Hindawi
Complexity
Volume 2022, Article ID 8996439, 15 pages
https://doi.org/10.1155/2022/8996439

Research Article

Synchronising Bus Bunching to the Spikes in Service Demand Reduces Commuters' Waiting Time

Luca Vismara[1,2] **Vee-Liem Saw**[2] **and Lock Yue Chew**[2]

[1]*Interdisciplinary Graduate Programme, 61 Nanyang Drive, Nanyang Technological University, Singapore*
[2]*Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, 21 Nanyang Link, Nanyang Technological University, Singapore*

Correspondence should be addressed to Lock Yue Chew; lockyue@ntu.edu.sg

Received 16 March 2022; Revised 12 July 2022; Accepted 1 August 2022; Published 30 August 2022

Academic Editor: Ning Cai

Bus bunching is ostensibly regarded as a detrimental phenomenon in bus systems. We study a bus loop with two bus stops, one regular bus stop and one spike bus stop, where bunched buses can outperform staggered buses. The spike bus stop models a bus stop connected to a train or metro service, where passengers arrive in groups at periodic intervals (spikes). We introduce the configuration of synchronised bunched buses, where bunched buses wait for the spike in demand. For a wide range of parameters, synchronised bunched buses outperform perfectly staggered buses in terms of minimising the waiting time of commuters. We present an analytical formulation of the average waiting time in the case of bunched buses, synchronised bunched buses, and perfectly staggered buses with the consideration of different passenger demands, number of buses, and bus capacity. We demonstrate the validity of our analytical results through an agent-based simulation of the bus loop system.

1. Introduction

Bus bunching is generally regarded as undesirable in bus systems [1–13]. As explained in References [14–16], waiting time for passengers is reduced by a staggered bus configuration when passengers arrive continuously at bus stops. A common technique used to achieve a staggered configuration is holding [2, 4, 6, 16, 17]. As the headway between two buses is diverging from the ideal staggered condition, the bus approaching holds at its current bus stop after all of the passengers board until the ideal headway is restored. Dynamically controlling buses is challenging as instructions need to be given in real-time to drivers, and specific infrastructure, such as tracking devices on buses, are needed for the system to work. It is therefore important to explore alternative nondynamical techniques that can reduce waiting time without the complications of dynamically adjusting bus behaviour. Important results are achieved in the context of route and service planning. A common strategy is stop-skipping [18–21] where buses do not board or let people alight at certain bus stops. Another solution is having a demand-responsive transit system [22–24] where passengers make a request for a transportation service from their location to their destination. Such techniques require assumptions regarding passengers' behaviours and knowing the origin-destination pattern of commuters. A recent result by our group [25] shows a significant reduction in waiting time using a bus assignment technique that does not require any assumption regarding passenger behaviour. In this work, we want to focus instead on an alternative strategy in bus deployment that leverages the time periodicity of spikes in demand and does not require dynamic controlling nor changing the assignment of buses to bus stops. In the transportation literature, the common assumption regarding passenger arrival is that the number of passengers arriving at a bus stop grows continuously and linearly with time [12, 16, 25] or according to a Poisson process [11]. Here, we want to study the case of a system with a *spike bus stop*.

Passengers arrive at a *spike bus stop* in batches of p at a time at periodic intervals of T_s unit times. A real-world example of a bus stop having this property is a bus stop connected to a rapid transit system or a train station with periodic service. Passengers waiting at the bus stop are still increasing with time, as in the cases examined by the literature, but they increase periodically in groups of p, rather than one by one or continuously. To the knowledge of the authors, this kind of bus stop has not been studied before in the literature. Allowing p to be stochastic makes the model more realistic, however, for analytical tractability and to give physical insights, we consider p as a constant in this paper.

In the literature, spikes in demand have been studied in terms of seasonal variations [26, 27] in the context of network planning.

In this work, we study the case of a bus loop with one *spike bus stop* and one *regular bus stop*, as defined in Section 2. The *regular bus stop* is analogous to the bus stops examined [2, 14, 25]. Through analytical models, we compute the average waiting time for three configurations: bunched buses, synchronised bunched buses (as defined in Section 2.2), and fully staggered buses. For the bunched configurations, we consider the steady state where buses are already bunched and we ignore the transient. The calculations are performed for the case of unlimited bus capacity in Section 2 and by explicitly considering buses with limited capacity in Section 3. We compare the three bus configurations in Section 4 and we show that bunched buses synchronising with the spike in passenger demand can outperform perfectly staggered buses in the regime of high demand in the *spike bus stop* and when bus capacity is limited. We also briefly generalise the results for a loop with multiple *regular bus stops* in Section 2.4 and for higher frequency spikes in Section 2.5. In Section 4.4, we validate the theoretical results with a simulation. In Section 5, we summarise the results obtained and draw the conclusion and final remarks.

2. Three Scenarios with a *Spike Bus Stop*

The system we set to study is a bus loop with two bus stops. At the first bus stop, passengers arrive at a constant rate of s per unit time. The second bus stop is called *spike bus stop* and passengers arrive in spikes of p people every T_s unit time. We aim to build an idealised system that can be solved analytically to draw insights from the bus interactions. We do not include important factors such as traffic as we focus on understanding the underlying mechanism of the interactions between passengers and buses when a *spike bus stop* is present. Nonetheless, this scenario is inspired by a real bus system: the shuttle buses in Nanyang Technological University (NTU) that serve the residential part and the academic side of the campus in a loop. As classes usually finish at the hour or half an hour mark, the few bus stops at the academic buildings get very crowded at regular intervals, while the demand is more constant at the bus stops towards the residential part of the campus. Another scenario of interest is modelling a bus loop serving a town or a district connected to the rest of the transportation network via a low-frequency train service. We also briefly explore the case of high-frequency spikes in the system. The first kind of bus stop, which we call *regular bus stop*, is modelled in the conventional way [1, 2, 6, 7, 12, 14, 16, 25, 28–33]. Passengers arrive at a constant rate s: defining Δ_t as the time elapsed since that bus stop was lastly served by a bus, the number of passengers waiting is $s \times \Delta_t$. While it is more realistic to model passenger arrival as a stochastic process [6–8, 10, 11, 28], however, for analytical tractability, in this work, we consider deterministic passenger arrivals at a constant rate s. In reference [25], we test with simulations how the analytical results obtained under the assumption of a constant rate of arrival s also apply in the case of Poisson arrivals with the same rate s. In the analysis presented in this paper, we fix the rate at which passengers board as l passengers per unit time and we assume passengers alight instantaneously at the opposite bus stop, not affecting the dwelling time. To simplify the expressions, it is convenient to work with the quantities $k = s/l$ and $P = p/l$ to represent passengers arriving at bus stops. Without loss of generality, we set l as one passenger per unit time (which effectively defines the unit of time) and refer to $k = s/l$ as the arrival constant of the *regular bus stop* and to P as the number of passengers at the *spike bus stop*. We use nondimensionalised units. In this section, the capacity of the buses is unconstrained, while Section 3 explicitly considers the implications of limited capacity. All of the buses move at the same speed and take time T to complete the loop without counting the dwelling time at bus stops. If more than a bus is at the same bus stop, such as in a case of bus bunching, the load of passengers is equally shared, effectively multiplying the boarding rate. Notation wise, we use the square symbol $\square$ to refer to the *regular bus stop* and the triangle symbol Δ for the *spike bus stop*. We measure distances in units of time, as the speed of the buses is fixed.

The following three sections explore the three configurations of buses tested in the bus loop defined above:

A: Bunched buses

B: *Synchronised bunching*

C: Perfectly staggered buses

Case B, "Synchronised bunching," is a novel approach that aims at being easier to implement compared with dynamic control techniques, exploiting the periodicity of the arrivals at the *spike bus stop*. The definition of this tactic is given in Section 2.2.

In the following sections, we calculate analytically the average waiting time for passengers at bus stops as a function of the crowdedness of the *regular bus stop* $k = s/l$, the number of people arriving at the *spike bus stop* $P = p/l$, and their period of arrival T_s for N buses. We consider spikes with a relatively long period T_s, such that at most one spike occurs during a revolution.

2.1. Bunched Buses. It is well known that uncontrolled buses bunch [3, 12, 14, 33], so this is a natural baseline and an important case to consider as bus bunching is not an uncommon occurrence in real-world bus systems.

The general idea to compute the average waiting time is to separate the contribution of the *regular* $\square$ and *spike* Δ bus stops.

For the *regular bus stop*, using the fact that passengers board in a FIFO way (the first to arrive is the first to board), the average waiting time at the *regular bus stop* is half the waiting time of the passenger who waited for the longest, who is the first to arrive at the bus stop after the last platoon of buses left the bus stop in the previous revolution. Passengers arrive and board in a linear fashion, and the last passenger to board is the one who arrives just before the platoon of buses leaves, giving a waiting time of zero; hence, the average is half the longest waiting time. The same idea is also used to compute the average waiting time in reference [25]. The linear nature of passenger arrival also implies that for a longer average waiting time at a *regular bus stop*, more passengers will be waiting, and the range of waiting times, from the longest to the shortest, is twice the average waiting time. The waiting time at that bus stop is

$$W_A^{\square} = \frac{1}{2}\left(\overline{T}_A - \tau_A^{\square}\right) = \frac{\overline{T}_A}{2}\left(1 - \frac{k}{N}\right). \tag{1}$$

The quantity $\overline{T}_A$ represents the effective time taken for the buses to complete the loop, including dwelling time. It is defined in equation (5). The time spent dwelling at the bus stop is $\tau_A^{\square}$ and it is proportional to the total passengers arrived, hence proportional to the time it takes to complete a revolution $\overline{T}_A$, times the arrival constant $k = s/l$. Having N buses bunching, the dwelling time is reduced by a factor $1/N$ as passengers can board in parallel the N buses speeding up the process. To compute the waiting time for the whole bus loop, the waiting time at each bus stop has to be weighted by the passengers boarding. To keep a consistent notation with k, we use the number of passengers boarding per revolution divided by the constant boarding rate l:

$$Ppl_A^{\square} = k\overline{T}_A. \tag{2}$$

Our calculations are in the steady state. However, it is possible to compute the waiting time dynamically for uncontrolled buses before they bunch. In reference [14], we present an analytical framework that allows us to compute waiting times dynamically for *regular bus stops*.

In the case of the *spike bus stop*, passengers arrive in a group of $P = p/l$. They all wait for the buses to arrive. At every revolution, the waiting time will be different, unless T_s and $\overline{T}_A$ happen to be exactly in resonance (in this work, we do not cover this possibility), so it is a reasonable assumption to consider an average value. Assuming that the buses are already in a bunched state, and for analysis purpose we have the platoon of buses distributed uniformly around the bus loop when the spike happens with the arrival of the passengers at the *spike bus stop*, the average time until the buses reach the bus stop is $\delta^{\Delta} = \overline{T}_A/2$. On top of δ^{Δ} time to wait, passengers need to board. The first passenger to board does not need to wait any longer, but the last passenger to board has to wait P/N units of extra time, as boarding is conducted in parallel with N buses, so, on average, the extra waiting is $1/2 \times P/N$:

$$W_A^{\Delta} = \delta^{\Delta} + \frac{P}{2N} = \frac{\overline{T}_A}{2} + \frac{P}{2N}. \tag{3}$$

Considering that passengers arrive at the *spike bus stop* every T_s, the average number of passengers boarding from this bus stop over a revolution is

$$Ppl_A^{\Delta} = P\frac{\overline{T}_A}{T_s}. \tag{4}$$

The final equation needed is for $\overline{T}_A$, the average time taken to complete a revolution, which comprises of the time T needed to drive, and the time spent at the bus stops boarding passengers $\tau_A^{\square}$ and τ_A^{Δ}:

$$\overline{T}_A = T + \tau_A^{\Delta} + \tau_A^{\square} = T + \frac{P\overline{T}_A}{NT_s} + \frac{k\overline{T}_A}{N} = \frac{T}{1 - P/NT_s - k/N}. \tag{5}$$

From the denominator of equation (5), we see the two conditions of maximum crowdedness at which the bunched buses get stuck at one or the other bus stop. For the *regular bus stop*, N buses can board N passengers per unit time, hence if passengers arrive at a rate $\geq N$ per unit time ($k \geq N$), the buses cannot finish boarding and $\overline{T}_A \longrightarrow \infty$. For the *spike bus stop*, as buses require P/N unit times to board all the passengers from the spike, if $P/N \geq T_s$ a new spike arrives before the *spike bus stop* is cleared, blocking the buses and causing $\overline{T}_A$ to diverge. The average waiting time in the bus loop is computed combining equations (1)–(5):

$$\begin{aligned} W_A &= \frac{Ppl_A^{\Delta} \times W_A^{\Delta} + Ppl_A^{\square} \times W_A^{\square}}{Ppl_A^{\Delta} + Ppl_A^{\square}} \\ &= \frac{P^2 + N\overline{T}_A\left(P + k(1 - k/N)\right)T_s}{2N\left(P + kT_s\right)}. \end{aligned} \tag{6}$$

2.2. Synchronised Bunched Buses. In this section, we introduce a novel approach to bus control. As dynamic control often requires specific infrastructure, complex algorithms, and nonlocal information such as the position of other buses [1–13, 20, 32–40], we propose a much simpler yet effective technique: hold the platoon of bunched buses at the *spike bus stop* until the spike happens. It is well known that buses tend to bunch, but bunched buses have a distinct advantage: they can distribute the load of passengers among them, speeding up boarding and reducing dwelling time. Another advantage is that N buses can accommodate N times as many passengers, although the problem of limited bus capacity is analysed in Section 3.2. This technique, which we call *synchronised bunching*, synchronises the effective time taken by buses to complete a loop (including dwelling time) $\overline{T}_B$ with the period of the spike T_s.

To calculate the waiting time in this scenario, we compute the waiting time at each bus stop and weight them by the number of passengers boarded from there. With the *regular bus stop*, the waiting time is computed in the same way as in equation (1), with the added condition that $\overline{T}_B = T_s$:

$$W_B^{\square} = \frac{1}{2}\left(T_s - \tau_B^{\square}\right) = \frac{T_s}{2}\left(1 - \frac{k}{N}\right). \tag{7}$$

Following the same idea as in equation (2),

$$Ppl_B^{\square} = kT_s. \tag{8}$$

The waiting time at the *spike bus stop* is very simple, as the passengers can start boarding immediately because the platoon of buses is waiting for the spike, so the only time to wait is the time it takes to board, which is equivalent to the second term in equation (3) as boarding happens in parallel for bunched buses.

$$W_B^{\Delta} = \frac{P}{2N}. \tag{9}$$

Given that the buses effectively take T_s time to complete a revolution, the number of passengers boarding per revolution is P.

$$Ppl_B^{\Delta} = P. \tag{10}$$

Combining the equations above, the average waiting time for passengers of the loop when the *synchronised bunching* technique is employed is

$$W_B = \frac{Ppl_B^{\Delta} \times W_B^{\Delta} + Ppl_B^{\square} \times W_B^{\square}}{Ppl_B^{\Delta} + Ppl_B^{\square}} = \frac{P^2 + kT_s^2 (N-k)}{2N(P + kT_s)}. \tag{11}$$

2.3. Perfectly Staggered Buses. The calculations in this section are performed under the assumption of perfectly staggered buses. As uncontrolled buses inevitably bunch, active control is necessary to keep buses staggered. There are several methods described in the literature to avoid bunching and keep a staggered configuration [1, 2, 4–8, 8–13, 17, 19, 35, 41–48]. The final result in terms of the average waiting time depends on the specific dynamic control technique used, so the average waiting time determined here is an approximation. In Section 4.4, we compare the analytical calculations with a time-based simulation, employing a headway-based holding control. We see very similar results in terms of waiting times in all but the most extreme cases with very high k where this approximation is less valid. In reference [14], we show that bunching can happen within a single revolution if the value of k is high enough that $n^* < 1$, where n^* is the number of revolutions needed for buses to bunch, hence invalidating the assumption that buses can stay staggered. Two buses starting from a perfectly staggered position in a loop with a single *regular bus stop* bunch in $n^* = \log(k/2)/(2\log(1-k))$ revolutions, according to reference [14]. For $k \geq 0.5$, uncontrolled buses bunch within the first revolution as $n^* \leq 1$. In this regime, keeping the buses staggered is very challenging, if not impossible. In Figure 1, the mismatch between a numerical simulation with dynamic control and the analytical results for the average waiting time is very significant when kT approaches $0.5T$, as buses cannot avoid bunching in the simulation.

The waiting time at the *regular bus stop* is computed as half the longest waiting time, in the same way as it was mentioned in Sections 2.1 and 2.2, but here the buses are assumed to be staggered, hence the bus stop is served N times every revolution.

$$W_C^{\square} = \frac{1}{2}\left(\frac{\overline{T}_C}{N} - \tau_C^{\square}\right) = \frac{1}{2}\frac{\overline{T}_C}{N}(1-k). \tag{12}$$

While boarding passengers from a bus stop, if other buses move, the system shifts away from a staggered configuration. To compute the average waiting time, we need to decide when the buses are considered perfectly staggered: either before or after they serve a bus stop. If buses are considered perfectly staggered *before* serving the bus stop, the expression for $W_C^{\square}$ is the one in equation (12). However, if buses are considered staggered *after* serving the bus stop, the average waiting time is $\tau_C^{\square}/2$ longer than equation (12). We choose the first option, as it leads to lower waiting time. Different techniques of dynamic control to keep the buses staggered could be better described by either of the two choices of buses staggered *before* or *after* serving a bus stop. This difference is significant in the regime of high $\tau_C^{\square} = k\overline{T}_C/N$ where the assumption that buses can be staggered also fails. Each revolution will take $\overline{T}_C$ as defined in equation (16). The number of passengers boarding from the *regular bus stop* during a revolution is computed as

$$Ppl_C^{\square} = k\overline{T}_C, \tag{13}$$

similarly to cases A and B in equations (2) and (8).

The average waiting time at the *spike bus stop* is calculated under the assumption that buses are fully staggered and uniformly distributed around the bus loop, hence the closest bus to the bus stop, when the spike arrives, is at an average distance (in units of time) $\delta_{\text{closest}}^{\Delta} = \overline{T}_C/(2N)$ since each bus is separated by $\overline{T}_C/N$ unit times.

$$W_C^{\Delta} = \delta_{\text{closest}}^{\Delta} + \frac{P}{2} = \frac{\overline{T}_C}{2N} + \frac{P}{2}. \tag{14}$$

The last term $P/2$ accounts for the time it takes to board the P passengers, since the last one to board has to wait an additional P units of time. The quantity W_C^{Δ} has to be weighted by the average number of passengers boarding from the *spike bus stop*:

$$Ppl_C^{\Delta} = P\frac{\overline{T}_C}{T_s}. \tag{15}$$

Finally, the average total time taken to complete a loop $\overline{T}_C$ is the sum of T and the average dwelling time at the bus stops:

$$\overline{T}_C = T + \tau_C^{\Delta} + \tau_C^{\square} = T + \frac{P\overline{T}_C}{T_s} + \frac{k\overline{T}_C}{N} = \frac{T}{1 - P/T_s - k/N}. \tag{16}$$

The dwelling time at the *spike bus stop*τ_C^{Δ} is counted, even though only one of the N buses boards passengers there. The reason for including τ_C^{Δ} is to account for active corrective actions by other buses, such as holding or slowing down, to maintain the perfectly staggered state. If that is not compensated for, the bus boarding at the *spike bus stop* would have a slower revolution as compared with the other buses, causing buses to not be staggered anymore. The average waiting time for passengers in the bus loop is computed as

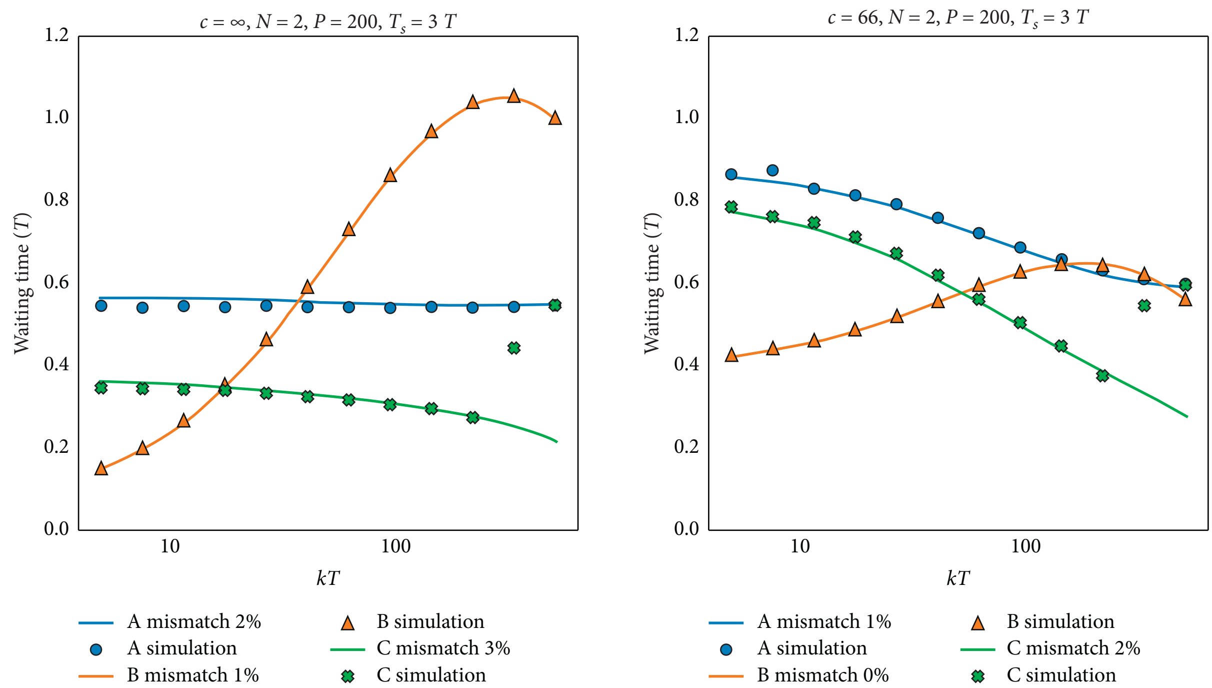

Figure 1: Comparison of the average waiting time of passengers computed with analytical formulas and with a numerical simulation. The plot on the left considers the case of unlimited capacity described in Section 2, while the plot on the right considers a limited capacity of $c = \lfloor P/3 \rfloor$ and uses the formulas from Section 3. The circles, triangles, and crosses represent the results from the simulation for the cases A, B, and C, respectively. Discrepancies are caused by simplified assumptions, discretisation of variables in the simulation, and limited running time. The percentage of mismatch in the legend is the median of the difference between the simulation and the analytical result across the range kT, rounded to the closest integer. As explained in Section 2.3, the hypothesis of perfectly staggered buses C fails at high values of k, as we see from the plot. In this example, there are two buses $(N = 2)$, the period of revolution is $T = 1000$ time units, the period of the spike is $T_s = 3T$, and $P = 200$.

the weighted average of the waiting time at the two bus stops, equations (12) and (14).

$$W_C = \frac{Ppl_C^{\Delta} \times W_C^{\Delta} + Ppl_C^{\square} \times W_C^{\square}}{Ppl_C^{\Delta} + Ppl_C^{\square}} = \frac{NP^2 + \overline{T}_C\left(P + k(1-k)T_s\right)}{2N\left(P + kT_s\right)}. \tag{17}$$

2.4. Generalisation for More Bus Stops. In Section 2, so far we have considered a scenario with only two bus stops for simplicity. It is possible to study a system with multiple *regular bus stops* with our technique. The general formula for the average waiting time with M*regular bus stops* and one *spike bus stop* is

$$W = \frac{Ppl^{\Delta} \times W^{\Delta} + \sum_{i=1}^{M} Ppl^{\square_i} \times W^{\square_i}}{Ppl^{\Delta} + \sum_{i=1}^{M} Ppl^{\square_i}}. \tag{18}$$

The quantities to compute are the average waiting time at bus stop i for all the M *regular bus stops*$W^{\square_i}$ and the people boarding from there $Ppl^{\square_i}$. Moreover, stopping in multiple bus stops means that the total time taken to complete a loop, $\overline{T}$, will need to take into account the extra time taken to board from more than one *regular bus stops*, except for the case of synchronised bunched buses in Section 2.2 where the total time to complete a loop is the period of the spike T_s. In the following part, we present the general expressions for the waiting time at the M *regular bus stops*, each of which with demand $s_i = k_i \times l$ for the three cases examined: bunched buses, synchronised bunching, and perfectly staggered buses.

2.4.1. Bunched Buses. Using the same reasoning as in equation (1), the waiting time at each *regular bus stops* in the case of N bunched buses is

$$W_A^{\square_i} = \frac{1}{2}\left(\overline{T}_A - \tau_A^{\square_i}\right) = \frac{\overline{T}_A}{2}\left(1 - \frac{k_i}{N}\right). \tag{19}$$

The passengers boarding at each bus stop are

$$Ppl_A^{\square_i} = k_i \overline{T}_A. \tag{20}$$

The total time taken to complete the loop is

$$\overline{T}_A = T + \tau_A^{\Delta} + \sum_{i=1}^{M} \tau_A^{\square_i} = T + \frac{P\overline{T}_A}{NT_s} + \sum_{i=1}^{M} \frac{k_i \overline{T}_A}{N}. \tag{21}$$

The waiting time at the *spike bus stop* is indirectly affected by $\overline{T}_A$. The average waiting time of passengers in a loop with M *regular bus stops* and one *spike bus stop* is computed via equation (18) by combining equations (19)–(21) along with W_A^{Δ} and Ppl_A^{Δ} from equations (3) and (4).

2.4.2. Synchronised Bunched Buses. Following the idea used in Section 2.2, the waiting time at each *regular bus stop* in the case of N synchronised bunched buses is

$$W_B^{\square_i} = \frac{T_s - \tau_B^{\square_i}}{2} = \frac{T_s}{2}\left(1 - \frac{k_i}{N}\right). \tag{22}$$

The passengers boarding at each bus stop are

$$Ppl_B^{\square_i} = k_i T_s. \tag{23}$$

The total time taken to complete the loop is still T_s, provided that the extra dwelling time does not slow down the revolution of the buses below the period of the spike T_s. The waiting time at the *spike bus stop* is not affected by the extra bus stops. By combining the equations in this section with equations (9) and (10) and substituting them in equation (18), we can compute the average waiting time of passengers in a loop with M *regular bus stops* and one *spike bus stop* in the case of synchronised bunched buses, generalising the result in Section 2.2.

2.4.3. Perfectly Staggered Buses. Analogously as how it is calculated in equation (12), the waiting time at *regular bus stops* in the case of N perfectly staggered buses is

$$W_C^{\square_i} = \frac{1}{2}\frac{\overline{T}_C}{N}(1 - k). \tag{24}$$

The passengers boarding at each bus stop are

$$Ppl_C^{\square_i} = k_i \overline{T}_C. \tag{25}$$

The total time taken to complete the loop is

$$\overline{T}_C = T + \tau_C^{\Delta} + \sum_{i=1}^{M} \tau_C^{\square_i} = T + \frac{P\overline{T}_C}{T_s} + \sum_{i=1}^{M} \frac{k_i \overline{T}_C}{N}. \tag{26}$$

The waiting time at the *spike bus stop* is indirectly affected by $\overline{T}_C$ but takes the same functional form as in equation (14). Combining the equations in this section along with equation (15) and substituting them in equation (18), it is possible to compute the average waiting time of passengers in a loop with M *regular bus stops* and one *spike bus stop* in the case of perfectly staggered buses, generalising the result in Section 2.3.

2.5. Generalisation for High-Frequency Spikes. We have considered low-frequency spikes $T_s > \overline{T}$ where $\overline{T}$ is the total time taken to complete a loop. However, there are realistic cases where spikes can occur at high frequency, such as next to a "Mass Rapid Transit" line (MRT). In this section, we briefly generalise the results for bunched buses (case A) and perfectly staggered buses (case C) for $T_s < \overline{T}$. As the frequency of the spikes is higher than the frequency of the buses, the synchronised bunched buses (case B) solution does not work in this setting as buses cannot synchronise with the spikes.

2.5.1. Bunched Buses. When buses are bunched, and more than one spike occurs during a loop, the resulting waiting time at the *spike bus stop* has to be corrected from the result in equation (3). The waiting time in equation (3) has two terms, the first, $\overline{T}_A/2$, is the waiting time due to the spike arriving when the buses are not at the *spike bus stop*, the second, $P/(2N)$, is due to the time passengers take to board the buses. Not knowing when the spikes arrive during the revolution of the platoon of buses, we cannot simply assume that the passengers arriving with the spike wait, on average, $\overline{T}_A/2$ in this case. A spike may occur while the buses are boarding at the *spike bus stop*, in which case we consider their waiting time contribution from the first term as zero as passengers only have to wait in line to board, which is accounted for by the second term of W_A^{Δ}. We split the first term of the waiting time into two parts: the contribution from the passengers arriving while the buses are on the road, hence equally likely to arrive at any time in $\overline{T}_A - \tau_A^{\Delta}$, and the contribution from passengers arriving while the bus is already boarding from the previous spikes, hence with zero waiting time from the first term. We define $\tilde{\tau}_A^{\Delta} = (\overline{T}_A/T_s - 1)P/N$ as the time spent to board from all the spikes minus one (the one that may occur during boarding at the *spike bus stop*). It is our assumption that no more than one spike can happen while the buses are boarding. We can compute the new first term for the average waiting time at the *spike bus stop* by averaging the two contributions with probability proportional to the time spent in each situation:

$$\begin{aligned} W_A^{\Delta}(\text{first term}) &= \frac{0 \times \tilde{\tau}_A^{\Delta} + \left(\overline{T}_A - \tau_A^{\Delta}/2\right) \times \left(\overline{T}_A - \tilde{\tau}_A^{\Delta}\right)}{\tilde{\tau}_A^{\Delta} + \left(\overline{T}_A - \tilde{\tau}_A^{\Delta}\right)} \\ &= \frac{1}{2}\frac{\left(\overline{T}_A - \tau_A^{\Delta}\right)\left(\overline{T}_A - \tilde{\tau}_A^{\Delta}\right)}{\overline{T}_A}. \end{aligned} \tag{27}$$

The second term of W_A^{Δ} accounts for the time taken to board the buses. Since more than P passengers board at the same time, it has to be corrected from equation (3) with a factor $\overline{T}_A/T_s$. The second term of the average waiting time at the *spike bus stop* becomes $(P\overline{T}_A/T_s)/(2N)$. The overall expression for W_A^{Δ} for frequent spikes $T_s < \overline{T}_A$ is:

$$W_A^{\Delta} = \frac{1}{2}\frac{\left(\overline{T}_A - \tau_A^{\Delta}\right)\left(\overline{T}_A - \tilde{\tau}_A^{\Delta}\right)}{\overline{T}_A} + \frac{\overline{T}_A}{T_s}\frac{P}{2N}. \tag{28}$$

All the other formulas in Section 2.1 are not affected by higher frequency spikes as the increased number of passengers is accounted for by the $\overline{T}_A/T_s$ factor in the expressions for the waiting time and total time to complete a revolution and for the number of passengers. The overall waiting time can be computed combining equations (1), (2), (4), (5), and (28).

The equations above are valid as long as a maximum of a single spike occurs while buses are boarding at the *spike bus stop*. It is interesting to explore the limit of infinitely many spikes $\overline{T}_A/T_s \longrightarrow \infty$, by $T_s \longrightarrow 0$, even though the equations are not exact in this regime. The total number of passengers arriving over a period $\overline{T}_A$ must be finite. The quantity $\widehat{P} = P/T_s$ is the number of passengers arriving at the *spike bus stop* per unit time. As spikes are infinitely frequent, the number of passengers arriving with each spike must be infinitesimal $P \longrightarrow 0$, effectively representing a continuous arrival of $\widehat{P}$ passengers per unit time. From the terms of equation (28), we see that, in the limit $T_s \longrightarrow 0$:

$$\begin{aligned} \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} \widetilde{\tau}_A^{\Delta} &= \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} \left(\frac{\overline{T}_A}{T_s} - 1\right) \\ \frac{P}{N} &= \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} \frac{\overline{T}_A}{T_s}\frac{P}{N} = \frac{\widehat{P}\overline{T}_A}{N} = \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} \tau_A^{\Delta}, \end{aligned} \tag{29}$$

where the last equation uses the definition of $\widehat{P}$. We can therefore express the waiting time at the *spike bus stop* in this limit as

$$\begin{aligned} \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} W_A^{\Delta} &= \lim_{\substack{T_s \longrightarrow 0 \\ P/T_s=\widehat{P}}} \frac{1}{2}\frac{\left(\overline{T}_A - \tau_A^{\Delta}\right)^2}{\overline{T}_A} \\ &+ \frac{\overline{T}_A\widehat{P}}{2N} = \frac{\overline{T}_A}{2}\left(1 - \frac{\widehat{P}}{N} + \frac{\widehat{P}^2}{N^2}\right). \end{aligned} \tag{30}$$

Compared with equation (1) for $W_A^{\square}$, there is an extra term $(\widehat{P}/N)^2$, due to our assumption that no more than one spike can occur during boarding at the *spike bus stop*. The term is generally small since $\widehat{P}/N < 1$ in order for buses to board passengers, as N buses can board N passengers per unit time. By comparing the analytical result with the waiting time computed via the simulation employed in Section 4.4, we verify that the correct limit for $W_A^{\square}$ is actually $\overline{T}_A/2 \times (1 - \widehat{P}/N)$, identical to $W_A^{\square}$ if $\widehat{P}$ is replaced by k, making $\widehat{P}$ effectively an arrival constant. The number of passengers boarding, from (4), becomes $Ppl_A^{\Delta} = \widehat{P}\overline{T}_A$ in this limit, analogous to $Ppl_A^{\square}$ replacing $\widehat{P}$ with k.

2.5.2. Perfectly Staggered Buses. For perfectly staggered buses, case C, the formulas for the waiting time at the *regular bus stop* and *spike bus stop* equations (12) and (14) are not directly affected. In particular, the waiting time and passengers at the *regular bus stop* do not directly depend on T_s or P, but they are indirectly affected through $\overline{T}_C$. The expression of the number of passengers at the *spike bus stop* already contains a term $\overline{T}_C/T_s$ that accounts for the frequency of the spikes. The waiting time at the *spike bus stop* is also not affected directly, as long as the number of spikes per loop is less than the number of buses, in which case the second term of $W_A^{\square}$, accounting for the waiting time due to boarding, $P/2$, has to be corrected. If $\overline{T}_C/T_s > N$, buses sometimes have to pick up passengers arrived from two or more different spikes; hence, the formula in equation (14) has to be modified. We do not consider this extreme case in this section. All the terms above, however, depend on $\overline{T}_C$. The expression for $\overline{T}_C$ in equation (16) has to be corrected for frequent spikes. In particular, as $\overline{T}_C = T + \tau_C^{\Delta} + \tau_C^{\square}$, the time spent at the *spike bus stop* τ_C^{Δ} changes. For low-frequency spikes, $\tau_C^{\Delta} = P\overline{T}_C/T_s$, meaning that every $\overline{T}_C/T_s$ revolution P passengers board. However if $\overline{T}_C/T_s > 1$, passengers from each spike still board a single bus, as passengers arrive P at a time, so the time to board τ_C^{Δ} becomes $\max(\overline{T}_C/T_s, 1)P = P$ time units, contributing to the total time to complete a revolution. The correct expression of $\overline{T}_C$ in the case of frequent spikes is therefore

$$\overline{T}_C = T + \tau_C^{\Delta} + \tau_C^{\square} = T + P + \frac{k\overline{T}_C}{N} = \frac{T + P}{1 - k/N}. \tag{31}$$

The average waiting time for the whole system W_C is obtained following equation (17) using $\overline{T}_C$ computed from equation (31) and the original equations (12)–(15).

It is well known [1–16] that perfectly staggered buses outperform bunched buses in terms of average waiting time when passengers arrive continuously at bus stops, i.e. in the limit $T_s \longrightarrow 0$ and $P/T_s = \widehat{P}$ finite. In Section 4, we compare the proposed solutions in the case of low-frequency spikes: bunched buses also perform poorly as compared with perfectly staggered buses, as it is shown in Figures 1–3. For this reason, we do not consider the case of high-frequency spikes in the rest of this work, as the only configuration that can outperform perfectly staggered buses is the synchronised bunched buses setting (case B), which is not possible in the presence of high-frequency spikes.

3. The Effect of Limited Capacity for the Buses

Real-world transportation systems have limited capacity regarding the number of commuters who can board. Such limitations can alter the optimal despatch of vehicles and headway [49–52], hence staggered solutions are not necessarily optimal even in the presence of only *regular bus stops*. Reference [53] deals more explicitly with bunching buses in the form of a newly proposed modular vehicle system. The modular vehicles can be combined to increase capacity, which is similar to the effect of bunched buses. The system however is studied in the context of a splitting route where different modules travel to different destinations: a passenger has to board a specific module and none of the cases cited above deals with loops or *spike bus stops*.

In this section, we only consider the capacity limit at the *spike bus stop*. All of the buses can board up to Q passengers at the *spike bus stop*. For dimensional consistency with our convention for $P = p/l$, we consider capacity over the fixed boarding rate l to simplify our equation with the quantity $c = Q/l$.

Limiting capacity has a stronger effect on staggered buses, as at any given time passengers board a single bus. Performance of perfectly staggered buses is affected when $c < P$ while for bunched buses, having a combined capacity Nc, no effect is seen until $c < P/N$. Figure 3 shows how, for lower values of c, synchronised bunched buses tend to outperform perfectly staggered buses. The major effect on the bus dynamics when capacity is limited is the inability of boarding passengers at the *spike bus stop* within one passage.

3.1. Bunched Buses. One of the advantages of bunched buses is the cumulative capacity of the platoon, which grows linearly with the number of buses. Limited capacity affects this scenario only when $c < P/N$, if $c \geq P/N$ the result in equation (6) applies. In this section, we present the result in the regime $P/(2N) < c < P/N$. In this case, the buses cannot board all of the passengers at the *spike bus stop*, so they need to complete another revolution to board passengers from there. For that to be possible, the spikes cannot be too frequent, so a necessary condition is $T_s \geq 2\overline{T}_{A_{\text{lim}}}$, where $\overline{T}_{A_{\text{lim}}}$ is the total time needed to complete a revolution. If the inequality above is not satisfied, more passengers are arriving than those whom the buses can pick up, hence the number of passengers waiting will grow at every revolution, leading to a diverging waiting time.

The major difference, as compared with the case with unlimited capacity in Section 2.1, is the waiting time at the *spike bus stop* as two passages are needed to board passengers at the *spike bus stop*. The waiting time has to be divided between the Nc passengers boarding at the first passage and the $P - Nc$ boarding at the second passage, approximately a revolution after the first group.

$$W^{\Delta}_{A_{\text{lim}}} = \frac{Nc}{P}\left(\delta^{\Delta}_{\text{first}} + \frac{Nc}{2N}\right) + \frac{P-Nc}{P}\left(\delta^{\Delta}_{\text{second}} + \frac{P-Nc}{2N}\right) = \frac{Nc}{P}\left(\frac{\overline{T}_{A_{\text{lim}}}}{2} + \frac{Nc}{2N}\right) + \frac{P-Nc}{P}\left(\frac{3\overline{T}_{A_{\text{lim}}}}{2} + \frac{P-Nc}{2N}\right). \tag{32}$$

Following the same idea as in Section 2.1, we define $\delta^{\Delta}_{\text{first}}$ and $\delta^{\Delta}_{\text{second}}$ as the distance (in units of time) of the buses when they serve the *spike bus stop* for the first and second time, respectively. Being the buses bunched, the second time the buses pick up passengers in the *spike bus stop* is a whole revolution after the first time, therefore there is a difference of $\overline{T}_{A_{\text{lim}}}$ between $\delta^{\Delta}_{\text{first}}$ and $\delta^{\Delta}_{\text{second}}$. The waiting time of passengers picked up by the buses between the two revolutions is weighted by the passengers boarded, Nc at the first revolution and $P - Nc$ in the second round. The average number of passengers boarded per revolution is

$$Ppl^{\Delta}_{A_{\text{lim}}} = P\frac{\overline{T}_{A_{\text{lim}}}}{T_s}, \tag{33}$$

as P passengers arrive once every $T_s/\overline{T}_{A_{\text{lim}}}$ revolution.

The values of $W^{\square}_{A_{\text{lim}}}$ and $Ppl^{\square}_{A_{\text{lim}}}$ are the same as in equations (1) and (2), with $\overline{T}_A$ replaced by $\overline{T}_{A_{\text{lim}}}$: $W^{\square}_{A_{\text{lim}}} = \overline{T}_{A_{\text{lim}}}/2(1 - k/N)$ and $Ppl^{\square}_{A_{\text{lim}}} = k\overline{T}_{A_{\text{lim}}}$. To compute the average time taken to complete a revolution, $\overline{T}_{A_{\text{lim}}}$, we consider the average time spent at the *spike bus stop* during a revolution. Every $T_s/\overline{T}_{A_{\text{lim}}}$ revolution, a spike happens, and for every spike, the buses need to stop to pick up Nc and $P - Nc$ passengers at the first and second revolution, respectively. Hence, the average time to complete a revolution is

$$\overline{T}_{A_{\text{lim}}} = T + \tau^{\Delta(1)}_{A_{\text{lim}}} + \tau^{\Delta(2)}_{A_{\text{lim}}} + \tau^{\square}_{A_{\text{lim}}} = T + \frac{\overline{T}_{A_{\text{lim}}}}{T_s}\frac{Nc}{N} + \frac{\overline{T}_{A_{\text{lim}}}}{T_s}\frac{P-Nc}{N} + \frac{k\overline{T}_{A_{\text{lim}}}}{N} = \frac{T}{1 - (P/\mathrm{NT}_s) - (k/N)}. \tag{34}$$

Combining equations (1), (2), and (32)–(34), the average waiting time for passengers in a scenario with bunched buses and reduced capacity $P/(2N) < c < P/N$ is

$$W_{A_{\text{lim}}} = \frac{N\overline{T}_{A_{\text{lim}}}\left(3P - 2Nc + kT_s(1 - k/N)\right) + (Nc)^2 + (P - Nc)^2}{2N(P + kT_s)}. \tag{35}$$

It is possible to generalise the results for $c < P/(2N)$ by adding extra terms in equation (32) using $\delta^{\Delta}_{i-th} = \overline{T}_{A_{\text{lim}}}/2 \times (2i - 1)$. The value of $\overline{T}_{A_{\text{lim}}}$ does not change since it does not depend on c, as seen in equation (34).

3.2. Synchronised Bunched Buses. Similarly to the bunched configuration in Section 3.1, limiting the capacity of the buses in this synchronised bunched setting affects the dynamics and the waiting time of buses only in the regime $c < P/N$. In this section, we consider only the case of $P/(2N) < c < P/N$ where the platoon of bunched buses waits for the spike of passengers at the *spike bus stop*, but the capacity is not enough to board them all, so another revolution is needed to pick up the remaining passengers. We also assume that the buses can always board all of the passengers at the *regular bus stop*. In the same way as the case in Section 3.1, the spikes need to have a long enough period T_s, in such a way that the buses can pick up all the passengers at the spike bus stop before a new spike happens. The formal condition is $T_s \geq \overline{T}^{(1)}_{B_{\text{lim}}} + \overline{T}^{(2)}_{B_{\text{lim}}}$ where $\overline{T}^{(1)}_{B_{\text{lim}}}$ represents the time taken to complete the first revolution after the spike and $\overline{T}^{(2)}_{B_{\text{lim}}}$ is the time taken to complete the second revolution, just before holding $\text{Hold}_{B_{\text{lim}}}$ unit times at the *spike bus stop*, waiting for the new spike. Those quantities are defined in equations (39) and (40) and related to T_s according to the equation $T_s = \overline{T}^{(1)}_{B_{\text{lim}}} + \overline{T}^{(1)}_{B_{\text{lim}}} + \text{Hold}_{B_{\text{lim}}}$.

To compute the waiting time, we can break the dynamics into three parts: first revolution, second revolution, and holding to synchronise with the next spike. The first part starts from the arrival of passengers at the *spike bus stop*, where the platoon of buses is waiting to board them. The first part ends when the *spike bus stop* is reached a second time. In our equations, it is indicated with the (1) notation and it lasts $\overline{T}^{(1)}_{B_{\text{lim}}}$ unit times. The second part starts when the platoon of buses picks up the remaining passengers at the *spike bus stop* and it ends a revolution later, when the buses reach the *spike bus stop* again, just before starting to hold to wait for the next spike. We indicate the quantities relative to this period with the index (2) and this revolution lasts $\overline{T}^{(2)}_{B_{\text{lim}}}$ unit times. The final phase consists of the buses holding at the *spike bus stop* waiting for a new spike to occur.

To compute the waiting time at the *regular bus stop*, we average the waiting time the first and the second time the buses reach this bus stop. At the first revolution (1), the bus stop has not been served for a time $\text{Hold}_{B_{\text{lim}}} + \overline{T}^{(1)}_{B_{\text{lim}}} - \tau^{\square(1)}_{B_{\text{lim}}} = T_s - \overline{T}^{(2)}_{B_{\text{lim}}} - \tau^{\square(2)}_{B_{\text{lim}}}$ and the average waiting time, under the hypothesis of constant arrival at rate $s = k \times l$, is half the longest time waited by a passenger, hence

$$W_{B_{\lim}}^{\square(1)} = \frac{1}{2}\left(T_s - \overline{T}_{B_{\lim}}^{(2)} - \tau_{B_{\lim}}^{\square(1)}\right) = \frac{T_s - \overline{T}_{B_{\lim}}^{(2)}}{2}\left(1 - \frac{k}{N}\right). \tag{36}$$

The passengers boarding in the first loop are $Ppl_{B_{\lim}}^{\square(1)} = k(T_s - \overline{T}_{B_{\lim}}^{(2)})$. The same procedure applies to the waiting time of the second period (2) where the *regular bus stop* has not been served for $\overline{T}_{B_{\lim}}^{(2)} - \tau_{B_{\lim}}^{\square(2)}$ units of time.

$$W_{B_{\lim}}^{\square(2)} = \frac{1}{2}\left(\overline{T}_{B_{\lim}}^{(2)} - \tau_{B_{\lim}}^{\square(2)}\right) = \frac{\overline{T}_{B_{\lim}}^{(2)}}{2}\left(1 - \frac{k}{N}\right), \tag{37}$$

with $Ppl_{B_{\lim}}^{\square(2)} = k\overline{T}_{B_{\lim}}^{(2)}$ passengers boarded, for a total of $Ppl_{B_{\lim}}^{\square} = kT_s$ passengers boarding from the *regular bus stop* during the two revolutions.

The waiting time at the *spike bus stop* also needs to be calculated averaging the waiting time $W_{B_{\lim}}^{\Delta(1)}$ of the first $Ppl_{B_{\lim}}^{\Delta(1)} = Nc$ passengers who can board immediately after the spike during revolution (1), so only the time taken to board is considered in the waiting time, and the waiting time $W_{B_{\lim}}^{\Delta(2)}$ of the remaining $Ppl_{B_{\lim}}^{\Delta(2)} = P - Nc$ that needs to wait $\overline{T}_{B_{\lim}}^{(1)}$ until the first revolution is completed. The average waiting time at the *spike bus stop* is computed as

$$\begin{aligned} W_{B_{\lim}}^{\Delta(1)} &= \frac{Nc}{2N}, \\ W_{B_{\lim}}^{\Delta(2)} &= \overline{T}_{B_{\lim}}^{(1)} + \frac{P - Nc}{2N}, \\ W_{B_{\lim}}^{\Delta} &= \frac{Nc}{P} W_{B_{\lim}}^{\Delta(1)} + \frac{P - Nc}{P} W_{B_{\lim}}^{\Delta(2)}. \end{aligned} \tag{38}$$

The waiting time computed requires an expression for the time taken to complete the revolutions (1) and (2). The first case is expressed as

$$\overline{T}_{B_{\lim}}^{(1)} = T + \tau_{B_{\lim}}^{\Delta(1)} + \tau_{B_{\lim}}^{\square(1)} = T + c + \frac{k}{N}\left(T_s - \overline{T}_{B_{\lim}}^{(2)}\right). \tag{39}$$

The revolution includes T unit times on the road, $\tau_{B_{\lim}}^{\Delta(1)}$, to board from the spike bus stop the first Nc passengers, and $\tau_{B_{\lim}}^{\square(1)}$ to board passengers at the *regular bus stop*. Passengers at the *regular bus stop* started arriving during the previous (2) revolution, after the platoon of buses finished to board at the *regular bus stop*. The number of passengers arriving at the *regular bus stop* during the first revolution (1) (and the dwelling time $\tau_{B_{\lim}}^{\square(1)}$) is proportional to the sum of the holding time waited to synchronise with the spike after revolution (2) and $\overline{T}_{B_{\lim}}^{(1)}$. The quantity $\overline{T}_{B_{\lim}}^{(1)}$ is defined as the time taken to board passengers at the *spike bus stop*, the time taken to board up to the last passengers at the *regular bus stop* during the first revolution (1), and the time to complete the loop itself, T: $\tau_{B_{\lim}}^{\square(1)} \propto \mathrm{Hold}_{B_{\lim}} + \overline{T}_{B_{\lim}}^{(1)}$. As the duration of the two revolutions (1) and (2) plus the time $\mathrm{Hold}_{B_{\lim}}$ waited at the *spike bus stop* must be T_s in order to synchronise with the spike, $\mathrm{Hold}_{B_{\lim}} + \overline{T}_{B_{\lim}}^{(1)} = T_s - \overline{T}_{B_{\lim}}^{(2)}$.

The time taken to complete the second revolution (2) before holding is

$$\begin{aligned} \overline{T}_{B_{\lim}}^{(2)} &= T + \tau_{B_{\lim}}^{\Delta(2)} + \tau_{B_{\lim}}^{\square(2)} \\ &= T + \frac{P - Nc}{N} + \frac{k}{N}\overline{T}_{B_{\lim}}^{(2)} \\ &= \frac{T + P - Nc/N}{1 - k/N}. \end{aligned} \tag{40}$$

The dwelling time taken to board at the *spike bus stop* is proportional to the number of passengers left $P - Nc$. For the case of the *regular bus stop*, the number of passengers, hence the dwelling time, are proportional to $\overline{T}_{B_{\lim}}^{(2)}$ itself, as there is no waiting for synchronisation in between serving the *regular bus stop* during the first (1) and second (2) revolution in this setting.

Combining the above expressions with $Ppl_{B_{\lim}}^{\square} = kT_s$ and $Ppl_{B_{\lim}}^{\Delta} = P$, the average waiting time for synchronised bunched buses with capacity $P/(2N) < c < P/N$ is found from

$$W_{B_{\lim}} = \frac{Ppl_{B_{\lim}}^{\Delta} \times W_{B_{\lim}}^{\Delta} + Ppl_{B_{\lim}}^{\square} \times W_{B_{\lim}}^{\square}}{Ppl_{B_{\lim}}^{\Delta} + Ppl_{B_{\lim}}^{\square}}. \tag{41}$$

The average waiting time for synchronised bunched buses is equal to that of a single bus with capacity Nc and boarding rate Nl. Assuming that it is possible to build and deploy such a bus, N synchronised bunched buses are still more versatile and adaptable to a change in demand. First, N buses can implement active control and move to a staggered configuration when the passengers demand at the bus stops changes. As we see from Figures 2 and 3, the more advantageous configuration can be either perfectly staggered buses or synchronised bunched buses, depending on the parameters of the bus loop. Another advantage of N bunched buses over a single bus with equivalent capacity and boarding rate is the option of adding and removing extra buses to match the passenger demand and capacity constraints as the number of commuters in the system changes, reducing operational cost when less capacity is needed.

3.3. Perfectly Staggered Buses. As the *spike bus stop* is being served by one bus at a time in this configuration of perfectly staggered buses, the limited capacity affects the system for any $c < P$. In the previous two cases in Sections 3.1 and 3.2, the limited capacity affects the system only for $c < P/N$.

We consider the case where m buses are needed to board the total number of passengers at the *spike bus stop* $(P/m < c < P/(m-1))$. The value for m is $\lceil P/c \rceil$. As this case requires m/N revolutions to clear the passengers at the spike bus stop, a necessary condition for the buses to serve this system is $T_s > (m/N)T$, otherwise the number of waiting commuters blows up. Following the same reasoning as in equations (14) and (32),

$$\begin{aligned} W_{C_{\lim}}^{\Delta} &= \frac{c}{P}\left(\delta_{1st}^{\Delta} + \frac{c}{2}\right) + \frac{c}{P}\left(\delta_{2n\,d}^{\Delta} + \frac{c}{2}\right) + + \ldots \\ &+ \frac{P - (m-1)c}{P}\left(\delta_{m-th}^{\Delta} + \frac{P - (m-1)c}{2}\right), \end{aligned} \tag{42}$$

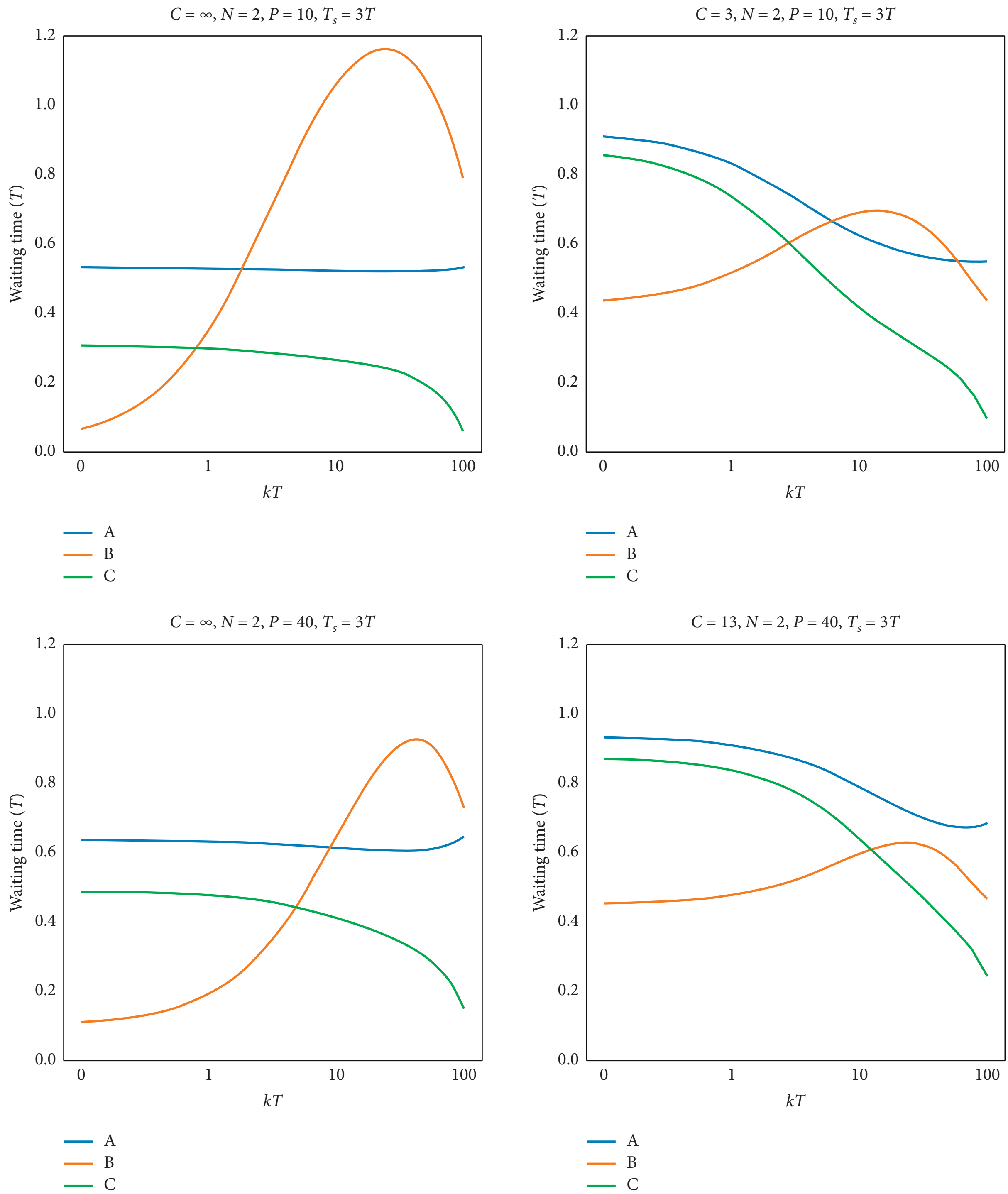

FIGURE 2: Average waiting time for two buses in the three configurations A, B, C described in Sections 2 and 3. The two upper plots represent a case of low demand P from the *spike bus stop* while the two lower plots represent high demand P. The plots on the left assume unlimited capacity of the buses, and the plots on the right explicitly limit the capacity of buses at $c = \lfloor P/3 \rfloor$ while boarding passengers from the *spike bus stop*. As discussed in Section 4, synchronised bunched buses (configuration B) outperform perfectly staggered buses (configuration C) when the demand of the *spike bus stop* is high as compared with the *regular bus stop*. Limiting capacity further increases the advantage of bunched buses. For this example, T is set to a value of 100 time units and $T_s = 3T$.

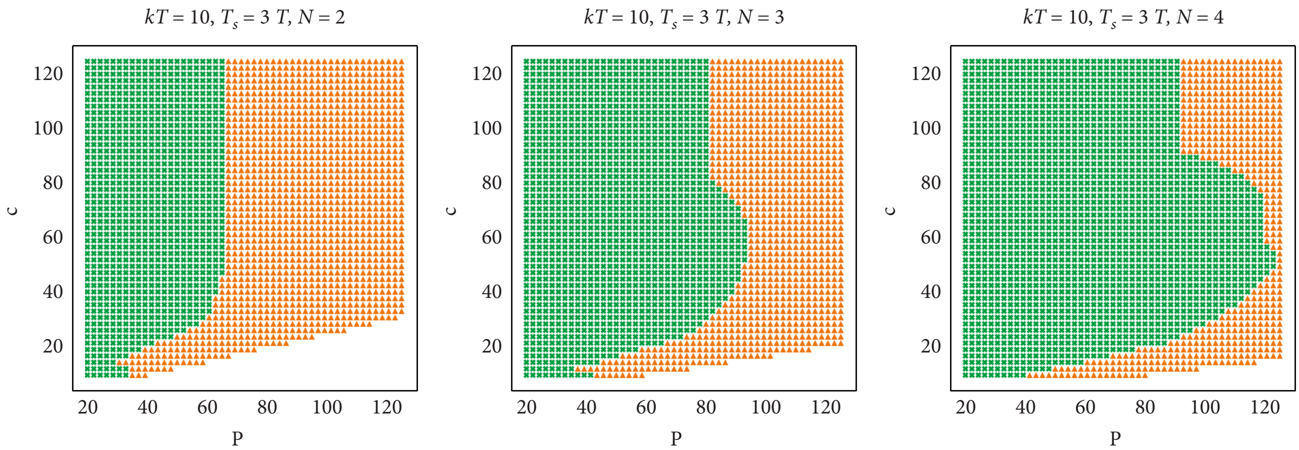

Figure 3: Comparison of the average waiting time for the three configurations as a function of the demand at the *spike bus stop* P and the capacity c. The orange triangles represent the set of parameters where synchronised bunched buses have the lowest average waiting time, while the green "x" symbols indicate where perfectly staggered buses have the lowest average waiting time among the three configurations. Bunched buses (case A) are never the best performers for the parameters explored. The three plots show how increasing the number of buses from 2 to 3 to 4 (from left to right) tends to favour the staggered bus configuration. Points in the region of high P and low c are missing, since in that regime more than two revolutions are needed to pick up all the passengers at the *spike bus stop*. This region is defined by $c < P/(2N)$. For this example, T is set to a value of 100 time units and $T_s = 3T$ while k is fixed at 0.1.

where $\delta^{\Delta}_{i-th} = \overline{T}_{C_{\lim}}(2i-1)/(2N)$ is the distance (in units of time) of the ith bus to serve the *spike bus stop*. As in the case with unlimited capacity, the number of people boarding is $Ppl^{\Delta}_{C_{\lim}} = P\overline{T}_{C_{\lim}}/T_s$.

The waiting time and people boarded from the *regular bus stop* are also the same as in equations (12) and (13): $W^{\square}_{C_{\lim}} = \overline{T}_{C_{\lim}}/(2N)(1-k)$ and $Ppl^{\square}_{C_{\lim}} = k\overline{T}_{C_{\lim}}$. To express the average waiting time in the whole system, the last remaining equation is for $\overline{T}_{C_{\lim}}$. As in the case of unlimited capacity in equation (16), boarding passengers at the *regular bus stop* takes $\tau^{\square}_{C_{\lim}} = k\overline{T}_{C_{\lim}}/N$ for each bus, under the assumption of perfectly staggered buses and no capacity constraints at the *regular bus stop*. At the *spike bus stop*, buses will dwell for either c or $P-(m-1)c$ unit times if there are passengers there. For perfectly staggered buses, the necessary condition is for all the buses to move with the same period, hence they all have to employ dynamic control to keep themselves staggered, slowing down at the pace of the slowest buses. Since $c > P-(m-1)c$, by definition of $m = \lceil P/c \rceil$, the slowest bus waits c unit times, and that happens $T_s/\overline{T}_{C_{\lim}}$ times a loop. The expression for the average time taken to complete a loop is then

$$\begin{aligned}\overline{T}_{C_{\lim}} &= T + \tau^{\Delta}_{C_{\lim}} + \tau^{\square}_{C_{\lim}} \\ &= T + \frac{c\overline{T}_{C_{\lim}}}{T_s} + \frac{k\overline{T}_{C_{\lim}}}{N} \\ &= \frac{T}{1 - c/T_s - k/N}.\end{aligned} \tag{43}$$

The general expression depends on the value of m via $W^{\Delta}_{C_{\lim}}$. We report the average waiting time in the case of $m = 2$, hence in the regime $P/2 < c < P$, but the equations above allow for a general calculation via equation (18).

$$W_{C_{\lim}} = \frac{N\left(c^2 + (P-c)^2\right) + \overline{T}_{C_{\lim}}\left(3P - 2c + k(1-k)T_s\right)}{2N\left(P + kT_s\right)}. \tag{44}$$

4. Comparison and Discussion

In a bus loop with a *spike bus stop* and a *regular bus stop*, six important variables affect the system: the crowdedness of the bus stops, through P and k, the period of the spikes T_s, the period for a revolution without dwelling T, the number of buses N, and the capacity of buses c.

4.1. Limit Cases of P and k. First, as the perfectly staggered buses are the best solution in terms of waiting time for a *regular bus stop* [14, 15], we expect that for lower demand of the *spike bus stop* as compared with the *regular bus stop*, the staggered configuration will outperform the two bunched configurations examined in this work. Computing the average waiting time of the three cases with unrestricted capacity in the limit $P = 0$ in equations (6), (11), and (17), we have

$$W_A(P=0) = \overline{T}_A\left(1 - \frac{k}{N}\right) = \frac{T}{2}, \tag{45}$$

$$W_B(P=0) = \frac{T_s}{2}\left(1 - \frac{k}{N}\right), \tag{46}$$

$$W_C(P=0) = \frac{\overline{T}_C}{2N}(1-k) = \frac{T}{2}\frac{1-k}{N-k}. \tag{47}$$

The waiting time for bunched buses, case A, is fixed as $T/2$ since as long as passengers can be boarded faster than they arrive ($k < N$), no passenger has to wait for more than a whole revolution T to board. The synchronised bunched buses in case B behave in the same way, except that having to wait at the *spike bus stop*, the time taken to complete a revolution is $T_s \geq \overline{T}_A$, where the inequality is a necessary condition for synchronised bunched buses, so $W_A(P=0) \leq W_B(P=0)$. In the last case of perfectly staggered buses, C, the waiting time is always the lowest, as expected. The quantity $(1-k)/(N-k)$ is always less than or equal to 1, since $k<1$ is a necessary condition for staggered buses, so $W_C(P=0) \leq W_A(P=0) \leq W_B(P=0)$. In the case of a single bus $N=1$, case A and case C are equivalent. Taking the limit $P=0$ leads to the same result as taking the limit for $T_s \longrightarrow \infty$ for the case of bunched buses A and perfectly staggered buses C. For synchronised bunched buses, this limit does not make sense, as the buses would have to wait indefinitely at the *spike bus stop*.

In the opposite limit where all demand is concentrated at the *spike bus stop* $(k=0)$, we expect the synchronised bunched buses to outperform the other configurations as the platoon of buses waits for the spike to arrive, eliminating any delay between arrival and beginning of boarding process. From equations (6), (11), and (17), in the setting of unconstrained capacity:

$$\begin{aligned} W_A(k=0) &= \frac{P+N\overline{T}_A}{2N} \\ &= \frac{P}{2N} + \frac{T}{2}\frac{1}{1-P/NT_s}, \end{aligned} \tag{48}$$

$$W_B(k=0) = \frac{P}{2N}, \tag{49}$$

$$\begin{aligned} W_C(k=0) &= \frac{NP+\overline{T}_C}{2N} \\ &= \frac{P}{2} + \frac{T}{2N}\frac{1}{1-P/T_s}. \end{aligned} \tag{50}$$

It takes P unit times to board P passengers for perfectly staggered buses C, so if another spike occurs at $T_s \leq P$, the number of passengers waiting grows with time, either diverging or causing the buses to bunch, against the hypothesis of perfectly staggered buses, hence we consider $T_s \geq P$. Synchronised buses B always perform the best among the three, since $1-P/(NT_s)$ and $1-P/T_s$ are necessarily positive. For cases A and C, in the limit of small $P \ll T < T_s$, the waiting time is dominated by the time needed for a bus to reach the bus stop, so bunched buses have the longest waiting time as, on average, buses are further away from the *spike bus stop* $W_A(k=0, P \ll T) \approx T/2 > W_C(k=0, P \ll T) \approx T/(2N)$. In the opposite limit, where $P \longrightarrow T_s$, perfectly staggered buses cannot keep up with the demand while staying staggered, hence $W_C(k=0, P \longrightarrow T_s) \longrightarrow \infty$. A similar analysis can be performed in the case of constrained capacity c from the equations of the waiting time in Section 3.

4.2. Intermediate Values of P and k. Intermediate cases of k and P are less intuitive, and introducing the capacity limit c makes the expressions less simple. To understand and compare the behaviour of the buses in such a scenario, we visualise the effect that the aforementioned variables have on the system. First, the average waiting time as a function of k is explored in Figure 2 where the three methods (bunched buses, synchronised bunched buses, and perfectly staggered buses) are compared in a case with high and low demand from the *spike bus stop P*, in two conditions of unlimited and limited capacity. As expected from the analytical comparison above, synchronised bunched buses outperform perfectly staggered buses in the regime of low demand kT from the *regular bus stop*. The comparison is done with kT to have dimensional consistency with P. The curve for the synchronised bunched buses B from Figure 2 is particularly interesting. The average waiting time increases with k for low values of k as expected, but after reaching a peak, the waiting time decreases in the regime of high k. The downward trend can be explained from equations (7) and (9). The contribution to the average waiting time from the *spike bus stop*, in the case of synchronised bunched buses B, does not depend on k, nor on the number of passengers boarded from the *spike bus stop P*. The number of passengers boarded at the *regular bus stop* $Ppl_B^{\square}\infty k$ increases with k, while P is constant, hence eventually the contribution at the *regular bus stop* dominates for high k, showing the downward trend $\infty(1-k/N)$ since $W_B^{\square}\infty(1-k/N)$. An intuitive explanation is that at larger k the buses spend more time at the *regular bus stop* and less time waiting for the spike at the *spike bus stop*. In all of the cases examined in Figure 2, bunching buses A are never the best performing configuration regarding minimising the average waiting time.

The impact of the limited capacity is visualised in Figure 3 where the three configurations of buses are compared, changing the number of buses N from 2 to 4. In the range of parameters explored, bunched buses are always outperformed by synchronised bunched buses or perfectly staggered buses, in terms of average waiting time. While the case of two buses in the leftmost plot in Figure 3 seems to follow the intuition that low capacity c and high spike demand P favour synchronised bunched buses over perfectly staggered buses, we see that for $N=3$ and $N=4$ the situation is less intuitive for intermediate values of c. The explanation of this phenomenon has to do with the fact that lowering the capacity c in the case of perfectly staggered buses lowers the average time that buses take to complete a loop $\overline{T}_C$ since each bus has to stop less time at the *spike bus stop*, according to equation (43). This reduction in $\overline{T}_C$ is directly proportional to the reduction in waiting time at the *regular bus stop* via equation (12) and indirectly reduces the waiting time at the *spike bus stop* via equation (42) since the δ^{Δ}_{i-th} is proportional to $\overline{T}_C$. This reduction of the waiting time is only effective up to a point where the reduction in waiting time due to a lower $\overline{T}_C$ is offset by the need for more and more buses to clear the passengers at the *spike bus stop*, through extra terms δ^{Δ}_{i-th} in equation (14). Fixing the maximum capacity c, synchronised bunched buses outperform perfectly staggered buses for values of P high enough.

4.3. Applicability. Another comparison point pertains to the applicability of the three different methods in the real world. Bunched buses are stable [3, 12], in the sense that perturbations in the system and in the initial conditions do not influence the long-term configuration as the buses will tend to bunch. Perfectly staggered buses, on the other hand, require active dynamic intervention to keep the headway between buses constant. Such intervention can be implemented through holding [1–13, 17, 19, 35, 41–48, 54], limiting boarding [28, 30, 55–58], or stop-skipping [18, 19, 21, 32, 59, 60]. Dynamic control requires specific infrastructure and it can be challenging to implement as instructions need to be provided in real-time to bus drivers, hence a technique that involves bunched buses has the advantages of being more robust to perturbations and, at the same time, easier to deploy.

4.4. Comparison with a Simulation. To validate our analytical results, we compare the average waiting time found with our approach in Sections 2 and 3 with a time-based numerical simulation of the scenarios described. The simulation is the same time-based simulation described in reference [25] and employed to test bus bunching in reference [14]. One of the challenges of the simulation is in translating a result built on continuous variables in a discrete simulation, where the smallest unit is the unit of time. A convenient way to do so is to use small units of time, hence we define T as 1000 units of time. The comparison with the simulation is given in Figure 1. The plot on the left considers unlimited capacity c while the plot on the right compares formulas and simulation for limited c. The main discrepancies with the simulation are at very low kT, where the discretisation of time still plays a role, and at very high kT for the perfectly staggered buses. This has to do with the assumption we made in Section 2.3 about the buses being perfectly staggered. In the simulation, this is implemented by holding buses at the *spike bus stop* whenever the headway is not perfectly staggered, which happens after the first bus picks up passengers at the *spike bus stop*. There are many choices and many ways of implementing dynamic control, one of which is described in reference [2], but none of them can keep the buses perfectly staggered at all times, making our assumption valid only approximately. Nonetheless, our simulation matches the waiting time predicted by the formulas within 3% for most of the range of parameters explored in all three configurations analysed, both for limited and unlimited capacity.

5. Conclusion

A bus loop with a *spike bus stop* creates a situation where perfectly staggered buses may not be the configuration that minimises the average waiting time of passengers. If the passenger demand from the *spike bus stop* is large enough, bunched buses synchronised with the spike perform better than staggered buses. The advantage is stronger when bus capacity constraints are considered. The edges of bunched buses are faster passenger boarding and have higher effective capacity. When the bunched buses wait at the *spike bus stop* until the spike occurs, the waiting time of the passengers arriving at the *spike bus stop* is minimised, at the expense of the waiting time of passengers arriving at the *regular bus stop*. As discussed in Section 4, bunched configurations also have the advantage of being robust to external perturbations, as well as being easier to implement as compared with staggered buses.

While our models are idealised, they take inspiration from real-world situations, such as a bus stop connected to a train station or mass rapid transit. We prove that bunched buses can outperform perfectly staggered buses in certain scenarios if the buses are synchronised with the spike in passenger demand. The analytical results of the waiting time for the three bus configurations can be generalised to more complicated bus loops, as explained in Sections 2.4 and 2.5.

Data Availability

No data were used to support this study.

Disclosure

A version of the manuscript was published as a preprint [61] based on the link https://arxiv.org/abs/2201.02343.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Joint WASP/NTU Programme (M4082189).

References

[1] C. F. Daganzo, "A headway-based approach to eliminate bus bunching: systematic analysis and comparisons," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 913–921, 2009.

[2] V.-L. Saw, L. Vismara, and L. Y. Chew, "Intelligent buses in a loop service: Emergence of No-boarding and holding strategies," *Complexity*, vol. 2020, Article ID 7274254, 2020.

[3] L. Y. Chew, V.-L. Saw, and Yi En I. Pang, "Stability of Anti-bunched buses and local Unidirectional Kuramoto Oscillators," in *Recent Trends in Chaotic, Nonlinear and Complex Dynamics*, pp. 429–454, World Scientific, aug 2019.

[4] M. Abkowitz, I. Engelstein, and I. E. Record, "Methods for maintaining transit service regularity," *Transportation Research Record*, vol. 961, pp. 1–8, 1984.

[5] M. D. Rossetti and T. Turitto, "Comparing static and dynamic threshold based control strategies," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 8, pp. 607–620, 1998.

[6] M. D. Hickman, "An analytic stochastic model for the transit vehicle holding problem," *Transportation Science*, vol. 35, no. 3, pp. 215–237, 2001.

[7] L. Fu and X. Yang, "Design and Implementation of bus-holding control strategies with real-time information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1791, no. 1, pp. 6–12, 2002.

[8] O. Cats, A. N. Larijani, H. N. Koutsopoulos, and W. Burghout, "Impacts of holding control strategies on transit performance," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2216, no. 1, pp. 51–58, 2011.

[9] J. J. Bartholdi and D. D. Eisenstein, "A self-coördinating bus route to resist bus bunching," *Transportation Research Part B: Methodological*, vol. 46, no. 4, pp. 481–491, 2012.

[10] L. Moreira-Matias, O. Cats, J. Gama, J. Mendes-Moreira, and J. F. de Sousa, "An online learning approach to eliminate Bus Bunching in real-time," *Applied Soft Computing*, vol. 47, pp. 460–482, 2016.

[11] W. Liang Quek, N. Chung, V.-L. Saw, and L. Y. Chew, "Analysis and simulation of intervention strategies against bus bunching by means of an empirical agent-based model," *Complexity*, p. 2021, 2021.

[12] V.-L. Saw, N. N. Chung, W. Quek, Yi En I. Pang, and L. Y. Chew, "Bus bunching as a synchronisation phenomenon," *Scientific Reports*, vol. 9, no. 1, p. 6887, 2019.

[13] J. Wang and L. Sun, "Dynamic holding control to avoid bus bunching: a multi-agent deep reinforcement learning framework," *Transportation Research Part C: Emerging Technologies*, vol. 116, Article ID 102661, 2020.

[14] L. Vismara, V.-L. Saw, and L. Y. Chew, "Bunching dynamics of buses in a loop," in *Proceedings of the 2021 6th International Conference on Intelligent Transportation Engineering (ICITE 2021)*, pp. 203–212, ICITE, Springer, Singapore, June 2022.

[15] M. A. Ramli, V. Jayaraman, H. Kwek, K. H. Tan, G. L. Lee Kee Khoon, and C. Monterola, "Improved estimation of commuter waiting times using headway and commuter boarding information," *Physica A: Statistical Mechanics and its Applications*, vol. 501, pp. 217–226, 2018.

[16] E. Osuna and G. F. Newell, "Control strategies for an Idealized public transportation system," *Transportation Science*, vol. 6, no. 1, pp. 52–72, feb 1972.

[17] C. Gershenson, "Self-organization leads to supraoptimal performance in public transportation systems," *PLoS One*, vol. 6, no. 6, pp. 1–6, oct 2011.

[18] Y. Li, J.-M. Rousseau, and M. Gendreau, "Real-time scheduling on a transit bus route: a 0-1 stochastic programming model," *Proceedings of the thirty-third annual meeting, Transportation Research Forum*, no. CRT-772, pp. 157–166, 1991.

[19] C. E. Cortés, S. Sáez, F. Milla, A. Núñez, and M. Riquelme, "Hybrid predictive control for real-time optimization of public transport systems' operations based on evolutionary multi-objective optimization," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 757–769, 2010.

[20] V. Chiraphadhanakul and C. Barnhart, "Incremental bus service design: combining limited-stop and local bus services," *Public Transport*, vol. 5, no. 1, pp. 53–78, 2013.

[21] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, 2013.

[22] H. N. Psaraftis, "A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem," *Transportation Science*, vol. 14, no. 2, pp. 130–154, 1980.

[23] X. Li and L. Quadrifoglio, "Optimal Zone design for Feeder transit services," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 1, pp. 100–108, 2009.

[24] N. Ronald, R. Thompson, and S. Winter, "Simulating demand-responsive transportation: a review of agent-based approaches," *Transport Reviews*, vol. 35, no. 4, pp. 404–421, 2015.

[25] L. Vismara, L. Y. Chew, and V.-L. L. Saw, "Optimal assignment of buses to bus stops in a loop by reinforcement learning," *Physica A: Statistical Mechanics and its Applications*, vol. 583, Article ID 126268, 2021.

[26] S. M. M. Amiripour, A. Ceder, and A. S. Mohaymany, "Hybrid method for bus network design with high seasonal demand variation," *Journal of Transportation Engineering*, vol. 140, no. 6, Article ID 04014015, 2014.

[27] W. Wu, Li Peng, R. Liu et al., "Predicting peak load of bus routes with supply optimization and scaled shepard interpolation: a newsvendor model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, Article ID 102041, 2020.

[28] V.-L. Vismara and L. Y. Chew, "Intelligent buses in a loop service: Emergence of No-boarding and holding strategies," *Complexity*, vol. 2020, no. 3, pp. 1–18, 2020.

[29] V.-L. Saw, L. Vismara, and L. Y. Chew, "Chaotic semi-express buses in a loop," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 2, Article ID 023122, 2021.

[30] V.-L. Saw and L. Y. Chew, "No-boarding buses: agents allowed to cooperate or defect," *Journal of Physics: Complexity*, vol. 1, no. 1, Article ID 15005, 2020.

[31] V.-L. Saw, L. Vismara, and L. Y. Chew, "Inferring origin-destination distribution of agent transfer in a complex network using deep gated recurrent units," 2022, https://arxiv.org/abs/2207.13375.

[32] L. Fu, Q. Liu, and C. Calamai, "Real-time optimization model for dynamic scheduling of transit operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1857, no. 1, pp. 48–55, 2003.

[33] G. F. Newell, "Control of Pairing of vehicles on a public transportation route, two vehicles, one control point," *Transportation Science*, vol. 8, no. 3, pp. 248–264, aug 1974.

[34] A. Barnett, "On controlling Randomness in transit operations," *Transportation Science*, vol. 8, no. 2, pp. 102–116, may 1974.

[35] F. Alesiani and K. Gkiotsalitis, "Reinforcement learning-based bus holding for high-frequency services," in *Proceedings of the 21st International Conference on Intelligent Transportation Systems*, pp. 3162–3168, IEEE, Maui, HI, USA, November 2018.

[36] C. Leiva, J. C. Muñoz, R. Giesen, and H. Larrain, "Design of limited-stop services for an urban bus corridor with capacity constraints," *Transportation Research Part B: Methodological*, vol. 44, no. 10, pp. 1186–1201, 2010.

[37] Y. Y. Ulusoy, S. I.-Jy Chien, and C.-H. Wei, "Optimal all-stop, short-Turn, and express transit services under Heterogeneous demand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2197, no. 1, pp. 8–18, 2010.

[38] J. Chen, Z. Liu, S. Zhu, and W. Wang, "Design of limited-stop bus service with capacity constraint and stochastic travel time," *Transportation Research Part E: Logistics and Transportation Review*, vol. 83, no. 1–15, 2015.

[39] H. Larrain, J. C. Muñoz, and R. Giesen, "Generation and design heuristics for zonal express services," *Transportation Research Part E: Logistics and Transportation Review*, vol. 79, pp. 201–212, 2015.

[40] G. Soto, H. Larrain, and J. C. Muñoz, "A new solution framework for the limited-stop bus service design problem," *Transportation Research Part B: Methodological*, vol. 105, pp. 67–85, 2017.

[41] Xu J. Eberlein, H. M. Wilson, and D. Bernstein, “The holding problem with real-time information Available,” *Transportation Science*, vol. 35, no. 1, pp. 1–18, 2001.

[42] B. Bin, Z. Zhongzhen, and B. Baozhen, “Bus arrival time prediction using support Vector Machines,” *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151–158, 2006.

[43] N. Mukai, T. Watanabe, and J. Feng, *Route Optimisation Using Q-Learning for On-Demand Bus Systems*, Springer, Berlin, 2008.

[44] C. X. Chen, W. Y. Chen, and Z. Y. Chen, “A Multi-Agent Reinforcement Learning approach for bus holding control strategies,” *Advances in Transportation Studies*, vol. 2, pp. 41–54, 2015.

[45] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, J. C. Muñoz, and J. C. Munoz, “Planning, operation, and control of bus transport systems: a literature review,” *Transportation Research Part B: Methodological*, vol. 77, no. 38, pp. 38–75, jul 2015.

[46] W. Chen, K. Zhou, and C. Chen, “Real-Time bus holding control on a transit corridor based on multi-Agent reinforcement learning,” in *Proceedings of the IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, IEEE, Rio de Janeiro, November 2016.

[47] P. Wang, X. Chen, W. Chen, L. Cheng, and Da Lei, “Provision of bus real-time information: Turning passengers from being Contributors of headway Irregularity to Controllers,” *Transportation Research Record*, vol. 2672(8), 2018.

[48] K. Menda, Y. Chen, J. Grana et al., “Deep reinforcement learning for Event-Driven multi-agent Decision Processes,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1259–1268, 2019.

[49] G. F. Newell, “Dispatching policies for a transportation route,” *Transportation Science*, vol. 5, no. 1, pp. 91–105, 1971.

[50] M. Sadrani, A. Tirachini, and C. Antoniou, “Vehicle dispatching plan for minimizing passenger waiting time in a corridor with buses of different sizes: model formulation and solution approaches,” *European Journal of Operational Research*, vol. 8, 2021.

[51] H. Niu and X. Zhou, “Optimizing urban rail timetable under time-dependent demand and oversaturated conditions,” *Transportation Research Part C: Emerging Technologies*, vol. 36, no. 212–230, p. 11, 2013.

[52] Y. Zhu, B. Mao, Y. Bai, and S. Chen, “A bi-level model for single-line rail timetable design with consideration of demand and capacity,” *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 211–233, 12 2017.

[53] X. Shi, Z. Chen, M. Pei, and X. Li, “Variable-capacity operations with modular transits for shared-use corridors,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 9, pp. 230–244, 2020.

[54] Y. Xuan, J. Argote, and C. F. Daganzo, “Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis,” *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1831–1845, dec 2011.

[55] F. Delgado, J. C. Muñoz, R. Giesen, and A. Cipriano, “Real-time control of buses in a transit corridor based on vehicle holding and boarding limits,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2090, no. 1, pp. 59–67, 2009.

[56] F. Delgado, J. C. Munoz, and R. Giesen, “How much can holding and/or limiting boarding improve transit performance?” *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1202–1217, 2012.

[57] S. Z. Zhao, C. X. Lu, S. D. Liang, and H. S. Liu, “A self-adjusting method to resist bus bunching based on boarding limits,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 8950209, 2016.

[58] W. Sun and J.-D. Schmöcker, “Considering passenger choices and overtaking in the bus bunching problem,” *Transportation Business: Transport Dynamics*, vol. 6, no. 2, pp. 151–168, 2018.

[59] X. J. Eberlein, “Real-time control strategies in transit operations: models and analysis,” Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1995.

[60] A. Sun and M. Hickman, “The real-time stop-skipping problem,” *Journal of Intelligent Transportation Systems*, vol. 9, no. 2, pp. 91–109, 2005.

[61] L. Vismara, V.-L. Saw, and L. Y. Chew, “Can bus bunching reduce waiting time?,” 2022, https://arxiv.org/abs/2201.02343.