# An analysis on the traffic processing efficiency of a combination of serial and parallel bottlenecks

Wei-Liang Quek[a,b,*], Ning Ning Chung[a,b], Lock Yue Chew[a,b]

[a] *Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore*
[b] *Complexity Institute, Nanyang Technological University, Singapore 637723*

**Abstract**

By means of the Nagel-Schreckenberg model, we have investigated into the maximum vehicular flow rate of traffic processing bottlenecks. The evaluated analytical form of this flow rate is found to give quantitative insights into the underlying physics of collective vehicular motions constrained by these bottlenecks. Our analysis shows that for large-scale expansion, a new class of processing bottleneck known as the serial bottleneck is more efficient than the conventional parallel bottleneck in the absence of human driving behavior. When characteristics such as slow-to-start is considered in the model, the consequential delay due to human reaction time not only degrade the overall efficiency, it also diminishes the efficacy of serial processing such that a serial bottleneck is no longer tenable for traffic processing. These results point to the fundamental importance of optimizing traffic efficiency, which we illustrate by elucidating the detailed mechanisms with which vehicles interact collectively in the bottlenecks. In particular, we demonstrate that by constructing combinations of serial and parallel bottlenecks, optimal efficiencies are achieved via configurations with few (many) lanes of a large (small) number of serial units when the processing time is short (long). A direct implication of these results is that autonomous self-driving vehicles could serve to improve the transportation capacity for the densely pop-

*Corresponding author
*Email addresses:* s130019@e.ntu.edu.sg (Wei-Liang Quek ), nnchung@ntu.edu.sg (Ning Ning Chung), lockyue@ntu.edu.sg (Lock Yue Chew)
*URL:* http://www.ntu.edu.sg/home/lockyue/ (Lock Yue Chew)

ulated urban cities of the future, due to the intrinsically more efficient collective vehicular motions through these bottlenecks.

*Keywords:* NaSch model, Cellular automaton, Multi-point tollbooth, Lane Expansion, Fundamental diagram

## 1. Introduction

The Nagel-Schreckenberg (NaSch) model [1] is a cellular automaton that had been used extensively for theoretical traffic modelling. While it is simplistic and does not describe the full spectrum of real-world traffic phenomenology, it does capture the essential aspect of the emergence of congestion when vehicular density increases past a critical value [2]. In particular, this transition is known as the absorbing-state phase transition [3] when the random deceleration parameter $p = 0$. Active research on critical transition in traffic system invariably employs the NaSch model due to its analytical tractability [4]. In fact, one example of such phase transition occurs at the presence of bottleneck, with a *plateau* exhibited in the fundamental diagram. Traffic states at the plateau are observed to display the characteristics of phase-segregation in the spatiotemporal trajectories of the vehicles, with a continuous change from the free flow to the congestion state as the traffic density increases. Interestingly, this change is similar to the separation of liquid phase and vapor phase in a van der Waals fluid, giving the study of NaSch model with bottleneck a direct relevance to statistical physics [2].

In real-world traffic system, the presence of traffic bottlenecks is fundamental and they are inevitable sources of congestion. Traffic bottlenecks appear in various forms: as accident sites, merging of lanes, or even slow moving vehicles. In general, traffic bottleneck is defined as a localized section of the road or highway at which vehicles experience reduced speeds, and vehicles may require more time to transverse through these sites. This reduction in flow rate results in an increase in vehicle density, which in turn leads to the instability of the free flow phase due to the collective vehicular interactions, eventually leading

to traffic congestions at the bottlenecks [5].

While traffic bottlenecks can arise spontaneously and randomly such as in the case of a traffic accident, a vast majority of bottlenecks are fixed in their locations. Examples of these latter bottlenecks include tollbooths, highway's on/off-ramps, pick-up/drop-off points for public transport, and many more. Nowadays, such fixed-site bottlenecks are common sights in the traffic landscape of modern societies. Their purpose is to process the mobility of traffic flow, which is of interest in this paper. Since bottlenecks serve only to increase vehicular density which in turn destablizes the free flow phase, an important conclusion is that the effects of a bottleneck on traffic flow is independent of the specific bottleneck mechanism [6, 2, 17, 15]. The consequence is a maximum outflow through the bottleneck, which is known as the bottleneck capacity. Bottleneck capacity corresponds to the maximum value of the macroscopic traffic flow captured by the fundamental diagram in a closed, or sufficiently long stretch of road. Its universal feature enables a general study of bottlenecks by solving the maximal outflow through it, or by the time-headway approach adopted in this paper.

At present, there are numerous studies on traffic bottlenecks using NaSch model, such as the effects of time-delay bottlenecks [6], on- and off-ramps [7], and lane expansion at bottlenecks [8, 9]. These studies have been further facilitated by full analytical solutions of the bottleneck capacity, and compared against the flow-density relation of the homogeneous NaSch model in the deterministic limit, or in the $v_{\max} = 1$ limit [10, 11]. Solutions based on mean-field approximation have also been explored for general parameter range [12, 13]. The analytical solutions from these studies allow for greater quantitative insights into the physical mechanisms underlying traffic bottlenecks, which are not tenable from empirically consistent traffic models since the complex parameterisation and stochasticity within these models render such solutions almost impossible [2]. The analytical consequences of NaSch model as exhibited in these research give strength to the approach, and explains its continued relevance in current traffic studies even when there exist traffic models that are more empirically

consistent [14, 15, 16].

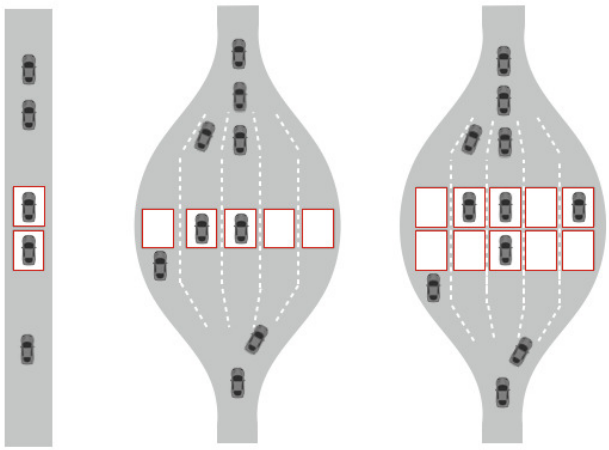


Figure 1: Schematic illustration of 3 types of bottleneck expansions: serial ($m = 2$); parallel ($n = 5$); and combination ($m = 2, n = 5$). Note that direction of travel is upwards

In this paper, we examine into a new type of bottleneck whose function is to process the mobility and transportation needs of modern society. In order to sustain the rapid growth of traffic demands, these *processing* bottlenecks are required to function efficiently. Our idea is a bottleneck that comprises multiple processing units that serve the same function, with both serial and parallel processing occurring at the same time. In other words, our proposed processing bottleneck combines two configurations: processing in parallel; and in serial (see Fig. 1).

Parallel expansion of traffic bottleneck involves splitting a lane into multiple lanes in the vicinity of a processing site, with each lane supporting a single processing unit. This is the typical method used to increase the processing capacity of tollbooths and immigration counters. Several studies were performed to probe the effect of parallel expansions of bottlenecks on traffic flow [18, 19, 8]. Notably, the capacity of a highway is found to be fully recoverable through parallel expansion of bottleneck [8]. On the other hand, serial expansion of traffic bottleneck allows multiple processing units to be co-located along a lane such that two or more vehicles are simultaneously processed. As an example,

a bus bay that extends over a length of two buses allows passengers from two buses to alight and board at the same time. This decreases the overall processing time of the two buses. Such an expansion of traffic bottleneck was explored in Ref. [20] to investigate the influence of manual versus electronic toll collection on traffic flow.

In the following sections, we first measure and compare the flow-density relation of a simplified NaSch model, modified with a parallel and a serial expansion of traffic bottlenecks. These values will be compared against the processing capacity of the different expansions, which would be solved analytically. Note that such characterization has yet to be achieved in the literature. Interestingly, the two expansions perform differently with each outperforming the other under certain conditions. One key advantage of a serial expansion is that it requires no expansion of lane to function. This raises the possibility of trading processing efficiency with the cost of lane expansion. For this, we examine into the mechanism underlying the discrepancy in efficiency of the two expansions. In addition, we quantify the difference between them in terms of various bottleneck parameters. After which, we proceed to analyze the performance of a traffic bottleneck that combines parallel and serial processing into a single multi-processing unit. In particular, we investigate into the conditions under which this combination bottleneck outperforms both the parallel and serial expansions in terms of traffic processing capacity.

## 2. Model and Results

In this paper, a bottleneck is defined as a section of road which comprises one or more processing units. The processing unit represents various functions and is modeled as a generic time-delay site. When a vehicle passes through the processing site, it would stop and be delayed for a time of $t_d$ before moving off. Traffic flow, on the other hand, is modeled based on NaSch cellular automata model [1] on a road of length $L$ cells, single-lane with periodic boundary condition. Each vehicle occupies a single site, and moves with a speed which is

less than or equal to the maximum velocity $v_m$. Vehicles with speed less than $v_m$ accelerate at a rate of 1 site/$s^2$ as long as there is sufficient unblocked cells ahead. To simplify mathematical analysis, we have set the random deceleration probability to $p = 0$. In all of our simulations, we let $L = 1000$ and $v_m = 4$. The flow-density measurements are determined after ignoring transients that occur within the first $T_0 = 5L$ time steps. Note that each bottleneck site is assumed to be equivalent.

### *2.1. Serial Bottleneck*

In a serial bottleneck, a traffic bottleneck is expanded into $m$ equivalent processing bottlenecks arranged in a series along a single lane (see Fig. 1 for the illustration of a serial bottleneck with 2 processing sites). Vehicles that pass into this expanded unit have to undergo processing by one (and only one) of the $m$ processing bottlenecks. The processing takes a time of $t_d$. The serial expansion is facilitated with a series of gantries to control the movement of vehicles passing by the bottleneck. A car will travel through a site if its gantry is open and will stop in front of a closed gantry. In addition, the gantry system is able to distinguish between processed and unprocessed cars and stops only unprocessed cars.

For efficient processing, an unprocessed car will always be directed to the furthest unoccupied and unblocked site before it is stopped by the gantry system. Specifically, if site 1 to site $k$ are empty while site $k + 1$ is occupied, the car will be directed to the $k$th site and occupy it. Here, site 1 denotes the first processing site which is nearest to the bottleneck entrance while site $m$ denotes the last processing site. A second car which follows the previous car will then be directed to the $(k-1)$th site. Note that sites further down the bottleneck (from $k + 2$ onwards) could be occupied or unoccupied. In this case, any unoccupied processing site further down the bottleneck cannot be utilized as it is blocked. In other words, if the first site is occupied while all the other processing sites are empty, vehicles arriving at the bottleneck will have to wait for the car at the first site to leave before they can enter the bottleneck. Once the car leaves, vehicles

that are waiting to enter the bottleneck are directed to the $m$th, $(m-1)$th, $(m-2)$th sites consecutively. While this delay is a disadvantage of the serial expansion over the parallel expansion, serial expansion can nonetheless be a better choice as it requires no expansion of lane to function. An illustration of vehicle trajectories resulting from the serial bottleneck is presented in the Appendix.

We first examine the relation between traffic flow rate ($J$) and traffic density ($c$) for serial bottlenecks with one, three, and eight gantries. The fundamental diagrams are shown in Fig. 2a, where plateaus of different heights are observed. Essentially, the fundamental diagram is made up of three parts: (i) the free-flow phase on the left; (ii) the fully-congested phase on the right; and (iii) the plateau phase which indicates the maximum allowable flows through the bottlenecks. As shown in Fig. 2a, the slopes for the free-flow and the fully-congested phases are given by $v_m$ and $-1$ respectively for all the three serial bottlenecks.

This qualitative shape of the fundamental diagram is consistent with studies of traffic bottleneck on two-phase traffic models. Under the assumption of a long time limit, the fundamental diagram plateaus in a range of densities which is determined by the flow-density relation of a homogenoues system and the outflow of the bottleneck, such that the net flow through the system remains the same. In other words, the serial expansion has minimal effect on the flow-density relation in either the free-flow or the fully-congested phases. Thus, we shall focus our study on the maximum flow rate at which the fundamental diagram plateaus. Indeed, we observe an increase in the maximum flow rate through the bottleneck as the number of processing bottlenecks in the serial expansion increases.

Next, we study analytically the maximum flow rate for serial bottlenecks with $m$ number of processing sites. Specifically, the maximum flow rate is determined through the time-headway approach (see appendix for details) with the following results:

$$J_{max} = \frac{m}{m + \Delta + t_d}\,, \tag{1}$$

where

$$\Delta = \begin{cases} \left\lceil \frac{\sqrt{1+8m}-1}{2} \right\rceil & m < \frac{v_m(v_m+1)}{2} \\ \left\lceil \frac{2m-v_m(v_m+1)}{2v_m} \right\rceil + v_m & m \geq \frac{v_m(v_m+1)}{2} \end{cases}. \quad (2)$$

Note that the above equation gives $J_{max} = 1/(2 + t_d)$ for $m = 1$. This corresponds to the maximum flow rate for a single bottleneck [1]. On the other hand, we have $J_{max} = v_m/(v_m+1)$ as $m$ approaches infinity. By considering the flow-density relation of a homogenous NaSch model, we determine that the fundamental diagram plateaus at densities that satisfies $J_{max}/v_m < \rho < 1 - J_{max}$, where $J_{max}$ is defined as in Eq. (1). This is consistent with results of our numerical simulations presented in Fig. 2a.

### *2.2. Parallel Bottleneck*

Vehicular processing can also be expanded in parallel by splitting the road into $n$ lanes. In our implementation, the splitting starts at the point $L_{1\to n}(x = 451)$; and it merges into a single lane at the point $L_{n\to 1}(x = 550)$. Here, each split lane is facilitated by one processing bottleneck, with the $n$ bottlenecks forming a parallel bottleneck as a whole. Upon arriving at the point $L_{1\to n}$, a vehicle has a probability of $1/n$ to move into any of the $n$ lanes. The vehicle would then experience a time-delay of $t_d$ in the processing bottleneck before moving off. After being processed, cars at different splitted lanes may arrive at the merging point simultaneously, upon which they move into the main lane with equal probability [26]. Similar to the serial bottleneck, we examine the relation between traffic flow rate and traffic density for parallel bottlenecks with one, three and eight processing sites based on the fundamental diagrams as shown in Fig. 2b. Once again, plateaus of different heights are observed. Interestingly, a parallel expansion does not always increase the maximum flow rate through the bottleneck. For the example shown in Fig. 2b, while the maximum flow rate through a three-laned bottleneck is higher than the maximum flow rate through a single bottleneck, an expansion into an eight-laned bottleneck reduces this enhancement, resulting in a maximum flow rate which is lower than that of the three-laned bottleneck.

For a bottleneck with $n$ parallel processing units, the maximum flow rate is solved analytically as follows:

$$J_{max} = min\left(\frac{n}{3 + t_d - 1/n}, \frac{nv_m}{1 + 2nv_m - v_m}\right). \tag{3}$$

Note that the first term is the combined outflow from the $n$ processing sites and is limited by processing delays. The second term, on the other hand, quantifies the maximum allowable flow when traffic merges back into a single lane at $L_{n\to1}$. Thus, in a parallel bottleneck, the traffic system essentially contains two different bottlenecks. When there is more than one bottleneck, the stronger one dominates [6]. For small value of $n$, the delay of the processing site dominates. At higher values of $n$, flows are limited mainly by the merging of traffic. In Fig. 3, we compare the maximum flow rate through a parallel bottleneck to that of a pure merging lane (no processing site). Indeed, $J_{max}$ is the same for the two cases when $n \geq 3$.

For $n = 1$, we again obtain the expected result of $J_{max} = 1/(2 + t_d)$ from Eq. (3). Meanwhile, with the probability of entering the merged lane being $1/n$, a car in the splitted lane can hardly move into the merged lane if $n$ is large. In this scenario, the average velocity for cars arriving at the merging point $L_{n\to1}$ is approximately zero. Due to finite acceleration of the NaSch model, cars moving out of $L_{n\to1}$ typically have a velocity of $v = 1$. Under the parallel updating procedure of the NaSch model, with a car at $v = 1$ on $L_{n\to1}$ at time $t$, the next car can only reach $L_{n\to1}$ at time $t + 2$. This would give an average flow rate of 0.5 at $L_{n\to1}$. In other words, we expect $J_{max} \to 0.5$ as $n \to \infty$. Moreover, we observe that the range of densities where the fundamental diagram plateaus satisfies $J_{max}/v_m < \rho < 1 - J_{max}$, which is similar to the serial expansion. However, $J_{max}$ is defined according to Eq. (3) in this case.

Note that due to the random interactions between vehicles that occur at the merging bottleneck, there is a difference in the numerical and analytical results for parallel expansion as shown in Fig. 4 (as well as figures related to the parallel bottleneck shown later). This is a consequence of the statistical approximation we have made in deriving the analytical expression (see Appendix

A.2 for details), with our approximation not taking into full account of the intricate vehicular interactions that happen at the merging point.

### *2.3. Comparison between Serial and Parallel Bottleneck*

We first examine the flow rates given by Eqs. (1) and (3). As expected, the flow through a bottleneck is limited mainly by the processing time, and $J_{max}$ is inversely proportional to $t_d$. Expanding the bottleneck introduces a multiplicative effect when two or more vehicles are processed simultaneously. With it, $J_{max}$ becomes proportional to $m$ (or $n$). These relationships illustrate the similarities between the serial and parallel bottlenecks in their response to traffic processing. Nonetheless, there are differences in their mechanisms which causes a difference in outflow. These differences are discussed in this section.

Under serial expansion, cars cannot move concurrently into or out of the processing sites. Processing sites are lined up along a single lane, and each car is required to enter sequentially into it. In particular, two cars that are waiting immediately upstream of the bottleneck do not move simultaneously into the bottleneck after the processing sites are cleared for entry. Instead, the second car moves only after the first car has moved off. Similarly, processed cars exit the bottleneck in tandem with the movement of a car being dependent on whether the car directly in front of it has vacated the site. Underlying this motion is the mechanism that stationary vehicles moves off consecutively due to the parallel updating rule of the NaSch model. This leads to a bottleneck mechanism that constrains the collective motion of the neighboring vehicles, suppressing the flow rate as the number of bottleneck sites increases. This mechanistic feature is captured by the $m + \Delta$ term in the denominator of Eq. (1).

In contrast, lane separation in parallel bottleneck has allowed vehicles to traverse the different processing sites simultaneously. However, the effect of a parallel bottleneck depends on whether it is dominated by time delay processing or lane merging. In the former case, bottleneck effect takes the form $3 + 1/n$; and for the latter case, it is $2n + 1/v_m - 1$ (see Eq. (3)).

A comparison on the processing efficiency of the serial and parallel bottleneck is shown in Fig. 4. At small-scale expansion, the parallel bottleneck functions well and is dominated by time delay processing. Under this circumstance, $m+\Delta > 3+1/n$ which implies that the outflow from parallel bottleneck is greater than that of the serial bottleneck. Hence, parallel bottleneck performs better. On the other hand, as the number of expansion site increases, the bottleneck due to merging dominates over time delay for the parallel bottleneck. At this regime of large-scale expansion, the flow rate under parallel bottleneck possesses an inflection point that arises from a competition between the multiplicative effect and the flow limiting effect of lane merging, whereupon the flow saturates at 0.5 as $n$ becomes large. Conversely, the processing capacity of a serial bottleneck rises steadily as the scale of expansion increases. Beyond a certain number of processing sites, when interaction accrued during serial processing becomes less significant than that from lane-merging: $2n + 1/v_m - 1 > m + \Delta$, the relative effectiveness between serial and parallel bottleneck interchanges. Serial bottleneck now outperforms the parallel expansion in terms of processing efficiency, with a $J_{max}$ that is larger than 0.5.

Finally, it is important to determine the optimum choice of either the serial or parallel bottlenecks as characterized by different parameters for traffic processing purposes. Here, we present the performance between the two expansions for general values of $v_m$ and $t_d$. Specifically, with the same number of processing units, we found that the serial expansion performs better than the parallel expansion for

$$m \geq \frac{2v_m^2 + v_m + 4v_m t_d}{4v_m - 4}\,. \tag{4}$$

Note that the term on the right of Eq. (4) represents a critical point, at which a serial expansion outperforms a parallel expansion in terms of flow rate. Furthermore, it highlights that this critical point is a function of $t_d$.

*2.4. Combination Bottleneck*

To provide further expansion and efficiency, we propose to combine the strength of both the serial and parallel bottlenecks. This creates a combination bottleneck with $n$ lanes, each of which is facilitated by $m$ processing units. In other words, there are a total of $mn$ processing bottlenecks as illustrated in Fig. 1. Figure 5 shows the fundamental diagram of a combination bottleneck with 4 two-gantry lanes and a combination bottleneck with 2 four-gantry lanes. Interestingly, the two configurations perform differently under the same conditions.

The maximum flow rate for a combination bottleneck with $n$ number of $m$-gantry lanes is expressed as follows:

$$J_{max} = min\left(\frac{n^2 m}{(2m + \Delta + t_d)n - m}, \frac{n v_m}{1 + 2n v_m - v_m}\right), \tag{5}$$

with $\Delta$ being defined in Eq. (2). The derivation of Eq. (5) is given in the Appendix.

In Figs. 6a and 6b, maximum traffic flow rate are shown for combination bottlenecks with 1, 2, 3, and 4 lanes. In general, the flow rate increases with an increase in the number of gantries, before it reaches the limit imposed by lane merging which is indicated by the appearance of a horizontal section of the curve. Note that an increase in the number of $m$-gantry lanes does not always leads to an increase in $J_{max}$.

Interestingly, at the regime of large-scale expansion, it is possible to optimize the outflow by combining the strengths of the serial and parallel bottleneck. In other words, there is an optimum configuration for a fixed number of processing units that gives the most efficient traffic flow. Such a combination bottleneck has a collective vehicular effect on the maximal flow ascribed by $2u/n - m/n + \Delta$, with $u = mn$ being the total number of sites (see Eq. (5)). It is significant that this collective effect of the combination bottleneck can be made smaller than that of the parallel or serial bottleneck with the same $u$, thus achieving a more optimal traffic processing efficiency than either of them. Let us illustrate

this fact by considering bottlenecks with 12 processing units which are organized among: (1) 12 one-gantry lanes; (2) 6 two-gantry lanes; (3) 4 three-gantry lanes; (4) 3 four-gantry lanes; (5) 2 six-gantry lanes; and (6) 1 twelve-gantry lane (see Fig. 7a). Depending on the processing time $t_d$, one of these configurations outperforms the others. While configuration (5) is a better choice when $t_d$ is small, configuration (2) performs better than the other configurations when $t_d$ is large. In particular, a more serial processing gives better performance for a smaller $t_d$. This observation is clearly borne out in Fig. 7a. As $t_d$ reduces, we notice that the successive optimal configuration goes from configuration (2) to configuration (3), then to configuration (4), to configuration (5), and finally to configuration (6). Similar results are shown in Fig. 7b for bottlenecks with 30 processing units. A common mechanistic feature that can be gleaned from Figs. 7a and 7b is that $J_{max}$ is rate-limited by lane merging for small $t_d$, where $J_{max}$ displays a constant value above 0.5. Beyond a critical $t_d$ whose value depends on the configuration of the combination bottleneck, the delay bottlenecks due to the processing sites start to dominate, and we observe a monotonic decrease in $J_{max}$ as $t_d$ increases.

### *2.5. Relation to current real-world traffic*

Thus far, our study had been based on an elementary traffic model. While the simplicity of the employed model has enabled an analytical study and the discernment of the underlying physics, it lacks the mechanisms that capture key features of empirical traffic flow. Examples of missed features are: (1) metastable high-flow states; (2) reduced outflow from jams; and (3) wide-scattering of flow-density states [2]. To put this study into a more realistic context, we have considered an additional traffic model: the Kerner-Klenov-Wolf (KKW) model [14], which has incorporated mechanisms that generate the above features of empirical traffic flow. Specifically, the slow-to-start (STS) [21] and velocity-dependent randomization (VDR) [22] rules in the model reproduces the fluctuation of deceleration patterns according to the current vehicular velocity. Based on the empirically-based three-phase theory [23], wide-scattering

is a consequence of the existence of a "synchronized gap" in which cars that are traveling with the same velocity on a single lane can take on an arbitrary space gap between each other. The KKW model thus includes mechanisms related to human driving behavior which gives rise to the synchronized gap, the VDR and the STS rules, to produce a model that is comparable to real traffic.

Figure 8a presents the equivalent of Fig. 4 using the KKW model. Note that bottleneck capacity in empirically-based traffic models spans a certain range [24], which we illustrate here in terms of the mean and standard error. While the relationship between $J_{max}$ and $m$ is qualitatively similar to those of the NaSch model, the resultant bottleneck capacities shown in Fig. 8a are much lower. This is a consequence of the incorporation of additional mechanisms which have further limited the flow of vehicle through the bottleneck. We deduce that the STS rule is the largest contributor to this reduction in flow rate among the mechanisms introduced into the KKW model. Stop-and-start motion is very common in processing bottlenecks and when it is coupled with the STS rule, becomes the main mechanism in enhancing the stochastic deceleration as the vehicle moves off from rest. It models and attributes the larger delay in moving off a vehicle to that of human reaction time [21]. In consequence, $J_{max}$ for the parallel bottleneck has become more heavily limited, changing from 0.5 to 0.34.

From these empirically-based models, parallel expansion is found to consistently outperform the serial expansion in terms of processing efficiency. More importantly, the increased human reaction time captured by the STS rule have enlarged the performance differences between the serial and parallel expansion. This can be understood from the fact that serial bottleneck tends to cause cars to stop and wait for the preceding vehicle to complete processing before moving off. Since the STS rule affects cars that are stopped, there is a higher frequency that this rule is applied in the serial bottleneck than the parallel bottleneck. As a result, the cumulative delay due to slow human response in starting off the vehicle from rest accounts for the relative inefficiency of the serial processing bottleneck, justifying the expenditure of lane-expansion cost for the implementation of parallel bottleneck. It also explains the observed abundance of parallel

bottleneck, and a scarcity of serial bottleneck, in current traffic systems.

## 3. Discussion and Conclusion

In order to improve the efficiency of the processing bottleneck, our analysis has revealed the need to optimize the collective motions of the vehicle. In particular, human reaction time and characteristics in mobilizing a vehicle, which are modeled by the STS rule in the last section, have caused a restriction in outflow which drastically decreases the efficiency of all processing bottlenecks as depicted in Figs. 8. By removing the human factor in the bottleneck system, as exemplified by the simplified NaSch model, we observe the feasibility of achieving a processing efficiency that goes beyond current technology and capabilities. Thus, our results point to the potential benefits of autonomous or self-driving vehicles in enhancing traffic capacity through optimizing the collective motions of vehicle by putting human drivers out of the loop. This is particular pertinent in the current era of urbanization, where more than half of the world population is living in cities. It was projected that by 2050, 66% of the world population will reside in urban areas [25]. With the increasing human density comes greater demand on our road systems. In order for the cities to be liveable, there is a need to enhance the capacity of current road networks and to improve the efficiency of traffic systems in handling the heavier vehicular flow. Ultimately, the purpose of the various technological solutions is to prevent the occurrence of severe traffic congestion.

In this respect, our proposed combination processing bottlenecks is directly applicable in addressing these traffic woes of the future. We have seen that for a sufficiently large number of processing units (see Eq. (4)), the new configuration of serial bottleneck becomes tenable for traffic processing. While this is true for small processing time, our results have demonstrated that as the processing time increases, a set of combination bottlenecks with progressively more lanes and less serial processing (assuming a fixed total number of processing units) is optimal for traffic processing. In fact, for large processing time, a multi-lane

bottleneck with at least two serial processing site is found to be optimal. These results show that in the processing of dense traffic flow, serial bottleneck is an essential component that can serve to produce greater efficiency in the system as long as delay due to human reaction time (as illustrated by the STS rule) has been amply minimized.

In summary, we have studied the efficiencies of processing bottlenecks expanded in different configurations. For small-scale expansion, parallel bottleneck is generally more efficient in increasing roadway capacity. On the other hand, serial expansion is a better choice when the number of processing units grows beyond a critical point. More significantly, by incorporating serial bottleneck into parallel bottleneck during large-scale expansion, we uncover an even more optimal alternative solution where a combination of these bottlenecks is able to process and manage very efficiently the heavy traffic flow of urban cities of the future.

**Acknowledgments**

This work was partially supported by MOE AcRF Tier 1 Grant No. RG93/15.

## Appendix A. The time-headway approach

We adopt the time-headway approach to determine the maximum flow rate through a bottleneck. The underlying assumption of this approach is that of a two-phase traffic, where at the critical point between the free flow and the state where the flow maximizes, vehicles flow in an organized way on the verge of interacting. At this point, the spatial-temporal profile of the vehicles assumes an average pattern in the steady state, allowing us to extract a mean period $\bar{h}_{tc}$ from the pattern. The maximum flow rate $J_{max}$ is then determined as $J_{max} = 1/\bar{h}_{tc}$.

### *Appendix A.1. Serial bottleneck*

For serial expansion, the $m$ vehicles are observed to move in a periodic configuration under critical density due to the deterministic nature of the flow

(see Fig. A.9). In this configuration, it is easy to see that the total time-headway is given by $t_d + m + \Delta$. This implies an average time-headway of

$$\bar{h}^s_{tc} = \frac{m + \Delta + t_d}{m}\,, \tag{A.1}$$

between each vehicle. The maximum flow rate $J_{max}$ for the serial expansion is then given by $J_{max} = 1/\bar{h}^s_{tc}$, which is Eq. (1).

From Fig. A.9, $\Delta$ is observed to be the duration requires for the vehicle delayed at the left-most site to leave the $m$ sites. Since a delayed vehicle starts from rest and accelerates at 1 site/$s^2$, the distance covered by this vehicle would increase systematically at each time step. In fact, the successive distance covered follows an arithmetic progression which indicates a total distance traversed of $d = (t/2)(t+1)$. By letting $d = m$ and $t = \Delta$, we solve for $\Delta = \left\lceil \left(\sqrt{1+8m} - 1\right)/2 \right\rceil$, where the ceiling function $\lceil \cdots \rceil$ is a consequence of the discretized time in the NaSch model and that this vehicle must travel at least $m$ sites. Because vehicles in the NaSch Model accelerates to a finite velocity, this equation is applicable only for vehicle velocity $v \leq v_m$. In terms of $m$, this condition translates to $m \leq v_m(v_m+1)/2$.

When $m > v_m(v_m+1)/2$, the vehicle reaches $v_m$ and maintains at this velocity in its subsequent motion. As a result, once past the first $v_m(v_m+1)/2$ sites (which takes $v_m$ time steps to traverse), the vehicle covers the remaining sites in $\left\lceil [2m - v_m(v_m+1)] \,/\, [2v_m] \right\rceil$ time steps. Combining these results on $\Delta$, Eq. (2) is obtained.

*Appendix A.2. Parallel bottleneck*

As mention in section 2.2, there are two regimes of maximum flow rate under parallel expansion, of which their time-headway have to be determined separately. Moreover, unlike serial expansion, there exists stochastic behavior in the parallel expansion. Nonetheless, by considering the aggregate behavior of a large number of vehicles, we can ascertain its average behavior from which we obtain the mean time-headway for the vehicular flow.

The multiplicative effects of parallel expansion have already been explored in [26], where the overall flow rate was estimated to be the combined flow rate of each of the lanes. The analytical results of [6] state that the maximum flow rate at the processing site is $1/(2+t_d)$, which is equivalent to a time-headway of $(2+t_d)$. The total flow rate of the $n$ parallel lanes is then given approximately as $n/(2+t_d)$. However, this evaluation assumes no interaction between the vehicles at the point of merging which is a gross over-estimation.

In the following, we improve this approximation by including the effect of interaction at the merged flow in a statistical way. The consequence is a flow rate that is less than the sum of the flow rate of each lane. When vehicles at different lanes approach the merging point, only one vehicle passes through unobstructed per unit time, while the rest of the $n-1$ vehicles have to wait one extra time step relative to the unobstructed vehicle to get through. This gives a probability of $1/n$ for a time-headway of $(2+t_d)$ for outflow without obstruction; and a probability of $1-1/n$ for time-headway of $(3+t_d)$ for outflow with vehicular interaction. Thus, the average total time-headway is given by $3+t_d-1/n$, which leads to an average time-headway for parallel bottleneck due to multiplicative effect to be

$$\bar{h}_{tc}^{pm} = \frac{3+t_d-\frac{1}{n}}{n}\,. \tag{A.2}$$

In the regime where the traffic flow is limited by lane merging, the effect of the processing bottleneck is negligible. This is analogous to ignoring the presence of processing site in each lane. According to Ref. [10], vehicles moving with no bottleneck has a maximum flow rate of $v_m/\left(v_m+1\right)$. Thus, we expect the outflow without obstruction to have a time-headway of $\left(v_m+1\right)/v_m$ for one vehicle. For the rest of the $n-1$ vehicles, vehicular interactions cause them to stop at the merging point. They later move off with a velocity of $v=1$ due to unit acceleration of the NaSch model, which gives each a time-headway of 2 after considering an extra time step as before. Putting these $n$ time-headway together lead to an average total time-headway of $\left(v_m+1\right)/v_m+2(n-1)$. The average

time-headway in consequence of flow limitation by lane-merging is therefore

$$\bar{h}_{tc}^{pf} = \frac{1 + 2nv_m - v_m}{nv_m} \,. \tag{A.3}$$

Finally, the maximum flow rate through the parallel bottleneck depends on the dominance of the processing bottleneck or the lane-merging bottleneck. The dominance of the former leads to $J_{max} = 1/\bar{h}_{tc}^{pm}$, while the latter $J_{max} = 1/\bar{h}_{tc}^{pf}$. These results together then gives Eq. (3) of the parallel bottleneck. Note that a space-time trajectories of vehicles moving through a parallel bottleneck is illustrated in Fig. A.10.

*Appendix A.3. Combination bottleneck*

The derivation of the maximum flow rate of the combination bottleneck employs both the arguments of the serial bottleneck and the parallel bottleneck. Similar to the parallel bottleneck, the combination bottleneck possesses two regimes: one dominated by the processing bottleneck, while the other by lane-merging bottleneck.

For the former case, we again consider the outflow without obstruction. This gives a time-headway of $\bar{h}_{tc}^{s}$ resulting from the serial bottleneck, and happens with a probability of $1/n$. On the other hand, we have a time-headway of $\bar{h}_{tc}^{s}+1$ which occurs with a probability of $1 - 1/n$ for the rest of the outflow. Note an extra time unit has been added to account for the effect of interactions. The average time-headway of the combination bottleneck for multiplicative effect is then

$$\bar{h}_{tc}^{cm} = \frac{(2m + \Delta + t_d)n - m}{n^2 m} \,. \tag{A.4}$$

In the case of the lane-merging bottleneck, we again ignore the processing sites. In consequence, the result here is exactly the same as that for the parallel bottleneck above, i.e. $\bar{h}_{tc}^{cf} = \bar{h}_{tc}^{pf}$, when lane-merging dominates.

The maximum flow rate through the combination bottleneck is then given by $J_{max} = 1/\bar{h}_{tc}^{cm}$ when processing bottleneck dominates, and $J_{max} = 1/\bar{h}_{tc}^{cf}$ when lane-merging bottleneck dominates. This proves Eq. (5).

## References


[1] K. Nagel and M. Schreckenberg, J. Phys. I France **2**, 2221 (1992).

[2] A. S. Debashish Chowdhury, Ludger Santen and A. Schadschneider, Physics Report **329** , 199 (2000).

[3] M. L. L. Iannini and Ronald Dickman, Phys. Rev. E **95**, 022106 (2017)

[4] Henrik M. Bette, M. Schreckenberg et. al, Phys. Rev. E **95** 012311 (2017)

[5] Y. Sugiyama et. al., New Journal of Physics 10 (2008) 033001

[6] K.Chung and P. Hui, Journal of the Physical Society of Japan **63**, 4338 (1994)

[7] Schadschneider et. al. Physica A **285** 101 (2000)

[8] D. W. Huang and W. N. Huang, Physica A: Statistical Mechanics and its Applications **312**, 597 (2002).

[9] R. Jiang, B. Jia, and Q.S. Wu, International Journal of Modern Physics C, **15**(05), 619628, (2004).

[10] K. Nagel and H.Herrmann, Physica A **199**, 254 (1993)

[11] A. Schadschneider, M.Schreckenberg Journal of Physics A: Mathematical and Theoretical, **26**(15) 7 (1993)

[12] Schadschneider, M. Schreckenberg Journal of Physics A: Mathematical and General, **35** (5), 13211322 (2002)

[13] A. Schadschneider, European Physical Journal B, **10**(3), 573582 (1999)

[14] B. Kerner et. al J. Phys. A: Math Gen. **35** 9971-10013 (2002)

[15] B. S. Kerner, S. Klenov, and M. Schreckenberg, Phys. Rev.E **84**, 046110 (2011).

[16] JF Tian et. al. Trans. Res. B **71** 138-157 (2015)

[17] B. S. Kerner, Journal of Physics A: Mathematical and Theoretical **41**, 215101 (2008).

[18] C.F. Daganzo and M.J. Cassidy, Transportation Research Part B: Methodological, **42**(10), 861872 (2008).

[19] K. Komada and T. Nagatani, Physica A: Statistical Mechanics and its Applications, **389**(11), 22682279 (2010).

[20] Y. S. Qian et. al. , Physica A: Statistical Mechanics and its Applications **392** 5874 (2013).

[21] S.C. Benjamin et. al. J. Phys. A **29**, 3119 (1996)

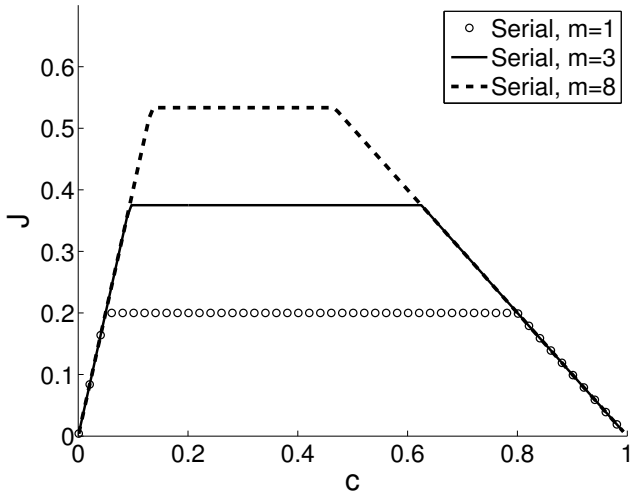
[22] R. Barlovic et. al. Eur. Phys. J. B **5** 793 (1998)

[23] B. Kerner *Introduction to Modern Traffic Flow Theory and Control: The long road to Three-phase Traffic Theory* (Springer, Berlin)

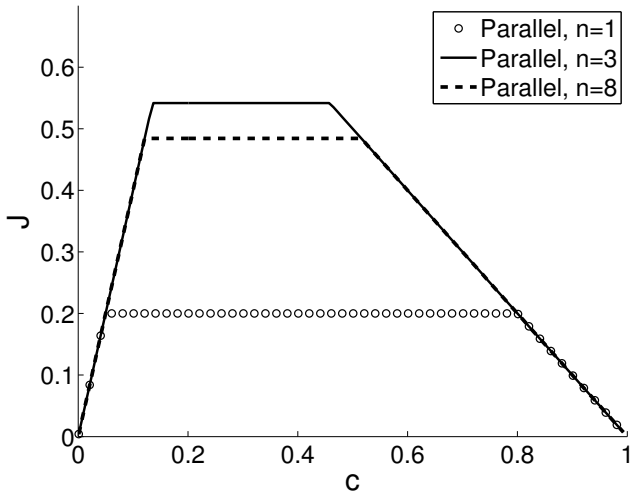
[24] B. Kerner Physica A **392** 5261–5282 (2013)

[25] Philip Ball, 2007 *Why Society is a Complex Matter* (Springer, Berlin).

[26] S. Xiao, L. Kong, and M. Liu, Acta Mechanica Sinica **21**, 305 (2005).




(a) Serial bottlenecks




(b) Parallel bottlenecks

Figure 2: Fundamental diagrams of the (a) serial and (b) parallel bottleneck configuration for $t_d = 3$. The results presented are for the mean flow rate $J$ at each traffic density $c$. Note that the serial bottleneck has consistently higher bottleneck capacity as $m$ increases. However, the bottleneck capacity for parallel bottleneck increases when $n$ goes from 1 to 3, but reduces as $n$ becomes 8. Note that for both serial and parallel bottleneck, the free flow phase and congestion phase exhibit the expected slope of $v_m$ and $-1$ respectively, in the fundamental diagram.

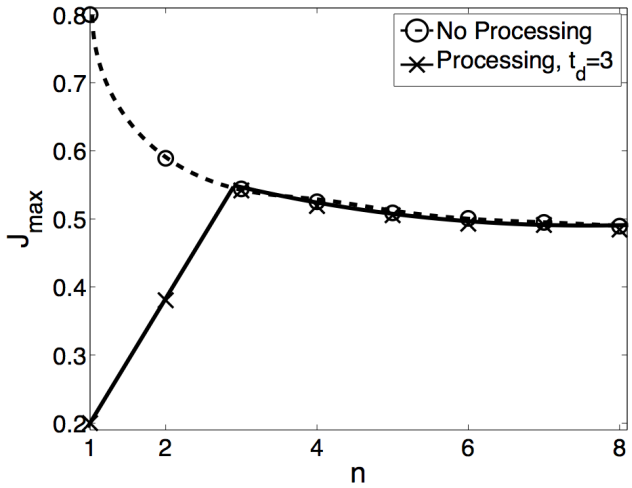



Figure 3: A comparison between maximum flow rate for parallel expansion without processing site and with processing sites for $t_d = 3$. While analytical plots are presented as a continuous line, only the results at integer values are relevant.

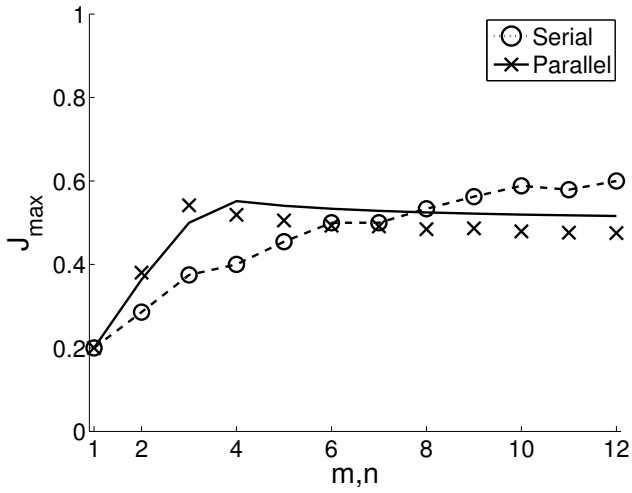
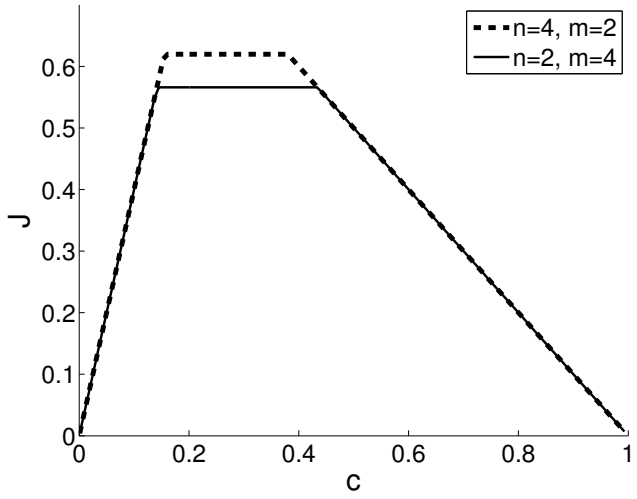
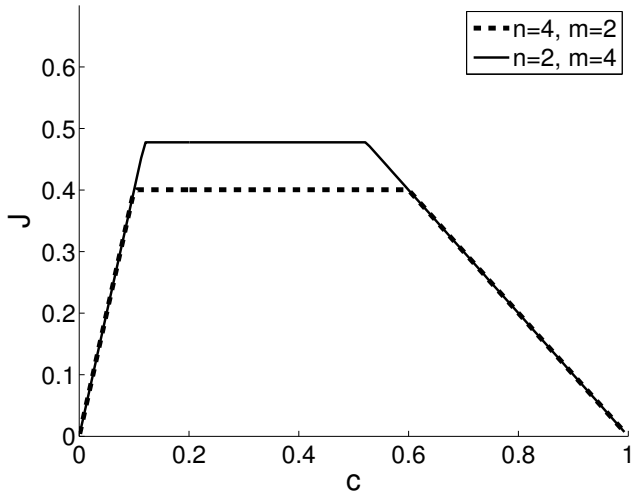



Figure 4: A plot of $J_{max}$ versus parameters $m$ and $n$ for serial and parallel expansion respectively for $v_m = 4$ and $t_d = 3$. Serial bottleneck is observed to perform better than parallel bottleneck when the number of processing sites is greater than 7. Note that the data points are the mean numerical results of 100 realizations, while the lines are from the analytical prediction of Eqs. (1) and (3). Error bars of numerical results are smaller than the marker size. While analytical plots are presented as a continuous line, only the results at integer values are valid. Note that this condition is applicable to all analytical plots in the paper.

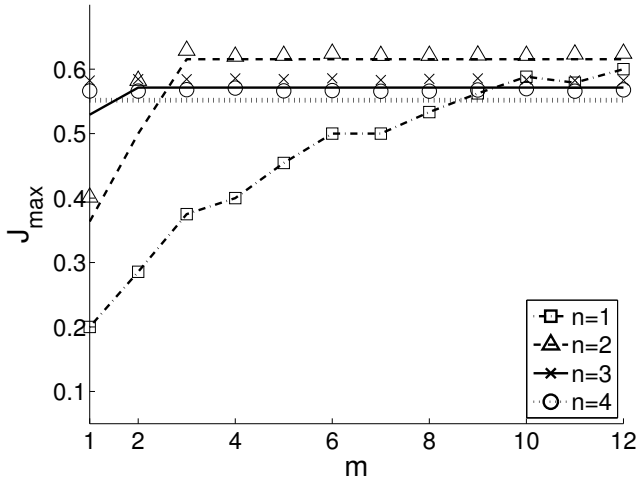



(a) $t_d = 3$

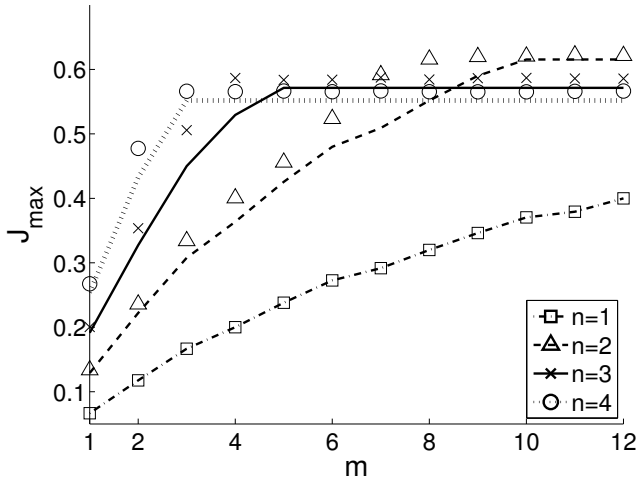



(b) $t_d = 13$

Figure 5: Fundamental diagrams of combination bottleneck for $t_d = 3$ and $t_d = 13$ for two different configurations with a total of 8 processing sites. The results are displayed for mean flow rate $J$ against traffic density $c$ based on an ensemble of 100 realizations.

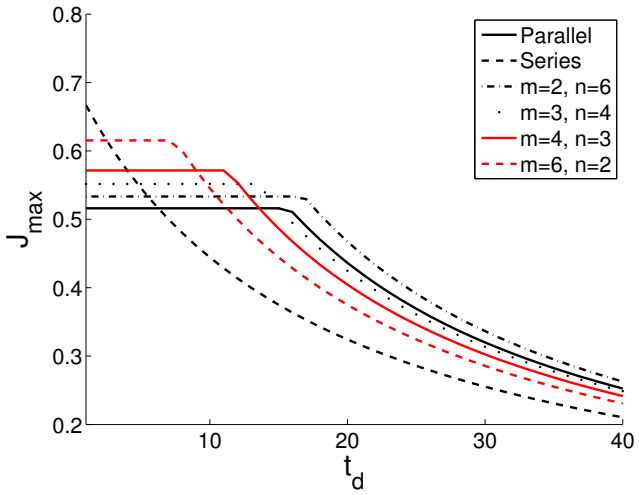



(a) $t_d = 3$

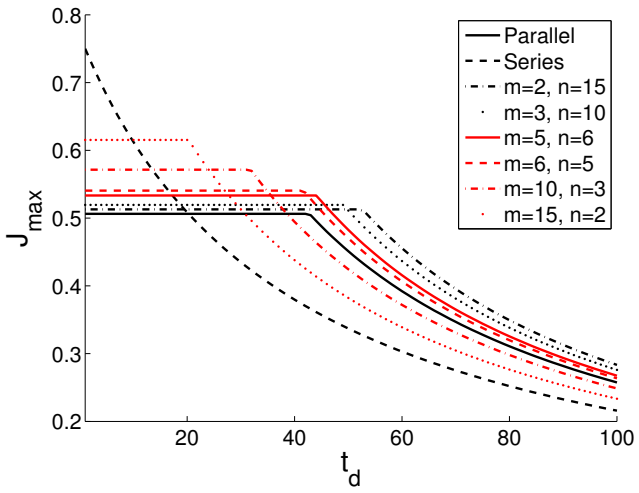



(b) $t_d = 13$

Figure 6: A plot of $J_{max}$ versus $m$ for combination bottleneck. Note that the data points are from numerical results and represent the mean of 100 realizations, while lines are plots from the analytical prediction of Eq. (5). Error bars of numerical results are smaller than the marker size. Only the integer values of the analytical plots, which are presented as continuous line, are relevant.

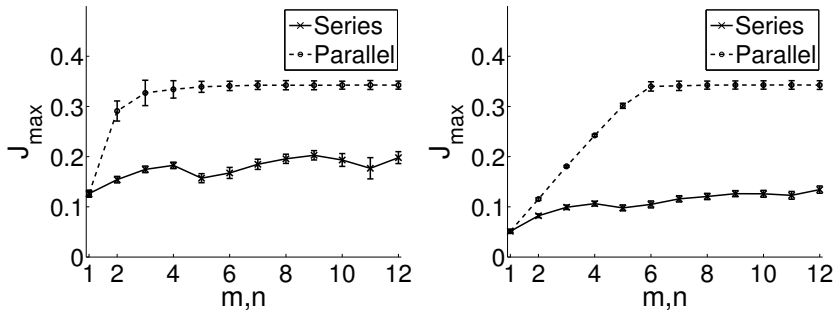



(a) 12 time-delay sites

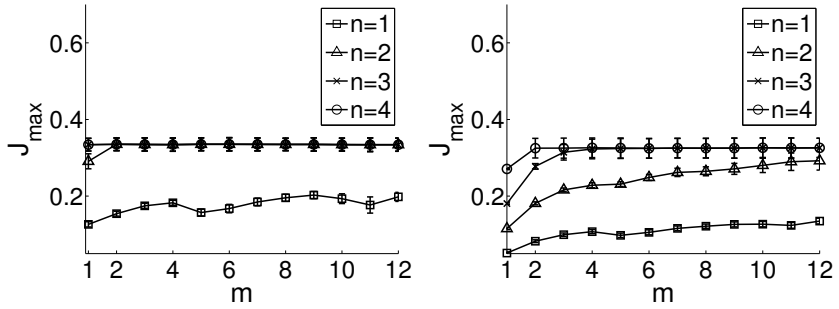



(b) 30 time-delay sites

Figure 7: A plot of $J_{max}$ versus $t_d$ for a combination expansion of 12 and 30 processing sites in various combinations of $m$ and $n$ values.




(a) Serial vs Parallel bottleneck of $t_d = 3$ (b) Serial vs Parallel bottleneck of $t_d = 13$




(c) Combination Bottleneck of $t_d = 3$ (d) Combination Bottleneck of $t_d = 13$

Figure 8: A plot of $J_{max}$ versus $m$ and $n$ for serial, parallel, and combination expansion of bottlenecks for the KKW-1 model. The parameters for the simulation are: $p = 0.04$, $p_0 = 0.425$, $p_{a1} = 0.2$, $p_{a2} = 0.052$, $k = 2.55$, $v_p = 3$, $t_d = 3$ and 13. Note that computations were made only at the data points with the lines between the points serve the sole purpose of illustration.

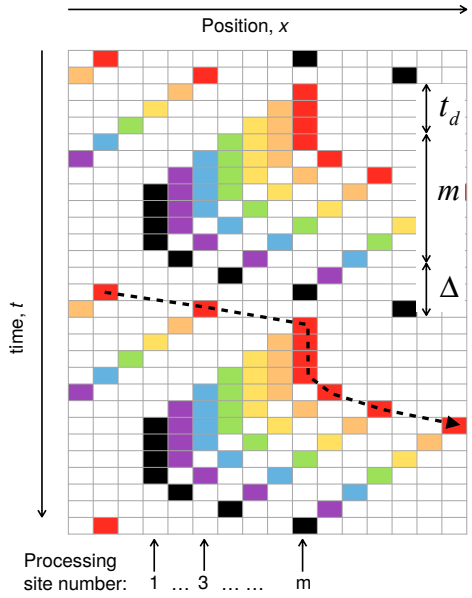



Figure A.9: A plot of the space-time trajectories of vehicles moving through a serial bottleneck at critical density. The parameters of the simulation are: $m = 7$, $t_d = 3$, and $v_m = 4$. Note that the different colors are used to differentiate each vehicle, and the repeated colors do not represent the same vehicle. The dotted line indicates the trajectory of one of these vehicles.

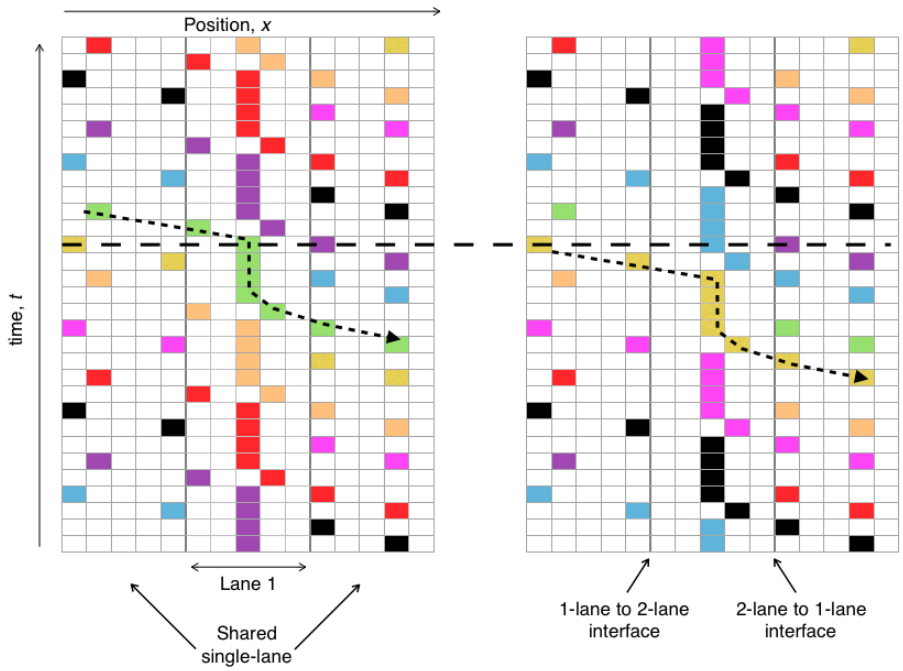



Figure A.10: A plot of the space-time trajectories of vehicles moving through a parallel bottleneck. Note that the first 5 and the last 5 cells of each row in each subfigures are common between the two subfigures, as they illustrate the single-laned road. The middle 5 cells show the 2 expanded lanes with individual processing bottlenecks. The parameters of the simulation are: $n = 2$, $t_d = 3$, and $v_m = 4$. We use different colors to differentiate between vehicles, while the dotted line indicates the trajectory of one of these vehicles. The dashed line serves to highlight the difference between the two lanes at a single point in time.