.

To frame things in a way more familiar to the study of traffic (Walters, 1961), define a declining inverse demand function $t = T(B; p)$ by the implicit relation:

$$B = D[p, T(B; p)] \quad \text{(pax/tu)}. \tag{11}$$

$T(B; p)$ is the value of t that entices B passengers per tu, given a fare p. T is well-defined, because D falls monotonically with t. In (Q, t) space, Fig. 2(b) plots two versions of $t = T(Q; p)$ and the parametric curve $\langle\alpha(n, V), \tau(n, V)\rangle$.

3.3. Comparative statics

Different equilibria vary in how they change in response to a rise in the fare. To see how, implicitly differentiate the equilibrium relationship

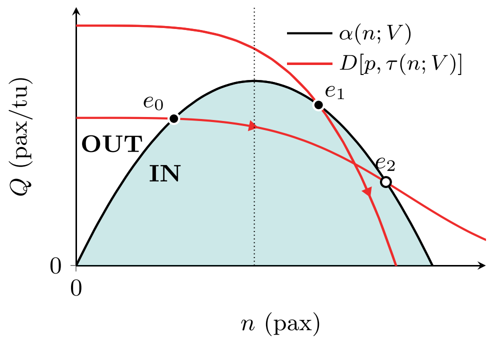
$$D\left[p, \tau\left(n^\circ, V\right)\right] = \alpha(n^\circ, V) \quad \text{(pax/tu)} \tag{12}$$

with respect to p:

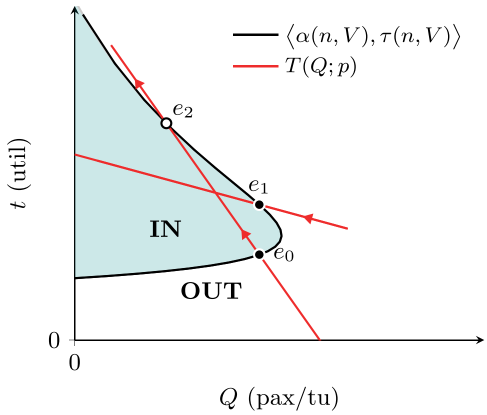
$$D_p + D_t\tau_n\frac{\mathrm{d}n^\circ}{\mathrm{d}p} = \alpha_n\frac{\mathrm{d}n^\circ}{\mathrm{d}p} \qquad \Rightarrow \frac{\mathrm{d}n^\circ}{\mathrm{d}p} = \frac{D_p}{\alpha_n - D_t\tau_n}. \tag{13}$$

Everywhere $D_p < 0$, but the sign of the denominator term $\alpha_n - D_t\tau_n$ varies among equilibria. To distinguish this visually, let **IN** and **OUT** indicate the regions of (n, Q) space, respectively, below and above $\alpha(n; V)$. Equivalently, **IN** and **OUT** are the regions of (Q, t) space respectively left of and right of the "supply curve" $\langle\alpha(n, V), \tau(n, V)\rangle$. These regions are shaded in Figs. 2(a) and 2(b). Ignoring degenerate cases where the two curves are exactly tangent, every equilibrium is a point where the demand curve $D(\cdot)$ crosses between the two regions (in either space). Call *inside-out*[5] those equilibria that cross from **IN** to **OUT** as n rises in (n, Q)

[5] Pandey et al. (2024) uses a similar system but calls outside-in "from above" and inside-out "from below". This system is better because the descriptions apply on both spaces.



(a) Equilibria in (n, Q) space



(b) Equilibria in (Q, t) space

Fig. 2. Equilibrium in two spaces.

space or (equivalently) as t rises in (Q, t) space. e_2 is inside-out in Fig. 2. An inside-out equilibrium has $\alpha_n - D_t\tau_n < 0$. Hence, a rise in the fare *increases* the accumulation and unit travel time. On the other hand, an *outside-in* equilibrium crosses **OUT** → **IN** and has $\alpha_n - D_t\tau_n > 0$, so a rise in p reduces accumulation and downstream variables. *Un*congested equilibria (with $\alpha_n > 0$) are necessarily *outside-in*, but *hyper*congested (with $\alpha_n < 0$) may be either.

4. Social optimum

Since Mohring (1972), many studies have tackled the problem of social welfare maximization for isotropic transit systems under different assumptions (Hörcher and Tirachini, 2021). This section derives optimal values of the fare and fleet size.

Let c (\$/veh-tu) be the cost of operating one vehicle for one time unit (including perhaps its amortized purchase price). In an equilibrium with fare p and fleet size V, social welfare (measured in money flows) is

$$S = \int_p^\infty D\left[P, \tau(n^\circ, V)\right] dP + p\alpha(n^\circ, V) - cV \quad (\$/\text{tu}) \tag{14}$$

$$\text{s.t. } \alpha(n^\circ, V) = D[p, \tau(n^\circ, V)]. \tag{15}$$

In (14), the integral term is the value of consumer surplus. The sum of the other two terms is the operator's profit (or loss). To proceed, let

$$\phi = -\int_p^\infty D_t(P, t) dP \quad (\$/\text{util-tu}) \tag{16}$$

stand for the welfare cost of a change in t. Now differentiate S with respect to fare:

$$\frac{\partial S}{\partial p} = \frac{\mathrm{d}n^\circ}{\mathrm{d}p}\left(p\alpha_n - \phi\tau_n\right) \tag{17}$$

Note that both delay and crowding costs are included in τ_n. Similarly, Turvey and Mohring (1975) and Kraus (1991) calculate optimal fares that reflect passenger congestion costs. From (17) alone (regardless of V), we may deduce:

Proposition 3. *No hypercongested equilibrium is socially optimal.*

To prove this, first note that if there is an "inside-out" hypercongested equilibrium then there must be an "outside-in" one with a lower accumulation. This is just to say that in order for the demand curve $D[p, \tau(n; V)]$ to leave **IN**, then it must have had to enter somewhere at a lower accumulation. Hence, an inside-out equilibrium can never be socially optimal: such an equilibrium is Pareto-dominated by the implied lower-density equilibrium, which has a higher flow, lower travel time index and the same fare and cost. Now consider a hypercongested outside-in equilibrium like e_1 in Fig. 2(a). At these, the RHS of (17) is strictly positive (because $\mathrm{d}n^\circ/\mathrm{d}p < 0, \alpha_n < 0, \tau_n > 0$). Hence, the social planner can do better by raising the fare. Eventually, as the fare rises (and the accumulation falls), the equilibrium will become uncongested. Therefore, hereafter we may assume that in considering the social optimum we speak of an uncongested equilibrium with $\alpha_n > 0$.

Setting $\partial W/\partial p = 0$ (the FOC on p) and solving for p yields

$$p = \phi\tau_n/\alpha_n \quad (\$/\text{pax}). \tag{18}$$

Turning now to the choice of V, we have:

$$\frac{\partial S}{\partial V} = \frac{\mathrm{d}n^\circ}{\mathrm{d}V}\left(p\alpha_n - \phi\tau_n\right) + p\alpha_V - \phi\tau_V - c \tag{19}$$

Setting $\partial S/\partial V = 0$ and substituting $\phi\tau_n/\alpha_n$ for p in (19) yields

$$c = -\phi\tau_V + p\alpha_V \quad (\$/\text{veh-tu}) \tag{20}$$

τ's partial derivatives are

$$\tau_n = \frac{\mu'}{V}\cdot\left[\omega\frac{R}{2V} + \ell\theta(n/V)\right] + \frac{\theta'}{V}\ell u \qquad \tau_V = -\frac{n}{V}\tau_n - \frac{Ru}{2V^2}, \tag{21}$$

where $u = \mu(n/V)$. Expanding τ_V in (20), evaluated at the equilibrium n° yields

$$c = \phi\frac{n^\circ}{V}\tau_n + \phi\omega\frac{Ru}{2V^2} + \phi\frac{\tau_n}{\alpha_n}\alpha_V = \phi\tau_n\underbrace{\left(\frac{n^\circ}{V} + \frac{\alpha_V}{\alpha_n}\right)}_{A/V\alpha_n} + \phi\omega\frac{Ru}{2V^2} \tag{22}$$

By taking α's partial derivatives, it can be shown that the term in parentheses simplifies to $A/V\alpha_n$, where $A = \alpha(n^\circ, V)$. Hence,

$$c = \phi\frac{\tau_n}{\alpha_n}\frac{A}{V} + \phi\omega\frac{Ru}{2V^2}. \tag{23}$$

In the social optimum, the first term on the RHS is just pA/V: how much money one vehicle earns per time unit. So the cost of a vehicle exceeds it earnings. It follows that the route loses money at the social optimum, and the size of the loss over the whole route is

$$\pi = pA - cV = \phi\frac{\tau_n}{\alpha_n}A - \phi\frac{\tau_n}{\alpha_n}A - \phi\omega\frac{Ru}{2V} = -\phi\omega\frac{Ru}{2V} \tag{24}$$

Recall $Ru/2V$ is the expected wait time. $\phi\omega$ is the cost to passengers (converted from utils to dollars) of a marginal increase in the wait time. Thus, the optimal total subsidy can be thought of as (approximately) the cumulative "value of time" that passengers spend waiting. The same is true in models of dispatch (Arnott, 1996) and street-hail (Lehe and Pandey, 2022b) taxi service.

5. Local asymptotic stability of equilibria

5.1. Context

In this section we analyze the equilibria's *stability*—what Watling (1998, p. 156) calls their "attractiveness...at points of disequilibrium". Stability is a concept from the study of dynamical systems, which refers to several types of stability. (See Strogatz, 2019 for an introduction to dynamical systems, stability and chaos.) These types have long been treated in the traffic assignment literature (e.g., Horowitz, 1984), and more recently been applied to bathtub traffic models. For example, Arnott and Inci (2010) and Haddad and Geroliminis (2012) examine the *global* stability of equilibria, while Laval et al. (2018, Sec. 2.2) and Pandey et al. (2024) analyze the *local asymptotic* stability. Like the latter references, this section considers local asymptotic stability in the following sense: Suppose the route's aggregate statistics slightly differ from the precise values they take at some equilibrium, e. Will the system's state (the collection of statistics such as accumulation, boarding rates, etc.) approach e over time? If so, then we say that the equilibrium is (locally asymptotically) stable.

Note that there is no contradiction between (i) the static setting; and (ii) considering the evolution of the system state over time. What makes the model static is that the demand function, $D(p,t)$, and all the physical parameters and functions (e.g., R, V, ℓ, δ etc.) are fixed and constant. The derived statistics n, A, B, u are *not* necessarily constant except at steady-state equilibria, where (by definition of a steady-state) they do not change. When they assume out-of-equilibrium values they will reasonably change (in ways described below). By contrast, in a *dynamic* version of our model, travelers might have preferences that change over time (e.g., around a morning work start time) or the physical system might change (e.g., if the fleet size is reduced or expanded at a certain time of day). A useful way to think about the equilibrium is to imagine it as transpiring during a morning rush period over which demands and service provision are approximately constant. See de Palma and Fosgerau (2010) for more discussion of the difference between static and dynamic traffic models.

5.2. Setup

Equilibria are not stable per se but rather in the context of a dynamical system which captures how the system's state evolves. Because the system's physics depend on accumulation, which may change from minute-to-minute, we apply *continuous-time* dynamics rather than the discrete dynamics often used for traffic assignment (Watling, 1998, 1999). A basic rule underlying these dynamics must be a conservation law of passengers:

$$\dot{n} = B - A. \tag{25}$$

Next we need rules for the motion of the passenger flows B and A. Regarding A, we assume the following:

Assumption A5 (*Bathtub Assumption*)**.** Given a unit travel time u and accumulation n, the flow of alighting passengers is instantaneously given by its steady-state value $A = n/\ell u$.

Arnott (2013, p. 112–113) contains a discussion of the strengths and weaknesses of this assumption for traffic. Jin (2020) discusses some of its implications, such as that trip lengths are distributed negative exponential, and provides a more exact "generalized bathtub model." Unfortunately, relaxing this assumption requires tracing the life of each trip through the network/route. In the traffic literature, this is often accomplished numerically via "trip-following" simulations (Daganzo and Lehe, 2015; Lamotte and Geroliminis, 2017; Mariotte and Leclercq, 2019), although Fosgerau (2015) presents a dynamic model which obtains an analytical solution. In some cases, the inexactness of the bathtub assumption is not as important: e.g., experiments and calculations in Laval (2023) suggest non-exponential trip length distributions have only a short-lived effect on system evolution. Trip-following simulations conducted for a bathtub traffic model with a lognormal distribution of trip lengths in Pandey et al. (2024) supported the same stability results derived under the bathtub assumption.

Since we are concerned with the stability of an equilibrium given V and p, these are not evolving state variables but rather are fixed. Hence we can write the u in terms of n and B. The rate of boardings and alightings per vehicle are, respectively, A/V and B/V (pax/veh-tu). Hence,

$$u = u^0 + d^a Au/V + d^b Bu/V \quad \text{(tu/du)} \tag{26}$$

Swap A with $n/\ell u$, per A5:

$$u = \frac{u^0 + d^a n/\ell V}{1 - d^b B/V} \tag{27}$$

Now swap $d^a = \delta^A(n/V)$ and $d^b = \delta^B(n/V)$ to write a new function

$$u = \hat{\mu}(n, B) := \frac{u^0 + \delta^a(n/V)n/\ell V}{1 - \delta^b(n/V)B/V} \quad \text{(tu/du).} \tag{28}$$

This function yields the instantaneous unit travel time, given a certain accumulation and boarding rate, under assumption A5. Similarly, the alighting rate may be written a function of accumulation and boarding rate:

$$A = \hat{\alpha}(n, B) := \frac{n}{\ell \hat{\mu}(n, B)} \quad \text{(pax/tu).} \tag{29}$$

So the conservation law becomes:

$$\dot{n} = B - \hat{\alpha}(n, B). \tag{30}$$

As for boardings, we assume they adjust gradually according to a form of "replicator dynamics", as used in Iryo (2019). (See Hofbauer and Sandholm, 2009 for an overview of various dynamics used in evolutionary game theory.)

Assumption A6 (*Replicator Dynamic*)**.** Given t, the boarding rate evolves by

$$\dot{B} = \zeta B \cdot [T(B; p) - t] \quad \text{where } \zeta > 0. \tag{31}$$

Here, ζ is some positive parameter which indicates the speed of adjustment of demand. When $\zeta = \infty$, the solution to (31) is $B = D(p, t)$. To close the loop, write t as a function similar to $\hat{\mu}$ and $\hat{\alpha}$:

$$t = \hat{\tau}(n, B) := \hat{\mu}(n, B) \left[\omega R/2V + \ell\theta(n/V)\right] \quad \text{(util)} \tag{32}$$

This completes the foundation for an autonomous dynamical system:

$$\dot{n} = B - \hat{\alpha}(n, B) \qquad \text{(conservation law)} \tag{33}$$

$$\dot{B} = \zeta B \left[T(B; p) - \hat{\tau}(n, B)\right] \qquad \text{(replicator dynamics)} \tag{34}$$

The first equation is the conservation law with A replaced by $\hat{\alpha}(n, B)$. The second equation is the replicator dynamic with t replaced by $\hat{\tau}(n, B)$.

5.3. Stability results

First affirm that the fixed points of the system are also equilibria of the static model. $\dot{n} = 0$ implies $B = \hat{\alpha}(n, B)$. With the system in a steady-state where boardings equal arrivals, the function $\alpha(n, V)$ holds; hence $B = \hat{\alpha}(n, B) = \alpha(n, V)$. In turn, $\dot{B} = 0$ implies $B = D[p, \hat{\tau}(n, B)]$. When the first is satisfied, the system is in a steady-state where $\hat{\tau}(n, B) = \tau(n, V)$. So we have it that a fixed point occurs at a value of n such that $\alpha(n, V) = D[p, \tau(n, V)]$, which is the equilibrium condition of the static model given in Section 3.

To check the stability of a fixed point, we follow the standard procedure for a 2D dynamical system. (See Strogatz, 2019, Ch. 5 for a clear explanation.) The first step is to inspect the system's Jacobian at some fixed point. Since $T_B = 1/D_t$, the Jacobian is

$$J = \begin{pmatrix} -\hat{\alpha}_n & 1 - \hat{\alpha}_B \\ -\zeta B \hat{\tau}_n & \zeta B \left(1/D_t - \hat{\tau}_B\right) \end{pmatrix} \tag{35}$$

For the fixed point to be stable, J must be negative definite, which implies its determinant is positive and its trace negative. These are

$$\mathrm{tr}(J) = -\hat{\alpha}_n + \frac{\zeta B}{D_t}(1 - D_t\hat{\tau}_B) \qquad \det(J) = -\frac{\zeta B}{D_t}\{\hat{\alpha}_n - D_t\left(\hat{\tau}_B\hat{\alpha}_n + \hat{\tau}_n - \hat{\tau}_n\hat{\alpha}_B\right)\} \tag{36}$$

Since $D_t < 0$, a positive trace implies

$$\zeta > \frac{\hat{\alpha}_n D_t}{B(1 - D_t\hat{\tau}_n)}. \tag{37}$$

Note $\alpha_n = \hat{\alpha}_n \cdot (1 - \delta^b(n/V)B/V)$. Thus, uncongested equilibria have $\alpha_n > 0$, and so the RHS of (37) is negative. By assumption, $\zeta > 0$ so all uncongested equilibria have a negative trace. By converse logic, at hypercongested equilibria the RHS is positive, and so ζ must be larger than a threshold which varies depending on properties of the equilibrium.

Since $D_t < 0$, the term $\zeta B/D_t$ is negative. Unpacking the derivatives reveals that, at a fixed point, the term in curly brackets on the RHS of (36) can also be written

$$\hat{\alpha}_n - D_t\left(\hat{\tau}_B\hat{\alpha}_n + \hat{\tau}_n - \hat{\tau}_n\hat{\alpha}_B\right) = \frac{\alpha_n - D_t\tau_n}{1 - \delta^b(n/V)B/V}. \tag{38}$$

Note the numerator on the RHS is what distinguishes "inside-out" from "outside-in" equilibria. Since $1-\delta^b(n/V)B/V > 0$, inside-out equilibria have negative determinants and outside-in ones have positive determinants. Recall that uncongested equilibria must be outside-in, but hypercongested can be outside-in or inside-out. Summing up:

Proposition 4. *Under Assumptions A5 and A6,*

- *uncongested equilibria are always stable;*
- *hypercongested equilibria are stable if and only if (i) they are "outside-in" and (ii) the speed-of-adjustment parameter ζ is exceeds a threshold (which depends on equilibrium properties).*

It follows that the social optimum (which is an uncongested equilibrium) is stable under these dynamics.

5.4. Relation to engineering results

Some engineering studies in the bathtub traffic literature also consider the stability of traffic states. These analyses tend to find hypercongested states are unstable (Haddad and Geroliminis, 2012; Laval et al., 2018). According to Daganzo (2007), the reason is that, at a congested state, a small and temporary rise in the in-flow of vehicles to a zone "sets in motion a positive feedback cycle where accumulation increases at a rate that itself grows with accumulation".

By contrast, we have concluded that some hypercongested equilibria can be stable. Why the difference? The reason does not lie in the different setting; Arnott and Inci (2010) and Pandey et al. (2024) reach the same results for a bathtub traffic model. Rather, the root is that, as Laval et al. (2018, p. 677, footnote 1) points out, the engineering literature usually assumes "demand" (inflow) to be exogenous. Such studies are thus depicting a special case in which $D_t = 0$: a horizontal demand curve. This would only yield *inside-out* hypercongested equilibria, which we have shown to be unstable. Hence our results agree with the instability results in the traffic literature.

6. Metered boarding

So far we have assumed that anyone who wants to board does so during the headway in which they arrive at the stop. In fact, two mechanisms may prevent that from happening and effectively "meter" the boarding flow. These may also prevent hypercongestion.

The first type of limit involves the operator scheduling vehicles to depart even if there are people at the stop who could board. This discipline may be formalized as a limit on the amount of time the door can be open per distance unit or, equivalently, by targeting a certain "minimum speed", $s_{\min}$ (du/tu). In either case, some passengers will be left behind when the vehicle departs, and by enforcing this discipline the operator is doing something analogous to "perimeter control" schemes proposed to limit the in-flow of vehicles to a zone. If the schedule is fixed to achieve a minimum speed $s_{\min}$, then we have $\mu(k) \le 1/s_{\min}$. Since $\mu' > 0$, a minimum speed implies a maximum load $\mu(\tilde{k}) = 1/s_{\min}$ and hence (given V) a maximum accumulation $\tilde{n} = V\tilde{k}$.

The second is a naturally-occurring capacity limit. For example, suppose that at loads higher than $\tilde{k}$ boarding is either impossible or else the passengers on the vehicle will not allow anyone else to board. In this case, at an accumulation of $\tilde{n} = V\tilde{k}$, one passenger will be able to board for each one that steps off. At $n \ge \tilde{n}$, passengers will only be able to board when *more than one* passenger alights, so such accumulations cannot be steady-states. Similarly, engineering studies (e.g., Geroliminis and Daganzo, 2007) have suggested there exists an "entry supply function" $I(n)$ which yields the maximum rate vehicles can enter a zone for each accumulation. It falls at high accumulations, when congestion around the border of the zones limits how quickly vehicles can enter. The accumulation where $I(n)$ intersects the Network Exit Function is what Daganzo (2007) calls the "equilibrium gridlock point" (though it is not exactly the same as "gridlock" in the colloquial sense). It represents the highest possible accumulation that can be achieved by fluctuations to traffic in-flow and is analogous to $\tilde{n}$.

Both mechanisms (accidental and purposeful limits) have the same result: given a fleet size, there exists a maximum accumulation which we have called $\tilde{n}$. Presumably an operator would choose a policy that results in an uncongested $\tilde{n}$. Moreover, the accidental

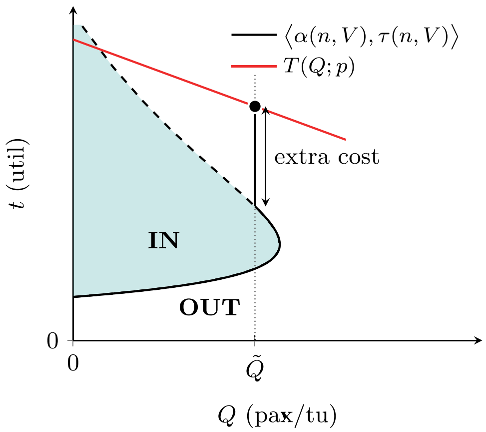


Fig. 3. Extra cost in a metered state with flow $\tilde{Q}$.

value of $\tilde{n}$ may be smaller than n^c. In traffic models, nothing especially requires that such $\tilde{n}$ occur in hypercongestion (see for example Model I of Mariotte and Leclercq, 2019). In either case, if $n^c > \tilde{n}$ then hypercongestion cannot arise. When n is limited to $\tilde{n}$, we say that the system is operating in a *metered state*. Everything about the metered state is determined: a passenger flow, $\tilde{Q} = \alpha(\tilde{n}, V)$, unit travel time $\tilde{u} = \mu(\tilde{n}/V)$, headway, $\tilde{H} = R\mu(\tilde{n}/V)$, and so on.

A question arises: what happens if

$$D[p, \tau(\tilde{n}, V)] > \tilde{Q}?$$

That is, what if the metered state invites excess demand? The analogous question arises in economic models of vehicles traversing a highway bottleneck of fixed capacity. The consensus among economists is that a queue forms upstream of the bottleneck, and that the queue grows until the disutility of traversing the bottleneck is sufficient that demand exactly equals the bottleneck's capacity (Verhoef, 2005; Small and Verhoef, 2007). Once this state is achieved, it is an equilibrium: vehicles arrive at the back of the queue at the exact rate they depart from the front of the queue through the bottleneck.

Ostensibly something similar applies to transit boarding: in a metered state, a queue forms at the transit stops which induces sufficient delay that demand falls to $\tilde{Q}$. Fig. 3 illustrates the extra waiting disutility necessary to limit demand. The details of how this queue might be formalized are left for future work. Does one have the same utility while standing in a queue as waiting around a stop? Also, what are the physics of the queue? Studies show that too many passengers standing at platforms interferes with boarding/alighting, just as standees on the vehicles themselves do. See for example Lam et al. (1999) on crowding in light rail stations in Hong Kong. To the degree this is so, then the metered transit route is *dis*similar to a highway bottleneck. Cars traverse highway bottlenecks at the road's maximum capacity regardless of the size of a queue. (See Anderson and Davis, 2020 for a review and new evidence.) If the queues extend boarding/alighting delays, then it would be accurate to model the route as a *two*-reservoir system where the vehicles are a metered reservoir and the stops collectively represent a second, unmetered reservoir (very similar to Haddad and Geroliminis, 2012). In turn, if access to the platforms are themselves metered (e.g., accidentally by the escalators in large subway stations or else purposefully by some gate control), then we could add a *third* reservoir, and so on.

7. Conclusion

7.1. Summary

This paper has developed an isotropic model of a transit route with elastic demand and "friction" between on-board passengers and boarding/alighting passengers. Due to this friction, we derive passenger flow as a strongly unimodal "Network Alighting Function" of passenger accumulation. These physics, in turn, determine the model's equilibria when demand declines with the fare and with a weighted index of wait and in-vehicle time. The weight on in-vehicle time depends on the number of passenger on each vehicle, to represent the disutility of crowding. We use the equilibrium derivations to find the model's socially optimal choice of the fare and fleet size, showing that the optimal choices yield an uncongested equilibrium with a deficit. While not optimal, certain hypercongested equilibria may be locally asymptotically stable under plausible dynamics, and these can be identified graphically. Finally, we explore what happens when flows are metered, whereupon the system achieves equilibrium by the growth of queues at stops—although we do not investigate the physics or economics of these queues.

The paper's main innovation has been to derive system-level relations in terms of accumulation. This leads to many correspondences between the model's physics and those used in bathtub traffic models. It is hoped the paper demonstrates that concepts refined in traffic theory could be useful in other domains, and also that it raises new possibilities for tractable models of transit. Hence, the paper is supposed to lay the foundations for future work.

7.2. Future work

In Section 6, we explained that the physics of queues at stations are a path for future work. Other progress is possible by relaxing some of our simplifications. Relaxing certain ones would probably increase complexity and realism without making fundamental progress. For example, for simplicity we have made the transit operator cost depend only on the fleet size, while in reality it depends on distance-traveled, too (Oldfield and Bly, 1988; Jansson, 1993). But incorporating distance-traveled into the cost expression would probably not yield qualitative insights. Below are our thoughts on what seem like the most promising avenues.

- There is no evidence our macroscopic relations $\mu(k)$ and $\alpha(n, V)$ exist nor that they have the shapes we derive—though the "crush loads" on many systems and the employment of professional pushers or "oshiya" (who push people onto crowded metro cars) are suggestive of hypercongestion. The shapes of μ and α are implied by dwell time research, but sometimes macroscopic relations do not behave in ways implied by microscopic relations. (For example, the quantum account of gravity is incompatible with general relativity.) Thus, actually taking and organizing data to test for the hypothesized relationships is essential. Interest in bathtub traffic models has grown owing to confirmation from measurements (Geroliminis and Daganzo, 2008) or in realistic simulations (Olmos et al., 2018; Ambühl et al., 2023). For traffic, some theory suggests that hypercongested network states may be chaotic (Laval, 2023). For bus systems, "bunching" adds an additional source of chaos unless controlled (Argote-Cabanero et al., 2015).
- While fixed average trip length is a common assumption in the bathtub traffic literature and in transit operation studies, reasonably it may be endogenous to the operator's decision variables. If the fare is per ride, then we ought to expect that a higher fare repels short trips and invites some longer trips (since the bus moves faster at a higher fare). Likewise, Lehe (2017) develops a bathtub traffic model in which a cordon toll does the same. Lehe and Pandey (2022a) gives a model in which average trip length changes with stop spacing, because longer trips take less time and shorter trips more time when spacings widen.
- We have implicitly ignored one impact of ridership on travel times: how often the bus stops. Bus routes often operate in what Kikuchi and Vuchic (1982) call an "on-call" stopping regime, whereby the bus bypasses stops if no one wants to board or alight there. Mohring (1972) develops an expression for the unit travel time which incorporates this effect, but it involves an exponential term which precludes solving for u as a explicit function of a and b. Thus, our model must represent an "all stop" regime or else an approximation that neglects this effect. Neglecting this effect is appropriate for routes with high-enough demand, which have boarding/alighting at every stop anyway, but not for systems with light demands.
- There is no road congestion in our model as would be relevant for a model of mixed-traffic buses or streetcars. The vehicles do not cause traffic congestion, and there is no congestion or queuing (by the buses) at stops themselves. See Tirachini (2014), Tirachini et al. (2014) for numerical analyses of these factors and their importance to design of high-frequency routes and stops. Relatedly, while we do not say what passengers do if they do not ride transit, reasonably many of them would drive and thus could possibly cause congestion by *not* taking the bus. In this case, there may be *additional* source of multiple equilibria due to a mechanism that Pandey and Lehe (2023) call "Congestive Mode-Switching". Finally, a one-mode bathtub model of auto traffic can exhibit multiple stable equilibria on its own due to hypercongestion (Arnott and Inci, 2010; Pandey et al., 2024). So, in summary, incorporating traffic conditions would seem to multiply the possibilities for what conditions might be stably observed.
- The operator's only decision variables are fare and fleet size. Obviously bus size is a decision variable, as considered in Jansson (1980), Oldfield and Bly (1988). The functions δ^a and δ^b could also depend on the fare collection system and boarding discipline (i.e., whether all-door or front-door) (Tirachini and Hensher, 2011; Jara-Díaz and Tirachini, 2013).

CRediT authorship contribution statement

Lewis J. Lehe: Conceptualization, Formal analysis, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing. **Ayush Pandey:** Conceptualization, Formal analysis, Writing – review & editing.

Acknowledgments

The authors would like to thank Vikash Gayah for comments as well as the National Science Foundation, United States for grant CMMI-2052512 which provided support.

Appendix

Proof of Proposition 2

$$\alpha_n = A^2 \left(\ell u^0/n^2 - \delta'/V^2\right). \tag{A.1}$$

α_n is positive over $(0, Vk^s]$ (because $\delta' = 0$ there). But as n rises higher, the positive term $\ell u^0/n^2$ falls monotonically *to* zero and the negative term $-\delta'/V^2$ is non-increasing. Hence, there must be some unique $n^c > Lk^s$ such that

$$\ell u^0/(n^c)^2 = \delta'(n^c/V)/V^2 \Leftrightarrow \alpha_n = 0 \tag{A.2}$$

Next, note

$$\alpha_{nn}/A^2 = 2A\left(\ell u^0/n^2 - \delta'/V^2\right)^2 - \left(2\ell u^0/n^3 + \delta''/V^3\right) \tag{A.3}$$

Suppose $0 < n < n^c$, which implies

$$0 < \delta'/V^2 \leq \ell u^0/n^2 \tag{A.4}$$

$$\left(\ell u^0/n^2 - \delta'/V^2\right)^2 \leq \left(\ell u^0/n^2\right)^2 \tag{A.5}$$

Hence from (A.3) we have

$$\alpha_{nn}/A^2 \leq 2A\left(\ell u^0/n^2 - \delta'/V^2\right)^2 - \left(2\ell u^0/n^3 + \delta''/V^3\right) \tag{A.6}$$

Since $\delta'' \geq 0 \ldots$

$$\alpha_{nn}/A^2 \leq 2A\left(\ell u^0/n^2\right)^2 - \left(2\ell u^0/n^3\right) \tag{A.7}$$

Expanding $A = n/\ell u$ on the RHS yields:

$$\frac{\alpha_{nn} n^3}{2\ell u^0 A^2} \leq \frac{u^0}{\mu(n/V)} - 1 \tag{A.8}$$

Since $u = \mu(n/V) > u^0$ for $n > 0$, the RHS and α_{nn} are strictly negative over $(0, n^c)$, making α concave.

References

Ambühl, L., Menendez, M., González, M.C., 2023. Understanding congestion propagation by combining percolation theory with the macroscopic fundamental diagram. Commun. Phys. 6 (1), 1–7. http://dx.doi.org/10.1038/s42005-023-01144-w.

Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of Bi-modal multi-region urban road networks. Transp. Res. B 104, 616–637. http://dx.doi.org/10.1016/j.trb.2017.05.007.

Anderson, M.L., Davis, L.W., 2020. An empirical test of hypercongestion in highway bottlenecks. J. Public Econ. 187, 104197. http://dx.doi.org/10.1016/j.jpubeco.2020.104197.

Argote-Cabanero, J., Daganzo, C.F., Lynn, J.W., 2015. Dynamic control of complex transit systems. Transp. Res. B 81, 146–160. http://dx.doi.org/10.1016/j.trb.2015.09.003.

Arnott, R., 1996. Taxi travel should be subsidized. J. Urban Econ. 40 (3), 316–333. http://dx.doi.org/10.1006/juec.1996.0035.

Arnott, R., 2013. A bathtub model of downtown traffic congestion. J. Urban Econ. 76, 110–121. http://dx.doi.org/10.1016/j.jue.2013.01.001.

Arnott, R., Inci, E., 2010. The stability of downtown parking and traffic congestion. J. Urban Econ. 68 (3), 260–276. http://dx.doi.org/10.1016/j.jue.2010.05.001.

Arnott, R., Kokoza, A., Naji, M., 2016. Equilibrium traffic dynamics in a bathtub model: A special case. Econ. Transp. 7–8, 38–52. http://dx.doi.org/10.1016/j.ecotra.2016.11.001.

Bao, Y., Verhoef, E.T., Koster, P., 2021. Leaving the tub: The nature and dynamics of hypercongestion in a bathtub model with a restricted downstream exit. Transp. Res. Part E: Logist. Transp. Rev. 152, 102389. http://dx.doi.org/10.1016/j.tre.2021.102389.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec.: J. Transp. Res. Board 2124, 127–136. http://dx.doi.org/10.3141/2124-12.

Castillo, J.C., 2022. Who Benefits from Surge Pricing?, (no. 3245533), Rochester, NY, URL https://papers.ssrn.com/abstract=3245533.

Castillo, J.C., Knoepfle, D.T., Weyl, E.G., 2023. Matching and Pricing in Ride Hailing: Wild Goose Chases and How to Solve Them, (no. 2890666), Rochester, NY, http://dx.doi.org/10.2139/ssrn.2890666.

Chiabaut, N., 2015. Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram. Transp. Res. B 81, 410–420. http://dx.doi.org/10.1016/j.trb.2015.02.005.

Daganzo, C.F., 1998. Queue spillovers in transportation networks with a route choice. Transp. Sci. 32 (1), 3–11. http://dx.doi.org/10.1287/trsc.32.1.3.

Daganzo, C.F., 2005. Improving city mobility through gridlock control: An approach and some ideas. URL https://escholarship.org/uc/item/7w6232wq.

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transp. Res. B 41 (1), 49–62. http://dx.doi.org/10.1016/j.trb.2006.03.001.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2012. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. EURO J. Transp. Logist. 1 (1), 47–65. http://dx.doi.org/10.1007/s13676-012-0003-z.

Daganzo, C.F., Lehe, L.J., 2015. Distance-dependent congestion pricing for downtown zones. Transp. Res. B 75, 89–99. http://dx.doi.org/10.1016/j.trb.2015.02.010.

Daganzo, C.F., Lehe, L.J., 2016. Traffic flow on signalized streets. Transp. Res. B 90, 56–69. http://dx.doi.org/10.1016/j.trb.2016.03.010.

de Palma, A., Fosgerau, M., 2010. Dynamic and static congestion models: A review. Working Papers no. hal-00539166. URL https://ideas.repec.org//p/hal/wpaper/hal-00539166.html.

de Palma, A., Kilani, M., Proost, S., 2015. Discomfort in mass transit and its implication for scheduling and pricing. Transp. Res. B 71, 1–18. http://dx.doi.org/10.1016/j.trb.2014.10.001.

Fernandez, R., 2011. Experimental study of bus boarding and alighting times.
Fletcher, G., El-Geneidy, A., 2013. Effects of fare payment types and crowding on dwell time: Fine-grained analysis. Transp. Res. Rec. 2351 (1), 124–132. http://dx.doi.org/10.3141/2351-14.
Fosgerau, M., 2015. Congestion in the bathtub. Econ. Transp. 4 (4), 241–255. http://dx.doi.org/10.1016/j.ecotra.2015.08.001.
Geroliminis, N., 2015. Cruising-for-parking in congested cities with an MFD representation. Econ. Transp. 4 (3), 156–165. http://dx.doi.org/10.1016/j.ecotra.2015.04.001.
Geroliminis, N., Daganzo, C.F., 2007. Macroscopic modeling of traffic in cities. In: Transportation Research Board 86th Annual MeetingTransportation Research Board, (no. 07–0413), URL https://trid.trb.org/view/801089.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transp. Res. B 42 (9), 759–770. http://dx.doi.org/10.1016/j.trb.2008.02.002.
Geroliminis, N., Levinson, D.M., 2009. Cordon pricing consistent with the physics of overcrowding. In: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee: Papers Selected for Presentation At ISTTT18, a Peer Reviewed Series Since 1959. Springer US, Boston, MA, pp. 219–240. http://dx.doi.org/10.1007/978-1-4419-0820-9_11.
Ghobrial, A., Daganzo, C.F., Kazimi, T., 1982. Baggage claim area congestion at airports: An empirical model of mechanized claim device performance. Transp. Sci. 16 (2), 246–260. http://dx.doi.org/10.1287/trsc.16.2.246.
Gonzales, E.J., Geroliminis, N., Cassidy, M.J., Daganzo, C.F., 2010. On the allocation of city space to multiple transport modes. Transp. Plan. Technol. 33 (8), 643–656. http://dx.doi.org/10.1080/03081060.2010.527171.
Haddad, J., Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. Transp. Res. B 46 (9), 1159–1176. http://dx.doi.org/10.1016/j.trb.2012.04.004.
Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: Who cares and why? Transp. Res. Part A: Policy Pract. 100, 215–227. http://dx.doi.org/10.1016/j.tra.2017.04.022.
Hofbauer, J., Sandholm, W.H., 2009. Stable games and their dynamics. J. Econom. Theory 144 (4), 1665–1693.e4. http://dx.doi.org/10.1016/j.jet.2009.01.007.
Hörcher, D., Tirachini, A., 2021. A review of public transport economics. Econ. Transp. 25 (July 2020), http://dx.doi.org/10.1016/j.ecotra.2021.100196.
Horowitz, J.L., 1984. The stability of stochastic equilibrium in a two-link transportation network. Transp. Res. B 18 (1), 13–28. http://dx.doi.org/10.1016/0191-2615(84)90003-1.
Iryo, T., 2019. Instability of departure time choice problem: A case with replicator dynamics. Transp. Res. B 126, 353–364. http://dx.doi.org/10.1016/j.trb.2018.08.005.
Jansson, J.O., 1980. Simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy 14 (1), 53–80, arXiv:20052563. URL https://www.jstor.org/stable/20052563.
Jansson, K., 1993. Optimal public transport price and service frequency. J. Transp. Econ. Policy 27 (1), 33–50, arXiv:20034976. URL https://www.jstor.org/stable/20034976.
Jara-Diaz, S.R., Gschwender, A., 2003. Towards a general microeconomic model for the operation of public transport. Transp. Rev. 23 (4), 453–469. http://dx.doi.org/10.1080/0144164032000048922.
Jara-Díaz, S., Tirachini, A., 2013. Urban bus transport: Open all doors for boarding. J. Transp. Econ. Policy (JTEP) 47 (1), 91–106.
Jin, W.-L., 2020. Generalized bathtub model of network trip flows. Transp. Res. B 136, 138–157. http://dx.doi.org/10.1016/j.trb.2020.04.002.
Johari, M., Keyvan-Ekbatani, M., Leclercq, L., Ngoduy, D., Mahmassani, H.S., 2021. Macroscopic network-level traffic models: Bridging fifty years of development toward the next era. Transp. Res. C 131, 103334. http://dx.doi.org/10.1016/j.trc.2021.103334.
Kikuchi, S., Vuchic, V.R., 1982. Transit vehicle stopping regimes and spacings. Transp. Sci. 16 (3), 311–331. http://dx.doi.org/10.1287/trsc.16.3.311.
Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. J. Urban Econ. 29 (2), 249–259. http://dx.doi.org/10.1016/0094-1190(91)90018-3.
Lam, W.H.K., Cheung, C.-Y., Lam, C.F., 1999. A study of crowding effects at the Hong Kong light rail transit stations. Transp. Res. Part A: Policy Pract. 33 (5), 401–415. http://dx.doi.org/10.1016/S0965-8564(98)00050-0.
Lamotte, R., Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. Transp. Res. Procedia 23, 591–611. http://dx.doi.org/10.1016/j.trpro.2017.05.033.
Laval, J.A., 2023. Self-organized criticality of traffic flow: Implications for congestion management technologies. Transp. Res. C 149, 104056. http://dx.doi.org/10.1016/j.trc.2023.104056.
Laval, J.A., Leclercq, L., Chiabaut, N., 2018. Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets revisited. Transp. Res. B 117, 676–686. http://dx.doi.org/10.1016/j.trb.2017.08.027.
Lehe, L.J., 2017. Downtown tolls and the distribution of trip lengths. Econ. Transp. 11–12 (October), 23–32. http://dx.doi.org/10.1016/j.ecotra.2017.10.003.
Lehe, L., Pandey, A., 2022a. Bus Stop Spacing with Heterogeneous Trip Lengths and Elastic Demand, (no. 4135394), Rochester, NY, http://dx.doi.org/10.2139/ssrn.4135394.
Lehe, L., Pandey, A., 2022b. Taxi service with heterogeneous drivers and a competitive medallion market. J. Urban Econ. 131, 103488. http://dx.doi.org/10.1016/j.jue.2022.103488.
Li, Z., Hensher, D.A., 2013. Crowding in public transport: A review of objective and subjective measures. J. Public Transp. 16 (2), 107–134. http://dx.doi.org/10.5038/2375-0901.16.2.6.
Lin, T.-m., Wilson, N.H.M., 1992. Dwell time relationships for light rail systems. Transp. Res. Rec. (1361), 287–295, URL https://trid.trb.org/view/370918.
Mariotte, G., Leclercq, L., 2019. Flow exchanges in multi-reservoir systems with spillbacks. Transp. Res. B 122, 327–349. http://dx.doi.org/10.1016/j.trb.2019.02.014.
Mariotte, G., Leclercq, L., Batista, S.F.A., Krug, J., Paipuri, M., 2020. Calibration and validation of multi-reservoir MFD models: A case study in lyon. Transp. Res. B 136, 62–86. http://dx.doi.org/10.1016/j.trb.2020.03.006.
Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: analytical and numerical comparisons of existing models. Transp. Res. B 101, 245–267. http://dx.doi.org/10.1016/j.trb.2017.04.002.
Milkovits, M.N., 2008. Modeling the factors affecting bus stop dwell time. Transp. Res. Rec. 2072 (1), 125–130. http://dx.doi.org/10.3141/2072-13.
Mohring, H., 1972. Optimization and scale economies in urban bus transportation. Amer. Econ. Rev. 62 (4), 591–604. http://dx.doi.org/10.2307/1806101, arXiv:1806101.
Oldfield, R.H., Bly, P.H., 1988. An analytic investigation of optimal bus size. Transp. Res. Part B 22 (5), 319–337. http://dx.doi.org/10.1016/0191-2615(88)90038-0.
Olmos, L.E., Çolak, S., Shafiei, S., Saberi, M., González, M.C., 2018. Macroscopic dynamics and the collapse of urban traffic. Proc. Natl. Acad. Sci. 115 (50), 12654–12661. http://dx.doi.org/10.1073/pnas.1800474115.
Paipuri, M., Leclercq, L., 2020. Bi-modal macroscopic traffic dynamics in a single region. Transp. Res. B 133, 257–290. http://dx.doi.org/10.1016/j.trb.2020.01.007.
Pandey, A., Lehe, L., 2023. Congestive Mode-Switching and Economies of Scale on a Bus Route, (no. 4471220), Rochester, NY, http://dx.doi.org/10.2139/ssrn.4471220.
Pandey, A., Lehe, L.J., Gayah, V.V., 2024. Local stability of traffic equilibria in an isotropic network. Transp. Res. B 179, 102873. http://dx.doi.org/10.1016/j.trb.2023.102873.

Safadi, Y., Fu, R., Quan, Q., Haddad, J., 2023. Macroscopic fundamental diagrams for low-altitude air city transport. Transp. Res. C 152, 104141. http://dx.doi.org/10.1016/j.trc.2023.104141.

Small, K.A., Chu, X., 2003. Hypercongestion. J. Transp. Econ. Policy (JTEP) 37 (3), 319–352.

Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation, second ed. Routledge, London, http://dx.doi.org/10.4324/9780203642306.

Strogatz, S.H., 2019. Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering, second ed. CRC Press, Boca Raton, http://dx.doi.org/10.1201/9780429492563.

Tirachini, A., 2013. Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers. Transport. A: Transp. Sci. 9 (1), 28–49. http://dx.doi.org/10.1080/18128602.2010.520277.

Tirachini, A., 2014. The economics and engineering of bus stops: Spacing, design and congestion. Transp. Res. Part A: Policy Pract. 59, 37–57. http://dx.doi.org/10.1016/j.tra.2013.10.010.

Tirachini, A., Hensher, D.A., 2011. Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. Transp. Res. B 45 (5), 828–844. http://dx.doi.org/10.1016/j.trb.2011.02.006.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. Transp. Res. Part A: Policy Pract. 53, 36–52. http://dx.doi.org/10.1016/j.tra.2013.06.005.

Tirachini, A., Hensher, D.A., Rose, J.M., 2014. Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding. Transp. Res. B 61, 33–54. http://dx.doi.org/10.1016/j.trb.2014.01.003.

Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A., 2017. Estimation of crowding discomfort in public transport: Results from santiago de Chile. Transp. Res. Part A: Policy Pract. 103, 311–326. http://dx.doi.org/10.1016/j.tra.2017.06.008.

Tsekeris, T., Geroliminis, N., 2013. City size, network structure and traffic congestion. J. Urban Econ. 76, 1–14. http://dx.doi.org/10.1016/j.jue.2013.01.002.

Turvey, R., Mohring, H., 1975. Optimal bus fares. J. Transp. Econ. Policy 9 (3), 280–286, arXiv:20052415. URL https://www.jstor.org/stable/20052415.

Verhoef, E.T., 2005. Speed-flow relations and cost functions for congested traffic: Theory and empirical analysis. Transp. Res. Part A: Policy Pract. 39 (7), 792–812. http://dx.doi.org/10.1016/j.tra.2005.02.023.

Vickrey, W., 2020. Congestion in midtown Manhattan in relation to marginal cost pricing. Econ. Transp. 21, 100152. http://dx.doi.org/10.1016/j.ecotra.2019.100152.

Walters, A.A., 1961. The theory and measurement of private and social cost of highway congestion. Econometrica 29 (4), 676–699. http://dx.doi.org/10.2307/1911814, arXiv:1911814.

Watling, D., 1998. Perturbation stability of the asymmetric stochastic equilibrium assignment model. Transp. Res. B 32 (3), 155–171. http://dx.doi.org/10.1016/S0191-2615(97)00022-2.

Watling, D., 1999. Stability of the stochastic equilibrium assignment problem: A dynamical systems approach. Transp. Res. B 33 (4), 281–312. http://dx.doi.org/10.1016/S0191-2615(98)00033-2.

Xu, G., Gayah, V.V., 2023. Non-unimodal and non-concave relationships in the network macroscopic fundamental diagram caused by hierarchical streets. Transp. Res. B 173, 203–227. http://dx.doi.org/10.1016/j.trb.2023.05.002.

Xu, Z., Yin, Y., Ye, J., 2020. On the supply curve of ride-hailing systems. Transp. Res. B 132, 29–43. http://dx.doi.org/10.1016/j.trb.2019.02.011.

Zhang, L., Garoni, T.M., de Gier, J., 2013. A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems. Transp. Res. B 49, 1–23. http://dx.doi.org/10.1016/j.trb.2012.12.002.

Zhang, J., Yang, H., Lindsey, R., Li, X., 2020. Modeling and managing congested transit service with heterogeneous users under monopoly. Transp. Res. B 132, 249–266. http://dx.doi.org/10.1016/j.trb.2019.04.012.

Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transp. Res. B 57, 326–341. http://dx.doi.org/10.1016/j.trb.2013.06.003.