Queuing model of a traffic bottleneck with bimodal arrival rate

MARKO WOELKI

Institute of Transportation Systems, German Aerospace Center, Rutherfordstraße 2
12489 Berlin, Germany
marko.woelki@dlr.de

This paper revisits the problem of tuning the density in a traffic bottleneck by reduction of the arrival rate when the queue length exceeds a certain threshold, studied recently for variants of TASEP and Burgers equation. In the present approach a simple finite queuing system is considered and its contrasting 'phase diagram' is derived. One observes one jammed region, one low-density region, and one where the queue length is equilibrated around the threshold. Despite the simplicity of the model the physics is in accordance with the previous approach: The density is tuned at the threshold if the exit rate lies in between the two arrival rates.

Keywords: queuing theory; traffic flow; TASEP.

1. Introduction

Queuing models have a long tradition as application to traffic flow problems [1], [2], [3]. In those models one typically tries to explain the formation and dynamics of a traffic jam. The most simple queuing model is the so-called $M/M/1$ queue [4]. In this model, customers arrive at the queue at a certain constant rate α and leave from the queue at a different constant service rate β. Obviously the stationary distribution takes a very simple form. The process has recently been revisited in the traffic context: Instead of modeling the dynamics of a compact jam, the $M/M/1$ queue serves in [5] as a model for a whole bottleneck in which the gaps between cars as well as the length of the bottleneck are neglected, only the number of cars (customers) is regarded. This was contrasted with the totally asymmetric exclusion process (TASEP) in which the cars move unidirectionally through a row of cells each of which is occupied by at most one car. For an extensive overview on TASEP and related processes, see [6], [7] and references therein. It was shown in [5] that the typical high- and low-density phases of the totally asymmetric exclusion process (TASEP) have an analogy in the $M/M/1$ queue.

Recently, for a generalization of TASEP it has been investigated under which circumstances the flow through a bottleneck in stochastic models of traffic flow can be optimized [8]. The strategy was to reduce (enhance) the arrival rate α, whenever

the car density is above (below) a certain threshold while the exit rate β is kept constant. This regulation, referred to in [8] as density-feedback control, serves as a traffic-management strategy that tries to prevent for high volume of traffic. One practical realization would be to demand a certain toll or to suggest different routings via the navigation system while the threshold density is exceeded. In this way the arrival rate could be reduced. The simplest idea to model the situation is to choose a bimodal arrival rate that switches between two constant values (α_- if the density ρ fulfills $\rho < \rho^*$ and α_+ otherwise). The threshold ρ^* is typically the density where the flow becomes maximal. In [8] the problem was studied on the example of the totally asymmetric exclusion process (TASEP) by solution of its mean-field equations which in turn are a discretization of the noisy Burgers equation with Diffusion constant 1/2, namely $\partial \rho_i / \partial t = \rho_{i-1}(1 - \rho_i) - \rho_i(1 - \rho_{i+1})$. The equation describes the time evolution of the density at site i ($i = 1, 2, \ldots L$) for a system with length L with right-hand reservoir-density $1 - \beta$ and modified left-hand density α. The diagram in Figure 1 summarizes most conveniently the corresponding results
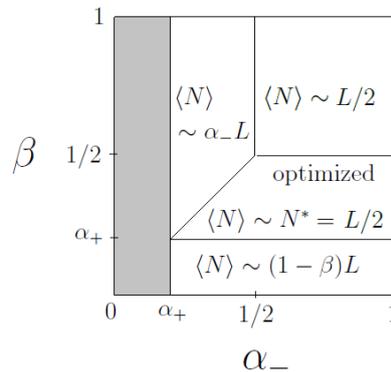
Fig. 1.   Phase diagram for a threshold density $\rho^* = 1/2$ resulting from mean-field theory for the TASEP. Shown is the average number of cars $\langle N \rangle$ in certain ranges of parameter space.

from [8] for $\rho^* = 1/2$. The region where $\langle N \rangle \sim \alpha_- L$ is a low-density phase and where α_- and β are larger than 1/2 is a maximum-current phase (there the flow takes its maximal value 1/4). Without regulation of the arrival rate the remaining two phases would be one high density phase. However due to the regulation mechanism a part of the high-density phase is transformed into a so-called coexistence phase where a shock moves through the system that separates a lattice part on the left at low-density from a high-density on the right. Here one has $\langle N \rangle \sim N^*$, thus the average car density is equilibrated at the threshold and the control mechanism leads to a maximization of the flow. Note that the grey shaded area is physically unreasonable in the present context but has applications in Langmuir kinetics [9].

Here we choose a simpler approach, namely a queuing model, in order to see whether the parameter range in which one finds the equilibration of the system at

the threshold density can be reproduced. We consider a bottleneck where the arrival rate α_N depends on the actual queue length $N = 0, 1, \ldots, L$ while the exit rate β is constant.

The queue-length dependence of the arrival rate can be interpreted as a sort of capacity constraint that has been studied in various socio-economic and biological applications [?], [10], [11]. However the special realization and the viewpoint of regulation in traffic management are completely different from previous investigations and results give a closer understanding of what happens physically to a traffic system under feedback control. The model is also related to the problem of two queues with finite capacity, for example two checkouts which together can hold L customers at most. For an overview of related problems, see [12].

The paper is organized as follows: First, we define the model and present its stationary distribution. We derive the average queue length and the flow and identify the phase diagram. Those results will finally be compared to recent investigations [8].

2. Model definition

We study a queue with length N, maximum length L (the length of the street) and a threshold N^*. The process is further defined by its transition rates: a constant output rate β and queue-length dependent arrival rate α_N. We consider continuous time, that is realized in a computer simulation as a random-sequential update by first flipping a coin to decide whether it is tried to insert or to remove a particle. Afterwards one draws a random number and decides according to the rate whether the move is executed or not.

The time evolution of the probability $P_{N,L}(t)$ to find a queue length N at time t is governed by

$$\dot{P}_{0,L}(t) = -\alpha_0 P_{0,L}(t) + \beta P_{1,L}(t), \tag{1}$$

$$\dot{P}_{N,L}(t) = -(\alpha_N + \beta)P_{N,L}(t) + \alpha_{N-1}P_{N-1,L}(t) \tag{2}$$

$$+ \beta P_{N+1,L}(t), \text{ for } 0 < N < L, \tag{3}$$

$$\dot{P}_{L,L}(t) = -\beta P_{L,L}(t) + \alpha_{L-1}P_{L-1,L}(t). \tag{4}$$

Note that the α_N and β are rates rather than probabilities. To obtain probabilities they have to be multiplied by a sufficiently small time interval Δt. However from now on we assume without loss of generality that they are already normalized so that all rates are smaller or equal than 1. Now we are going to analyze one choice of the arrival rate in more detail:

$$\alpha_N = \begin{cases} \alpha_-, \text{ for } 0 \leq N < N^* \\ \alpha_+, \text{ for } N^* \leq N < L \end{cases} \tag{5}$$

and we restrict ourselves to $\alpha_+ < \alpha_-$ (The case $\alpha_+ > \alpha_-$ makes no sense in the present context). Hence the probability that a car arrives at the queue takes a

smaller (or larger) value if the actual queue length N is above (or below) a threshold N^*.

3. The stationary distribution

Since (1-4) are the evolution equations of a disordered random walk, the exact solution in the long time limit is well-known [13],[14]. One finds with (5) and the more convenient variables

$$x_- = \alpha_-/\beta \text{ and } x_+ = \alpha_+/\beta \tag{6}$$

the probability distribution

$$P_{N,L} = \begin{cases} Z_L^{-1} x_-^N, & \text{for } N < N^*, \\ Z_L^{-1} x_-^{N^*} x_+^{N-N^*}, & \text{for } N \geq N^*, \end{cases} \tag{7}$$

with the normalization

$$Z_L = \frac{(1 - x_+)(1 - x_-^{N^*}) + (1 - x_-)x_-^{N^*}(1 - x_+^{L-N^*+1})}{(1 - x_-)(1 - x_+)} \tag{8}$$

where the latter follows from probability conservation $1 = \sum_{N=0}^L P_{N,L}$. Note that for a better readability, we skip the subscript N^* in all the quantities. Eq. (7) says that if there are N cars in the queue then each car below the threshold N^* contributes a factor x_- while other cars contribute a factor x_+ to the stationary distribution. It is convenient to rescale the queue length and its distribution to obtain with the help of the scaling variable $z = N/N^*$ an expression independent of N^* and L:

$$F(z) := (Z_L \cdot P_{N,L})^{z/N} = \begin{cases} x_-^z, & \text{for } z < 1 \\ x_- x_+^{z-1}, & \text{for } z \geq 1. \end{cases} \tag{9}$$

The results can nicely be distinguished if we define a large (low) rate as larger (smaller) than $1/2$ and reference to the change in arrival rates from α_- to α_+. Then we obtain the following three cases:

Case 1: Arrival rate changed from large to small ($\alpha_+ < 1/2 < \alpha_-$)
For $\alpha_+ < 1/2 < \alpha_-$ one has $1 - \alpha_- < 1/2$ and therefore $\alpha_+ < \beta \leq 1 - \alpha_-$ is accessible for β. In this region one has $x_+ < 1 \leq x_-$. Therefore queues with $N < N^*$ tend to increase (positive drift) and queues with $N \geq N^*$ tend to decrease (negative drift). This case is illustrated in Figure 2 (left). The green curve marks the border to the unphysical region. Here one has $x_+ = 1$ and the rescaled distribution takes the value α_-/α_+ for $N \geq N^*$.

Case 2: Small arrival rate reduced further ($\alpha_+ < \alpha_- < 1/2$)
For the choice $\alpha_+ < \alpha_- < 1/2$ one has for $\alpha_+ < \beta < \alpha_-$ a positive drift ($x_- > 1$), i.e. queues smaller than N^* tend to increase and for $\alpha_- < \beta < 1 - \alpha_-$ a negative drift ($x_- < 1$), thus queues smaller than N^* tend to shrink. Note that queues larger
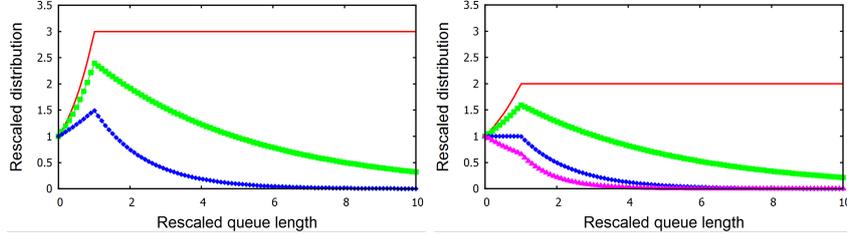
Fig. 2. x-axis: the rescaled queue length $z = N/N^*$. y-axis: the rescaled queue-length distribution $F(z)$. Case 1 left ($\alpha_- = 0.6$) and case 2 right ($\alpha_- = 0.4$), both for $\alpha_+ = 0.2$. The curves correspond from top to bottom to $\beta = 0.2$, $\beta = 0.25$, $\beta = 0.4$ and on the right additionally $\beta = 0.6$.
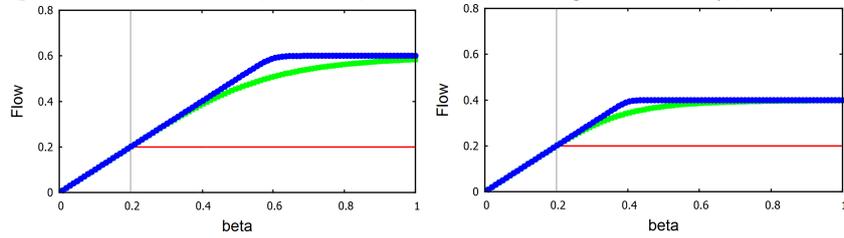
Fig. 3.   Flow J versus β for $\alpha_+ = 0.2$ with $\alpha_- = 0.6$ (left, case 1) and $\alpha_- = 0.4$ (right, case 2). From bottom to top the curves correspond to different threshold densities: Straight (red) line: $\rho^* = 0$, squares (green): $\rho^* = 1/20$, diamonds (blue): $\rho^* = 1/2$. The maximal queue length is $L = 100$. Note that the rounding of the transition at $\beta = \alpha_-$ is a finite site effect that vanishes if one increases L.

than N^* also tend to shrink ($x_+ < 1$). See Figure 2 (right).

Case 3: Reduction to a still large arrival rate ($1/2 < \alpha_+ < \alpha_-$)

In this case one has $\beta < 1 - \alpha_- < 1/2 < \alpha_-$ and therefore the queue length is of order L (for finite L or unbounded otherwise) for all β in this range.

4. Flow and average queue length

In this section we calculate the average queue length $\langle N \rangle$ and the flow which will lead in the next section to an interesting phase diagram. This is compared to a corresponding diagram for a recent approach to the same problem [8]. In Figure 3 the flow $J = \beta(1 - P_0)$ is plotted in dependence of β for various threshold densities $\rho^* = N^*/L$ and for finite length $L = 100$ of the bottleneck. For the $M/M/1$ queue the flow is constantly equal to the arrival rate. Instead here one sees that the flow obeys

$$J = \begin{cases} \beta & \text{, for } \alpha_+ < \beta \leq \alpha_-, \\ \alpha_-, & \text{, for } \alpha_- < \beta \leq 1. \end{cases} \tag{10}$$

The left plots shows case 1. Here the model is obviously not able to reach the flow of the simple $M/M/1$ queue with arrival rate α_-. However in case 2 (right plots) one sees that the maximum flow is reached for $\alpha_- < \beta \leq 1 - \alpha_-$. Concluding one can say that the switch to a lower arrival rate yields in general to a reduction of the

6   *Marko Woelki*

flow. However, if a small arrival rate ($\alpha_- < 1/2$) is reduced further then for β in the range $\beta \in [\alpha_-, 1 - \alpha_-]$ the maximum flow is kept alive for large N^*. The reason is the negative drift $x_- < 1$ that leads to the fact that queues smaller than N^* tend to shrink so that the system is dominated by α_- (which equals the maximum flow). The average queue length $\langle N \rangle = \sum_{N=1}^{L} N P_{N,L}$ can easily be calculated from the stationary distribution which yields

$$
\begin{aligned}
\langle N \rangle = & Z_L^{-1} \frac{x_-}{(x_- - 1)^2} \\
& + Z_L^{-1} \frac{x_-^{N^*}[N^*(x_- - 1) - x_-]}{(x_- - 1)^2} \\
& + Z_L^{-1} \frac{x_-^{N^*} x_+^{L-N^*+1}[L(x_+ - 1) - 1]}{(x_+ - 1)^2} \\
& - Z_L^{-1} \frac{x_-^{N^*}[N^*(x_+ - 1) - x_+]}{(x_+ - 1)^2}.
\end{aligned}
\tag{11}
$$

5. Limiting behaviour of large systems

Now the limit of large N^* and L is considered. For N^*, L sufficiently large, one finds for the average queue length from (11):

$$
\langle N \rangle \text{ is dominated by}
\begin{cases}
\text{Third term} & \text{, for } 0 < \beta \le \alpha_+, \\
\text{Second and fourth term} & \text{, for } \alpha_+ < \beta \le \alpha_-, \\
\text{First term,} & \text{, for } \alpha_- < \beta \le 1.
\end{cases}
\tag{12}
$$

Those three regions define the different regions in the parameter space (α_-, β), the 'phase diagram':

$$
\langle N \rangle \sim
\begin{cases}
L - \frac{\beta}{\alpha_+ - \beta} & \text{, for } 0 < \beta < \alpha_+, \\
\frac{\beta - \alpha_+}{\alpha_+ - \alpha_-} \frac{N^*(\alpha_- - \beta) - \alpha_-}{\beta - \alpha_-} + \frac{\beta - \alpha_-}{\alpha_- - \alpha_+} \frac{N^*(\alpha_+ - \beta) - \alpha_+}{\beta - \alpha_+} & \text{, for } \alpha_+ < \beta < \alpha_-, \\
\frac{\alpha_-}{\beta - \alpha_-}, & \text{, for } \alpha_- < \beta < 1.
\end{cases}
\tag{13}
$$

Here we assumed that N^* is not close to 0 or L and β is not close to α_+ or α_-. So it turns out that in the regime given in the second line the average number of particles depends only on N^*, not on L. This is true as long as $x_+^{L-N^*+1}$ can be neglected. The other results follow from similar arguments. The first region is a high-density region where the queue has typically maximal length. The region given in the lowest line is of low density and recovers the usual M/M/1 queue with $\langle N \rangle = \alpha/(\beta - \alpha)$ for $\alpha := \alpha_-$. Results can be summarized most easily with the help of Figure 4. The diagram shows the parameter space of exit rate β and arrival rate α_- for a reduced arrival rate $\alpha_+ < 1/2$. In this way cases 1 and 2 can be described altogether. Again, the grey-shaded area not covered by the model definition. One distinguishes three regions: one high-density phase where $\alpha_- < \beta < 1$, one low-density phase where $0 < \beta < \alpha_+$ and one phase at optimized density where $\alpha_+ < \beta < \alpha_-$.
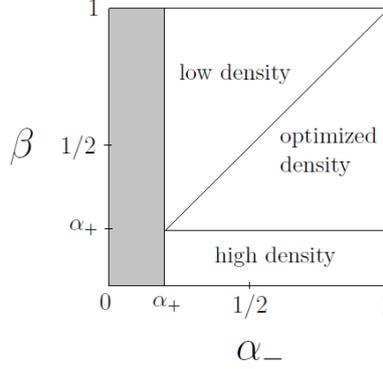
Fig. 4.   Different regimes in parameter space resulting from (13) the average queue length $\langle N \rangle$ in the queuing model. Grey shaded: physically unreasonable.

One observes that the different regimes and the physics of the average number of cars $\langle N \rangle$ from the approach in [8] (see Figure 1) are in accordance with the queuing model in the physical range $\beta \leq 1 - \alpha_-$. In the region where $\langle N \rangle \sim \alpha_-(\beta - \alpha_-)^{-1}$ (the low-density phase) the flow is $J = \alpha_-$ and in TASEP $J = \alpha_-(1 - \alpha_-)$. In the region where $\langle N \rangle \sim L$ (the high-density phase) the flow is $J = \beta$ (in TASEP: $J = \beta(1 - \beta)$). However, one main difference in TASEP is that one is able to maximize the flow by reduction of the arrival rate. This is possible since the flow-density relation is $J = \rho(1-\rho)$; reducing a high density to $1/2$ thus increases the flow to its maximum. In the corresponding regime in the queuing model one has $J = \beta$ instead, since the simple queuing dynamics does not allow for flow optimization. With (12) one finds for N^*, L large from (7):

$$P_{0,L} = Z_L^{-1} = \begin{cases} \frac{x_+ - 1}{x_-^{N^*} x_+^{L - N^* + 1}} & , \text{ for } 1 < x_+ < x_- \ (0 < \beta < \alpha_+), \\ \frac{(x_- - 1)(1 - x_+)}{x_-^{N^*}(x_- - x_+)} & , \text{ for } x_+ < 1 < x_- \ (\alpha_+ < \beta < \alpha_-), \\ 1 - x_-, & , \text{ for } x_+ < x_- \leq 1 \ (\alpha_- < \beta < 1), \end{cases} \quad (14)$$

and

$$P_{N,L} = \begin{cases} \frac{x_+ - 1}{x_+^{L - N + 1}} & 1 < x_+ < x_-, \\ \frac{(x_- - 1)(1 - x_+)}{(x_- - x_+)} \left[x_-^{-\Delta N} \theta(\Delta N) + x_+^{-\Delta N} \bar{\theta}(\Delta N) \right], & x_+ < 1 < x_-, \\ (1 - x_-) x_-^N, & x_+ < x_- \leq 1, \end{cases} \quad (15)$$

with $\Delta N = N^* - N$, the discrete Heaviside function $\theta(n) = 0$ for $n \leq 0$ and $\theta(n) = 1$ for $n \geq 1$, and $\bar{\theta} = 1 - \theta$. Here it was assumed that the particle number in the low-density (high-density) regime never increases (falls short of) the threshold N^*.

Finally consider the even stronger 'thermodynamic' limit $N^*, L \to \infty$ while keeping $\rho^* := N^*/L$ fixed: Note that the normalization $Z_{N^*/\rho^*} = \sum_{N=0}^{N^*} x_-^N + (x_-/x_+)^{N^*} \sum_{N=N^*+1}^{N^*/\rho^*} x_+^N$ diverges in the high-density regime $1 \leq x_+ < x_-$ and

no stationary state exists. Therefore a stationary flow can formally not be defined. That is the reason why in (10) the lower bound on β is explicitly given. Concluding, the high-density region is formally transferred into a physically unreachable region. However the same argument holds for the regime with optimized density $(x_+ < 1 < x_-)$, where $\sum_{N=1}^{N^*} x_-^N$ diverges so that no stationary optimization takes place. So in the thermodynamical limit the process is defined only in the low-density region $x_+ < x_- \leq 1$. As a consequence, there is a critical line that cuts the stationary regime towards the (non-stationary) optimized density regime at $x_- = 1$ $(\alpha_- = \beta)$, see next section.

6. Relaxation times

Figure 5 shows the time evolution of the queue length when starting from an empty queue (see Figure caption) for one stochastic realization each. In general queues grow with velocity $x_-/2$ from $N = 0$ at $t = 0$ until $N = N^*$. The sample from the high-density regime then continues growth with velocity $x_+/2$ until the queue saturates at $\langle N \rangle$. The corresponding time is $T = 2N^*/x_- + 2(\langle N \rangle - N^*)/x_+$. The other samples obviously lack the second kink – they grow with velocity $x_-/2$ and saturate at $\langle N \rangle$ after $T = 2N^*/x_-$. However, as expected on the critical line between divergent and convergent phases, for the second lowest curve (with pink squares, $\beta = \alpha_- = 0.6$) one observes very large fluctuations. The queue length takes any value between $\langle N \rangle_{\text{high density}}$ and $\langle N \rangle_{\text{opt. density}}$ with the same frequency so that on average one has approximately $\langle N \rangle_{\text{crit. line}} = N^*/2$. One sees that for the right
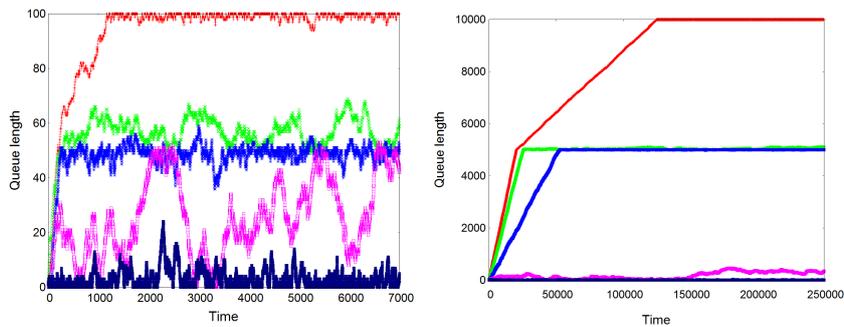
Fig. 5.   Five individual realizations of the dynamics: Development of the queue length for $\alpha_- = 0.6$, $\alpha_+ = 0.2$. From top to bottom graphs correspond to $\beta = 0.1$ (high density phase), $\beta = 0.2$ (line between high density and optimized density phase), $\beta = 0.4$ (optimized density phase), $\beta = 0.6$ (critical line between optimized density and high density phase) and $\beta = 0.8$ (high density phase). Left: $L = 100$, $N^* = 50$ and right $L = 10000$ and $N^* = 5000$.

realization the process struggles to reach the average queue length of 2500 which underlines that on this critical line the relaxation time diverges.

7. Conclusion

This paper studied a queueing model of a traffic bottleneck. While typically in queuing theory the queue length is unbounded, here a maximum possible size L was considered that represents the length of the bottleneck itself. At the left end cars arrive at a lower rate α_+ (instead of α_-) if the queue length N exceeds a certain value N^* and always leave at the right end of the system at constant exit rate β. For the TASEP it was shown recently [8] that this mechanism leads to regions in parameter space in which the density approaches the threshold N^*/L. It has been shown in the course of this paper that this regime can also be obtained from the simple queuing approach. The density can be tuned if the unreduced arrival rate is larger than the exit rate. Then choosing the reduced arrival rate smaller than the exit rate (but otherwise arbitrary) leads to the desired queue length. However in contrast to TASEP the flow can not be maximized by the simple queuing model since the queuing approach neglects the distribution of cars and empty space. The distribution of queue lengths has been obtained exactly and turns out to be sharply peaked at the threshold N^* where it takes its maximum value. Further the average queue length and flow have been calculated. For large system sizes the queueing process turned out to provide different density regimes that remind on the phase diagram found for the TASEP. However if the maximum possible queue length is taken to infinity then the region that is stationary accessible shrinks considerably. It was argued that the remaining convergent phase is separated from a divergent phase by a critical line on which the relaxation time diverges.

References

1. van Woensel T. and Vandaele N.: Modelling traffic flows with queuing models: a review. Asia-Pacific Journal of Operational Research, 24(4), 435-461 (2007)
2. Vandaele N., van Woensel T., and Verbruggen, A.: A queueing based traffic flow model. Transportation Research Part D 5 121-135 (2001)
3. Mahnke R. and Peret N. Stochastic master-equation approach to aggregation in freeway traffic. *Phys. Rev.* E 56 2666-2671 (1997)
4. Kleinrock, L. Queueing Systems Volume 1: Theory. ISBN 0471491101 (1975)
5. Arita C. Queuing process with excluded-volume effect. *Phys. Rev.* E 80, 051119 (2009)
6. K Mallick: Some exact results for the exclusion process. J. Stat. Mech., P01024 (2011)
7. Blythe, R.A., Evans, M.R.: Nonequilibrium steady states of matrix product form: A solver's guide. J Phys. A 40, R333 (2007)
8. Woelki M. Density feedback control in traffic and transport far from equilibrium. *Phys. Rev.* E 87, 062818 (2013)
9. Schultens C., Schadschneider A., and Arita C.: "Ergodicity breaking in an exclusion process with varying system length" arXiv:1409.0329v1 [cond-mat.stat-mech] 1 Sep 2014
10. Adams, D.A., Schmittmann, B., Zia, R.K.P.: Far-from-equilibrium transport with constrained resources. *J. Stat. Mech.* P06009 (2007)
11. Cook, L. J., Zia, R. K. P.: Feedback and Fluctuations in a Totally Asymmetric Simple Exclusion Process with Finite Resources. *J. Stat. Mech.* P02012 (2009)
12. Varghese, T. V. Analysis of some stochastic inventory systems subject to decay and

disaster, Thesis at Cochin University of Science and Technology, India (1998)
13. van Kampen N. G. Stochastic processes in physics and chemistry, Elsevier, The Netherlands. ISBN-13: 978-0-444-52965-7 (2007)
14. Klumpp, S. and Lipowsky, R. Cooperative cargo transport by several molecular motors. PNAS 102, 17284 - 17289 (2005).