Stochastic Systems
2017, Vol. 7, No. 1, 197–236
DOI: 10.1214/15-SSY208

WAVES IN A SPATIAL QUEUE

By David Aldous*

U.C. Berkeley

Envisaging a physical queue of humans, we model a long queue by a continuous-space model in which, when a customer moves forward, they stop a random distance behind the previous customer, but do not move at all if their distance behind the previous customer is below a threshold. The latter assumption leads to "waves" of motion in which only some random number W of customers move. We prove that $\mathbb{P}(W > k)$ decreases as order $k^{-1/2}$; in other words, for large k the k'th customer moves on average only once every order $k^{1/2}$ service times. A more refined analysis relies on a non-obvious asymptotic relation to the coalescing Brownian motion process; we give a careful outline of such an analysis without attending to all the technical details.

1. Introduction. Imagine you are the 100th person in line at an airport security checkpoint. As people reach the front of the line they are being processed steadily, at rate 1 per unit time. But you move less frequently, and when you do move, you typically move several units of distance, where 1 unit distance is the average distance between successive people standing in the line.

This phenomenon is easy to understand qualitatively. When a person leaves the checkpoint, the next person moves up to the checkpoint, the next person moves up and stops behind the now-first person, and so on, but this "wave" of motion often does not extend through the entire long line; instead, some person will move only a short distance, and the person behind will decide not to move at all. Intuitively, when you are around the k'th position in line, there must be some number $a(k)$ representing both the average time between your moves and the average distance you move when you do move – these are equal because you are moving forwards at average speed 1. In other words, the number W of people who move at a typical step has distribution $\mathbb{P}(W \geq k) = 1/a(k)$. This immediately suggests the question of how fast $a(k)$ grows with k. In this paper we will present a stochastic model and prove that, in the model, $a(k)$ grows as order $k^{1/2}$.

Received November 2015.
*Research supported by NSF Grants DMS-106998 and 1504802.
MSC 2010 subject classifications: Primary 60K25; secondary 60J05, 60J70.
Keywords and phrases: Coalescing Brownian motion, scaling limit, spatial queue.

In classical *queueing theory* [8], randomness enters via assumed randomness of arrival and service times. In contrast, even though we are modeling a literal queue, randomness in our model arises in a quite different way, via each customer's choice of exactly how far behind the preceding customer they choose to stand, after each move. That is, we assume that "how far behind" is chosen (independently for each person and time) from a given probability measure μ on an interval $[c_-, c^+]$ where $c_- > 0$. We interpret this interval as a "comfort zone" for proximity to other people. By scaling we may assume μ has mean 1, and then (excluding the deterministic case) μ has some variance $0 < \sigma^2 < \infty$.

In words, the model is

> when the person in front of you moves forward to a new position, then you move to a new position at a random distance (chosen from distribution μ) behind them, unless their new position is less than distance c^+ in front of your existing position, in which case you don't move, and therefore nobody behind you moves.

This is of course a "toy model" which for various reasons (some noted in section 1.5) we do not expect to provide a numerically accurate fit to data from real-world airport security queues. We have one informally collected data set, shown in section 5.1, showing features consistent with the model, but it would be interesting to collect more systematic data. Although we use the language of "human customers standing in line" throughout, one can imagine other contexts such as vehicle toll gate queues where similar models might be appropriate.

For the precise mathematical setup it is easiest to consider an infinite queue. A state $\mathbf{x} = (x_0, x_1, x_2, \ldots)$ of the process is a configuration

$$0 = x_0 < x_1 < x_2 < x_3 \ldots, \quad x_i - x_{i-1} \in [c_-, c^+]$$

representing the positions of the customers on the half-line $[0, \infty)$. The state space $\mathbb{X}$ is the set of such configurations. The process we study – let us call it the *spatial queue* process – is the discrete-time $\mathbb{X}$-valued Markov chain $(\mathbf{X}(t), t = 0, 1, 2, \ldots)$ whose transitions from an arbitrary initial state $\mathbf{x}$ to $\mathbf{X}(1)$ have the following distribution. Take i.i.d. $(\xi_i, i \geq 1)$ with distribution μ and define

$$\begin{aligned} X_0(1) &= 0 \\ W(1) &= \min\{i : \xi_1 + \ldots + \xi_i \geq x_{i+1} - c^+\} \\ X_i(1) &= \xi_1 + \ldots + \xi_i, \ 1 \leq i < W(1) \\ &= x_{i+1}, \ i \geq W(1). \end{aligned}$$

This simply formalizes the verbal description above. Note $W(1)$ is the number of customers who moved (including the customer moving to the service position 0) in that step; call $W(1)$ the *wave length* at step 1. As more terminology: in configuration $\mathbf{x} = (x_i, i \geq 0)$ the customer at *position* x_i has *rank* i. So in each step, a customer's rank will decrease by 1 whereas their position may stay the same or may decrease by some distance in $(0, \infty)$; recall that the random variables ξ do not directly indicate the distance a customer moves, but instead indicate the distance at which the customer stops relative to the customer ahead. For a general step of the process we take i.i.d. $(\xi_i(t), i \geq 1, t \geq 1)$ with distribution μ and define

$$
\begin{aligned}
(1) \qquad W(t) &= \min\{i : \xi_1(t) + \ldots + \xi_i(t) \geq X_{i+1}(t-1) - c^+\} \\
(2) \qquad X_i(t) &= \xi_1(t) + \ldots + \xi_i(t), \ 1 \leq i < W(t) \\
(3) \qquad &= X_{i+1}(t-1), \ i \geq W(t).
\end{aligned}
$$

Consistent with (2) we can set $X_0(t) = 0$; the "rank 0" customer is the customer being served, positioned at the origin of the spatial line. At each step the rank-1 customer moves to the server; the rank-2 customer becomes the rank-1 customer and will typically move forward (in which case $X_1(t) = \xi_1(t)$ because $W(t) > 1$) but sometimes will not move (in which case $X_1(t) = X_2(t-1)$ because $W(t) = 1$). Recall that $X_i(t)$ is the position of the rank-i customer at time t. The trajectory of a specific customer, say the customer initially at rank s, is therefore given by

$$X_s^*(t) = X_{s-t}(t), \ 0 \leq t \leq s$$

because at time t they have rank $s - t$.

Before continuing we should point out something we cannot prove, but strongly believe.

Conjecture 1. *The spatial queue process has a unique stationary distribution,* $\mathbf{X}(\infty)$ *say, and from any initial configuration we have* $\mathbf{X}(t) \to_d \mathbf{X}(\infty)$ *as* $t \to \infty$.

The "density component" assumption serves to discard cases where μ is supported on, say, $\{1, 2, 3\}$, in which case the behavior of the process with an initial configuration $\mathbf{x} = (x_i)$ with integer-valued x_i will be different from the behavior with a generic real-valued initial configuration.

Understanding the space-time trajectory of a typical customer far away from the head of the queue is essentially equivalent to understanding the large values of the induced process $(W(t), t \geq 1)$ of wave lengths, so that will be our focus. Intuitively we study properties of the stationary distribution, or of a typical customer in the stationary process, but we are forced to phrase results in a time-averaged way. Our fundamental result concerns the order of magnitude of long waves. Define

$$\rho^+(j) = \limsup_{\tau\to\infty} \tau^{-1} \sum_{t=1}^{\tau} \mathbb{P}(W(t) > j), \ \rho_-(j) = \liminf_{\tau\to\infty} \tau^{-1} \sum_{t=1}^{\tau} \mathbb{P}(W(t) > j).$$

PROPOSITION 2. *There exist constants* $0 < a_- < a^+ < \infty$ *(depending on* μ*) such that*

$$a_- j^{-1/2} \le \rho_-(j) \le \rho^+(j) \le a^+ j^{-1/2}, \ j \ge 1. \tag{4}$$

In other words for a rank-j customer, for large j, the average time between moves is order $j^{1/2}$. This result is proved in sections 2 and 3 as Proposition 11 and Corollary 15.

Building upon this order-of-magnitude result (Proposition 2), sharper results can be obtained by exploiting a connection with the coalescing Brownian motion (CBM) process, which we start to explain next.

1.1. *Graphics of realizations.* The graphics below not only illustrate the qualitative behavior of the process but also, via a non-obvious representation, suggest the main result.

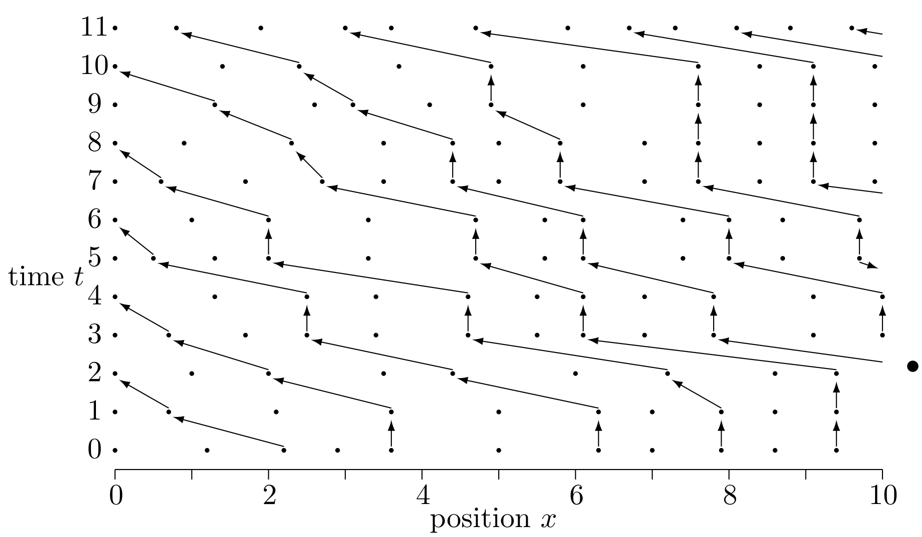

FIG 1. *A realization of the process, showing space-time trajectories of alternate customers. The customer with initial rank 12 is marked by* $\bullet$.

Figure 1 shows a realization of the process over the time interval $0 \leq t \leq 11$ and the spatial range $0 \leq x \leq 10$. Time increases upwards, and the head of the queue is on the left. The $\bullet$ indicate customer positions at each time, and the lines indicate the space-time trajectories of individual customers (for visual clarity, only alternate customers' trajectories are shown).

Although Figure 1 seems the intuitively natural way to draw a realization, a different graphic is more suggestive for mathematical analysis. A configuration $\mathbf{x} = (0 = x_0 < x_1 < x_2 < x_3 \ldots)$ can be represented by its centered counting function

$$F(x) := \max\{k : \ x_k \leq x\} \ - x, \quad 0 \leq x < \infty. \tag{5}$$

as illustrated in Figure 2.

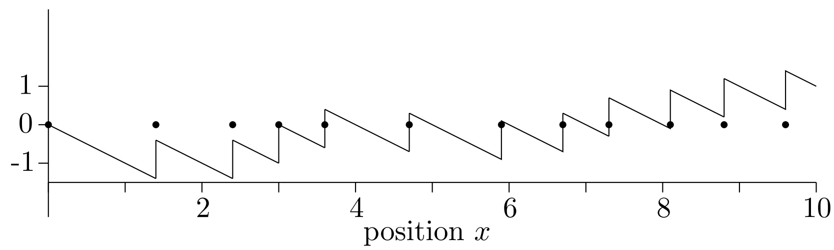

FIG 2. *The centered counting function associated with a configuration* $\bullet$.

At each time t, let us consider the centered counting function $F_t(x)$ and plot the graph of the upward-translated function

$$x \to G(t, x) := t + F_t(x). \tag{6}$$

In other words, we draw the function starting at the point $(0, t)$ instead of the origin. Taking the same realization as in Figure 1, and superimposing all these graphs, gives Figure 3.

Clearly the graphs in Figure 3 are "coalescing" in some sense. What is precisely true is the following. The assertion

> the rank k person (for some k) at time t is at position x^* and remains at that position x^* at time $t+1$

is equivalent to the assertion

> the graphs starting at $(t, 0)$ and at $(t+1, 0)$ both contain the same step of the form (s, x^*) to $(s+1, x^*)$ (for some s).

The former assertion implies that customers at ranks greater than k also do not move, and therefore the graphs in the second assertion coincide for all

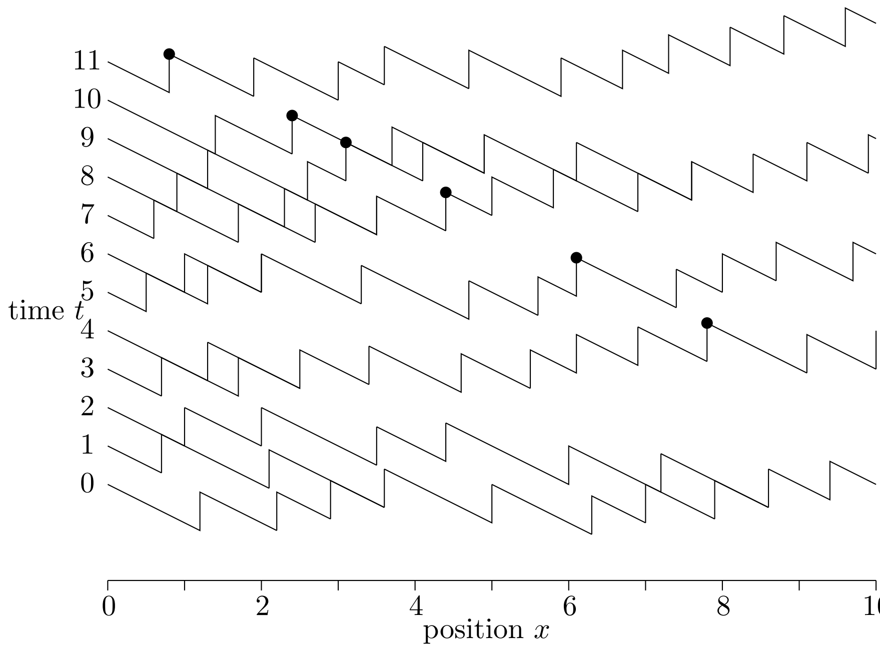

FIG 3. *A realization of the process $G(t,x)$. The • indicate the points associated with the moves of the customer with initial rank 12; compare with the space-time trajectory in Figure 1.*

$x \geq x^*$. So, even though the graphs may not coalesce upon first meeting (as is clear from the left region of Figure 3), they do coalesce at the first point where they make the same jump.

Many readers will immediately perceive that Figure 3 shows "something like" coalescing random walks in one dimension. Our goal is to formalize this as an asymptotic theorem. But note that, for a graphic of a model of coalescing random walks which is drawn to resemble Figure 3, the horizontal axis would be "time" and the vertical axis would be "space". So in making the analogy between our process and coalescing random walks, we are interchanging the roles of space and time. In particular, in Figure 3 our process is (by definition) Markov in the vertical direction, whereas models of coalescing random walks pictured in Figure 4 below are (by definition) Markov in the horizontal direction. These means that in conditioning arguments we will need to be careful, in fact somewhat obsessive, to be clear about what exactly is being conditioned upon.

1.2. *Coalescing Brownian motion on the line.* Given $\tau > 0$ and a random locally finite subset $\eta_\tau \subset \mathbb{R}$, it is straightforward to define a process

$(B_y(t), \tau \le t < \infty, y \in \eta_\tau)$ which formalizes the idea

(i) at time τ, start independent Brownian motions $(B^0_y(t), t \ge \tau)$ on $\mathbb{R}$ from each $y \in \eta_\tau$;
(ii) when two of these diffusing "particles" meet, they coalesce into one particle which continues diffusing as Brownian motion.

So $B_y(t)$ is the time-t position of the particle containing the particle started at y. Call this process the CBM(η_τ, τ) process. A line of research starting with Arratia [1] (see also [10]) shows this construction can be extended to the *standard CBM (coalescing Brownian motion) process* $(B(y,t), -\infty < y < \infty,\ t \ge 0)$ which is determined by the properties

(iii) given the point process $\eta_\tau = \{B(y,\tau),\ y \in \mathbb{R}\}$ of time-τ particle positions, the process subsequently evolves as CBM(η_τ, τ)
(iv) $\lim_{\tau \downarrow 0} \mathbb{P}(\eta_\tau \cap I = \emptyset) = 0$ for each open $I \subset \mathbb{R}$.

The latter requirement formalizes the idea that the process starts at time 0 with a particle at *every* point $y \in \mathbb{R}$. In particular, the point process $\eta_t = \{B(y,t),\ y \in \mathbb{R}\}$ of time-t particle positions inherits the scaling and translation-invariance properties of Brownian motion:

(v) For each $t > 0$, the point process $\eta_t(\cdot)$ is translation-invariant and so has some mean density ρ_t.
(vi) The distribution of the family $(\eta_t(\cdot), 0 < t < \infty)$ is invariant under the scaling map $(t,y) \to (ct, c^{1/2}y)$.
(vii) The distribution of $\eta_t(\cdot)$ is the push-forward of the distribution of $\eta_1(\cdot)$ under the map $y \to t^{1/2}y$ and so

$$\rho_t = t^{-1/2}\rho_1. \tag{7}$$

What will be important in this paper is an "embedding property". If we take the CBM$(\tilde{\eta}^0, 0)$ process, starting at time 0 with some translation-invariant locally finite point process $\tilde{\eta}^0$ with mean density $0 < \tilde{\rho}_0 < \infty$, and embed this process into standard CBM in the natural way, it is straightforward to see that the time-asymptotics of CBM$(\tilde{\eta}^0, 0)$ are the same as the time-asymptotics of standard CBM.

1.3. *What does the CBM limit tell us?.* Figure 3 suggests an asymptotic result of the form

> the spatial queue process, in the Figure 3 representation as a process $G(t,x)$, converges in distribution (in the usual random walk to Brownian motion scaling limit) to the CBM process.

We formalize this in section 4 as Theorem 17, by working on a "coalescing continuous functions" state space. Our formalization requires a double limit: we study the process as it affects customers beyond position x, in a time $t \to \infty$ limit, and show the $x \to \infty$ rescaled limit is a CBM process. And as in our earlier statement of Theorem 3, we can only prove this for time-averaged limits.

Now the Figure 3 representation as a process $G(t, x)$ is a somewhat "coded" version of the more natural Figure 1 graphic of customers' trajectories, so to interpret the CBM scaling limit we need to "decode" Figure 3. The wave lengths, measured by position of first unmoved customer, appear in Figure 3 as the positions x' where two functions $G(t', \cdot), G(t'', \cdot)$ coalesce. In other words the time-asymptotic rate of waves of length $> x$ equals the time-asymptotic rate at which distinct functions $G(t, \cdot)$ cross over x. In the space-asymptotic limit (of rank j or position x), as a consequence of results used to prove the order of magnitude bounds (Proposition 2) we have that $x/j \to 1$. Now the scaling limit identifies "distinct functions $G(t, \cdot)$ crossing over x" with CBM particles at time x, and so Theorem 17 will imply a result for the first-order asymptotics of wave length, measured by either rank or position, which sharpens Proposition 2. Recall σ is the s.d. of μ.

THEOREM 3. *For any initial distribution,*

$$\limsup_{\tau \to \infty} \tau^{-1} \sum_{t=1}^{\tau} \mathbb{P}(W(t) > i) \sim \rho_1 \sigma^{-1} i^{-1/2} \text{ as } i \to \infty$$

$$\liminf_{\tau \to \infty} \tau^{-1} \sum_{t=1}^{\tau} \mathbb{P}(W(t) > i) \sim \rho_1 \sigma^{-1} i^{-1/2} \text{ as } i \to \infty$$

where ρ_1 is the numerical constant (7) associated with CBM.

This will be a consequence of Corollary 21. As discussed earlier, if we knew there was convergence to stationarity we could restate this in the simpler way

$$\text{at stationarity, } \mathbb{P}(W(t) > i) \sim \rho_1 \sigma^{-1} i^{-1/2} \text{ as } i \to \infty. \tag{8}$$

The implications of the CBM scaling limit for the space-time trajectory of a typical customer are rather more subtle. As suggested by Figure 8, the trajectory is deterministic to first order, so the CBM limit must be telling us something about the second-order behavior.

Look at Figure 4, which shows CBM over "time" $0 \le x \le 1$ and "space" $-\infty < t < \infty$. There are positions $(w_u,\ u \in \mathbb{Z})$ of the particles at time $x = 1$; and

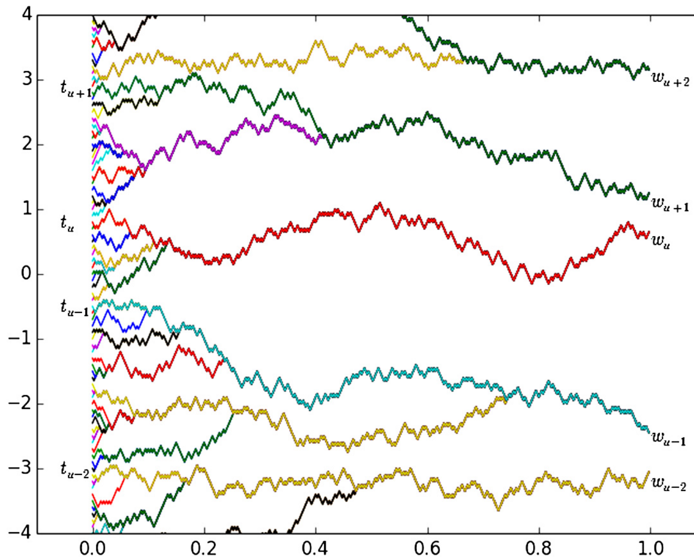

FIG 4. *Simulation of CBM over the time interval* $[0, 1]$.

(*) each such particle is the coalescence of the particles which started at $x = 0$ in some interval (t_{u-1}, t_u).

So we have two stationary point processes on $\mathbb{R}$, the processes $(w_u,\ u \in \mathbb{Z})$ and $(t_u,\ u \in \mathbb{Z})$, which must have the same rate ρ_1, and in fact have the same distribution, described in [12, 11].

In the queue model, as Figure 3 illustrates, the motion of one particular individual can be seen within the family of graphs as line segments within one diagonal line, To elaborate, write (x_u, s_u) for the coordinates of the top end points (the $\bullet$'s in Figure 3) of these segments, after the u'th move of the individual. The u'th move happens at some time t_u, when the individual moves to position x_u, and so $s_u = G(t_u, x_u)$. Until the next move, that is for $t_u \leq t < t_{u+1}$, we have (from the definition of G) that $G(t, x_u) = s_u$. At the time t_{u+1} of the next move, the individual moves to some position $x_{u+1} < x_u$. From the definition of G we have

$$s_{u+1} - s_u = G(t_{u+1}, x_{u+1}) - G(t_u, x_u) = -x_{u+1} + x_u.$$

So the points (x_u, s_u) lie on a diagonal line

$$x_u + s_u = \text{ constant.}$$

In fact the constant is the initial rank of the individual, or equivalently the time at which the individual will reach the head of the queue. What is important to note for our analysis is that the times t_u of the moves are not indicated directly by the points (x_u, s_u), but instead are specified by

$$G(t, x_u) = s_u,\ t_u \le t < t_{u+1}. \tag{9}$$

In words,

> (**) the t_u are the end-times of the intervals of t for which the functions $G(t, \cdot)$ have coalesced before the position of the individual.

So consider a customer at position approximately n, and hence with rank approximately n, for large n. From the order-of-magnitude results for wave lengths, the time between moves, and the distances moved, for this customer are both order $n^{1/2}$. But in the Brownian rescaling which takes Figure 3 to Figure 4, the horizontal axis is rescaled by $1/n$ and the vertical axis by $1/n^{1/2}$. That is, 1 unit time in the CBM model corresponds to n space units in the queue model, and 1 space unit in the CBM model corresponds to $n^{1/2}$ time units in the queue model. So when our position-n customer moves, in the CBM approximation the horizontal distance moved is only $O(n^{-1/2})$ whereas the vertical distance is order 1. So in the scaling limit the Figure 3 points corresponding to that customer become the points w_u on the vertical line in Figure 4. One consequence of this limit is provided by (*) and (**): the sequence of times of moves of that customer becomes, in the scaling limit, the sequence (t_u) in the CBM process indicating the spatial intervals of the time-0 particles that have coalesced into the time-1 particles. Precisely, from Theorem 17 we can deduce (section 4.2 gives the outline) the following corollary, in which the first component formalizes the assertion above.

Let U_τ have uniform distribution on $\{1, 2, \dots, \tau\}$.

Corollary 4. *In the queue model, consider the individual at time U_{τ_n} nearest to position n. Write $((t_u^{(n)}, w_u^{(n)}),\ u \ge 1)$ for the times t and end-positions w of moves of that individual. If $\tau_n \to \infty$ sufficiently fast then*

$$\left(\frac{t_u^{(n)} - U_{\tau_n}}{n^{1/2}}\ ,\ \frac{n - w_u^{(n)}}{n^{1/2}} \right)_{u \ge 1} \xrightarrow{d} (\sigma t_u, \sigma w_u)_{-\infty < u < \infty}$$

in the usual sense [6] of convergence of point processes on $\mathbb{R}^2$.

If we knew there was convergence to stationarity in the queue model then we could use τ_n in place of U_{τ_n}. Note that the phrase *for all $\tau_n \to \infty$ sufficiently fast ...* has its usual precise meaning: *there exist $\tau_n^* \to \infty$ such that for all sequences (τ_n) with $\tau_n \ge \tau_n^*$ ....*

How does the second component in Corollary 4 arise? Return to the setting of (**), and consider the graphs $G(t_{u+1}, \cdot)$ and $G(t_u, \cdot)$ coding configurations at successive moves of the individual under consideration. From the definition of G, the increment $G(t_{u+1}, x_u) - G(t_u, x_u)$ equals (up to $\pm O(1)$) the number (m, say) of customers who cross position x_u in the time-(t_{u+1}) wave. So the distance moved by the individual under consideration. is essentially the sum of these m inter-customer spacings. These spacings are i.i.d.(μ) subject to a certain conditioning, but the conditioning event does not have small probability, so the distance moved by that customer is $(1 \pm o(1))m$. The size of an increment $G(t_{u+1}, x) - G(t_u, x)$ as a function of x is first-order constant over the spatial interval $x \in (x_u \pm O(x_u^{1/2}))$ where the individual will be for the next $O(1)$ moves. So for large n and t, the successive distances moves by a rank n customer around time t are to first order the distances between the successive distinct values of $t \to G(t, n)$. In the scaling limit these become the successive increments of (w_u) (Figure 4) in the CBM process, explaining the second component in Corollary 4.

We will add some details to this outline in section 4.2.

1.4. *Outline of proofs.* Our proofs are rather complicated, in part because of the subtleties, mentioned earlier, arising in formalizing conditioning arguments. Simpler proofs may well be discovered in future.

The main goal of this paper is to prove Proposition 2, that waves of lengths $> k$ occur at rate of order $k^{-1/2}$, not faster or slower. As one methodology, we can analyze the distribution at a large time by looking backwards, because the current configuration must consist of successive "blocks" of customers, each block having moved to their current position at the same past time. In particular, because the inter-customer distances within a block are roughly i.i.d.(μ) (only *roughly* because of conditioning effects), and because block lengths tend to grow, we expect that (in our conjectured stationary limit) the space-asymptotic inter-customer distances should become approximately i.i.d.(μ). One formalization of this idea is given in Lemma 12, and this leads (section 3.2) to an order $k^{-1/2}$ lower bound on the rate of waves of lengths $> k$; but for technical reasons we first need to know the corresponding upper bound. Our proof of the upper bound (outlined at the start of section 2) is intricate, but the key starting point is the observation (section 2.1) that, after a long wave, the inter-customer spacings of the moved customers tend to be shorter than i.i.d.(μ); this and classical random walk estimates lead to an order $k^{1/2}$ lower bound on the mean time between successive waves of lengths $> k$, which is an order $k^{-1/2}$ upper bound on the rate of such waves. Although intricate in detail, all these proofs essentially rely only on explicit bounds, in contrast to those mentioned below.

In section 4 we develop the CBM scaling limit results, and their implications, discussed in section 1.3. Given the order of magnitude result (Proposition 2), this further development is essentially “soft”, involving weak convergence and compactness on an unfamiliar space (of coalescing continuous functions), combined with basic properties of CBM and classical weak convergence of random walks to Brownian motion. We give a careful outline of such an analysis without attending to all the technical details.

1.5. *Remarks on the model and related work.* Conceptually the model seems fairly realistic and to provide a plausible explanation for what the reader might have observed in long queues such as airport security lines. See section 5.1 for some data. In reality there are other reasons why a customer might not move when the “wave” reaches them – not paying attention because of talking or texting, for instance. Two assumptions of our model – that service times are constant, and that the successive customers in a wave move simultaneously – are unrealistic, but these are irrelevant to the particular feature (length of waves of motion) that we are studying. Intuitively the model behavior should also be robust to different customers having different parameters μ, c_-, c^+, but we have not investigated this carefully.

We do not know any previous work on closely related models. In one way, our model could be regarded as a continuum, syncronous-moves analog of the (discrete-space, asynchronous moves) TASEP which has been intensively studied [4] owing to its connections with other statistical physics models. The small literature on such analogs of TASEP (see [3] and citations therein) has focused on ergodic properties for processes on the doubly-infinite line. In another way, our phenomenon could be regarded as analogous to stop-and-go motion in traffic jams, for which realistic models would need to consider both positions and velocities of vehicles, although discrete-space discrete-time synchronous models have also been studied [7]. Academic literature such as [5] studying real world security lines does not address the specific “wave” feature studied in this paper.

2. Upper bounding the rate of long waves. In this section we show that the rate of waves of length $> j$ is $O(j^{-1/2})$. Informally, we do this by considering the progress of an individual from rank $j + o(j)$ to rank $j - o(j)$. The individual only moves when there is a wave of length $> j \pm o(j)$. The key idea is

> If the individual moves a distance $o(j^{1/2})$ then the next wave that moves the individual is unlikely to have length $> 2j$.

The argument is intricate and in fact uses that idea somewhat indirectly (Proposition 9). The remaining argument is outlined in section 2.2.

Figure 3 seems the most intuitive way to picture waves, in terms of their spatial extent, but for proofs it is more convenient to work with ranks. To track the progress of an individual we use the notation

$$X_s^*(t) = X_{s-t}(t),\ 0 \le t \le s.$$

That is, $X_s^*(\cdot)$ tracks the position of "individual s", the individual who starts at rank s and will reach the service position at time s. Write

$$\mathbf{X}_{[s]}(t) = (X_i(t),\ 0 \le i \le s-t)$$

for the positions of "individual s" and the customers ahead of that individual in the queue. Write

$$\mathcal{F}(t) = \sigma(\mathbf{X}(u), 0 \le u \le t)$$

for the natural filtration of our process, which is constructed in terms of i.i.d(μ) random variables $\{\xi_i(t), i \ge 1, t \ge 1\}$ as at (1–3).

2.1. *Close waves coalesce quickly.* The event that individual s is moved in the time-t wave is the event

$$\{X_s^*(t) < X_s^*(t-1)\} = \{W(t) > s-t\}.$$

Such a wave may be a "long" wave, in that it extends at least j customers past individual s. Our first goal is to show, roughly speaking, that after one long wave, and before individual s moves a distance $o(j^{1/2})$:

either (i) there is unlikely to be another long wave;
or (ii) there are likely to be many waves that move individual s.

This is formalized by Propositions 9 and 10, though complicated by the fact we will need to consider "good" long waves.

How far one wave reaches depends on the growth rate of the partial sums $\sum_{i=1}^k \xi_i$ associated with this new wave, relative to the partial sums associated with the previous long wave. Because we are interested in the difference, what is ultimately relevant is the distribution of the maximum of the symmetrized (hence mean zero) random walk

$$S_k^{sym} = \sum_{i=1}^k (\xi_i' - \xi_i'') \tag{10}$$

where (ξ_i') and (ξ_i'') denote independent i.i.d.(μ) sequences. Define

$$q(j,y) = \mathbb{P}(\max_{1 \le k \le j} S_k^{sym} \le y). \tag{11}$$

For later use, note that Donsker's theorem and the fact that the distribution of the maximum of Brownian motion over $0 \le t \le 1$ has bounded density imply

LEMMA 5. *There exists a constant C and $\delta_j \downarrow 0$, depending only on the distribution μ, such that*

$$q(j,y) \le Cyj^{-1/2} + \delta_j, \quad y \ge 1, j \ge 1.$$

Now consider

$$q_{j,y}(x_1,\dots,x_j) = \mathbb{P}\left(\max_{1\le k\le j} \sum_{i=1}^{k} (\xi'_i - x_i) \le y\right). \tag{12}$$

Say a sequence $(x_1,\dots,x_j)$ is (j,y)-*good* if $q_{j,y}(x_1,\dots,x_j) \le 2q(j,y)$, and note that for an i.i.d.(μ) sequence (ξ_i)

$$\begin{aligned}
\mathbb{P}((\xi_1,\dots,\xi_j) \text{ is not } (j,y)\text{-good }) &= \mathbb{P}(q_{j,y}(\xi_1,\dots,\xi_j) > 2q(j,y)) \\
&\le \frac{\mathbb{E}q_{j,y}(\xi_1,\dots,\xi_j)}{2q(j,y)} \\
&= \frac{q(j,y)}{2q(j,y)} = \frac{1}{2}. \qquad (13)
\end{aligned}$$

The proof of the next lemma exploits a certain monotonicity property: the spacings between customers tend to be shorter than usual, for the customers who have just moved in a long wave.

LEMMA 6. *For $\mathbf{x} \in \mathbb{X}, x' < x_{s-t_0+1}(t_0-1), 1 \le t_0 < s, \ j \ge 1$,*

$$\begin{aligned}
&\mathbb{P}\left[(\xi_{s-t_0+1}(t_0),\dots,\xi_{s-t_0+j}(t_0)) \textit{ is } (j,y)\textit{-good}\right. \\
&\qquad\qquad \left.|\mathbf{X}(t_0-1) = \mathbf{x}, X_{s-t_0}(t_0) = x', W(t_0) > s-t_0+j\right] \ge \frac{1}{2}.
\end{aligned}$$

PROOF. Given $\mathbf{X}(t_0-1) = \mathbf{x}$, the event $\{W(t_0) > s-t_0+j\}$ is the event $\{\sum_{i=1}^{k} \xi_i(t_0) < x_{k+1} - c^+, 1 \le k \le s-t_0+j\}$. This implies that, given $\mathbf{X}(t_0-1) = \mathbf{x}$ and $X_{s-t_0}(t_0) = x' < x_{s-t+1}(t_0-1)$ (the latter is equivalent to $X^*_s(t_0) = x' < X^*_s(t_0-1)$), the event $\{W(t_0) > s-t_0+j\}$ is the event

$$\{x' + \sum_{i=1}^{k} \xi_{s-t_0+i}(t_0) < x_{k+1} - c^+, 1 \le k \le j\}. \tag{14}$$

Write $(\hat{\xi}_i, 1 \le i \le j)$ for random variables with the conditional distribution of $(\xi_{s-t_0+i}(t_0), 1 \le i \le j)$ given event (14). Write $(\xi_i, 1 \le i \le j)$ for an independent i.i.d.(μ) sequence. The conditional probability of event (14) given $\xi_{s-t_0+1}(t_0)$ is clearly a decreasing function of $\xi_{s-t_0+1}(t_0)$. From this and Bayes rule we see that the distribution of $\hat{\xi}_1$ is stochastically smaller than μ, in other words we can couple $\hat{\xi}_1$ and ξ_1 so that $\hat{\xi}_1 \le \xi_1$ a.s. Repeating this argument inductively on i, we can couple $(\hat{\xi}_i, 1 \le i \le j)$ and $(\xi_i, 1 \le i \le j)$ so that $\hat{\xi}_i \le \xi_i$, $1 \le i \le j$. Now the function $q_{j,y}(x_1, \ldots, x_j)$ is monotone increasing in each argument x_i, so

$$\text{if } (\xi_i, 1 \le i \le j) \text{ is } (j,y)\text{-good then } (\hat{\xi}_i, 1 \le i \le j) \text{ is } (j,y)\text{-good.}$$

Now (13) implies that $(\hat{\xi}_i, 1 \le i \le j)$ is (j,y)-good with probability $\ge 1/2$. But this is the assertion of the lemma. □

The next lemma starts to address the "key idea" stated at the start of section 2. Given that a time-t_0 wave extends at least j places past individual s, and given that a time-t_0+t wave reaches individual s, we can upper bound the probability this time-t_0+t wave extends at least j places past individual s, and the upper bound is small if individual s has moved only distance $o(j^{1/2})$ between times t_0 and t_0+t.

LEMMA 7. *For* $t_0 \ge 1, t \ge 1, s > t_0 + t, j \ge 1$,

$$\begin{aligned}\mathbb{P}(W(t_0+t) > s - t_0 - t + j \,|\, \mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1)) \\ \le q_{j, X^*_s(t_0) - X^*_s(t_0+t)}(\xi_{s-t_0+1}(t_0), \ldots, \xi_{s-t_0+j}(t_0)) \\ \text{on } \{W(t_0+t) > s - t_0 - t\} \cap \{W(t_0) > s - t_0 + j\}.\end{aligned}$$

Note that the event $\{W(t_0+t) > s - t_0 - t\}$ is the event $\{X^*_s(t_0+t) < X^*_s(t_0+t-1)\}$ and so is indeed in $\sigma(\mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1))$. Also on the event $\{W(t_0) > s - t_0 + j\}$ we have (for $1 \le i \le j$) that $\xi_{s-t_0+i}(t_0) = X_{s-t_0+i}(t_0) - X_{s-t_0+i-1}(t_0)$ and so $\xi_{s-t_0+i}(t_0)$ is $\mathcal{F}(t_0)$-measurable.

PROOF. On the event $\{W(t_0+t) > s - t_0 - t\}$ we have $\{X^*_s(t_0+t) < X^*_s(t_0+t-1)\}$, and given this, the event $\{W(t_0+t) > s - t_0 - t + j\}$ is the event

$$\{X^*_s(t_0+t) + \sum_{i=1}^{k} \xi_{s-t_0-t+i}(t_0+t) < X^*_{s+k}(t_0+t-1) - c^+, 1 \le k \le j\}.$$

Now $X^*_{s+k}(t_0+t-1) \le X^*_{s+k}(t_0)$ and $-c^+ < 0$, so the event above is a subset of the event

$$\{\sum_{i=1}^{k} \xi_{s-t_0-t+i}(t_0+t) < X^*_{s+k}(t_0) - X^*_s(t_0+t), 1 \le k \le j\}. \tag{15}$$

Restricting further to the event $\{W(t_0) > s - t_0 + j\}$, on which $X^*_{s+k}(t_0) - X^*_s(t_0) = \sum_{i=1}^k \xi_{s-t_0+i}(t_0),\ 1 \le k \le j$, we can rewrite (15) as the event

$$\{\max_{1\le k\le j} \sum_{i=1}^{k} (\xi_{s-t_0-t+i}(t_0+t) - \xi_{s-t_0+i}(t_0)) < X^*_s(t_0) - X^*_s(t_0+t)\}. \tag{16}$$

So the conditional probability in the statement of the lemma is bounded by the conditional probability, given $\mathcal{H} := \sigma(\mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1))$, of event (16). The random variables $\xi_{s-t_0+i}(t_0)$, $X^*_s(t_0)$, $X^*_s(t_0+t)$ are $\mathcal{H}$-measurable and the $\xi_{s-t_0-t+i}(t_0+t)$ are independent of them. So the conditional probability of event (16) given $\mathcal{H}$ on the subsets specified equals precisely the bound stated in the lemma, by definition (12) of $q_{j,y}(\cdot)$. $\square$

The way we will combine these lemmas is abstracted in the next lemma, in which $S \wedge T$ denotes $\min(S,T)$. Also, we rather pedantically write $\sigma(A_t, \mathcal{F}_{t-1})$ in the conditioning statement to emphasize that we are conditioning on a σ-field rather than an event.

LEMMA 8. *Let $1 \le T < \infty$ be a stopping time for a filtration $(\mathcal{F}_t, t = 0,1,2,\ldots)$. Let A_t and B_t be events in $\mathcal{F}_t$ satisfying*

$$B_t \subseteq A_t \tag{17}$$

$$\mathbb{P}(B_t|\sigma(A_t,\mathcal{F}_{t-1})) \ \le\ \delta \text{ on } A_t \cap \{T \ge t\},\ t \ge 1. \tag{18}$$

Define

$$N(\tau) = \sum_{t=1}^{\tau} \mathbb{1}_{A_t}$$

$$S = \min\{t \ge 1 :\ B_t \text{ occurs}\}.$$

Then $\mathbb{P}(T \le S) \le \delta \mathbb{E} N(S \wedge T)$.

PROOF. Rewriting (18) using indicator random variables

$$\mathbb{P}(B_t|\sigma(A_t,\mathcal{F}_{t-1}))\mathbb{1}_{\{T\ge t\}}\mathbb{1}_{A_t} \le \delta \mathbb{1}_{\{T\ge t\}}\mathbb{1}_{A_t}.$$

Because (17) implies $\mathbb{P}(B_t|\sigma(A_t,\mathcal{F}_{t-1})) = 0$ on A_t^c we can delete the 1_{A_t} term from the left side of the inequality above and write

$$\mathbb{P}(B_t|\sigma(A_t,\mathcal{F}_{t-1}))\mathbb{1}_{\{T\ge t\}} \le \delta \mathbb{1}_{\{T\ge t\}}\mathbb{1}_{A_t}.$$

Conditioning on $\mathcal{F}_{t-1}$ and noting $\{T \ge t\} \in \mathcal{F}_{t-1}$ we have

$$\mathbb{P}(B_t|\mathcal{F}_{t-1})\mathbb{1}_{\{T\ge t\}} \le \delta \mathbb{P}(A_t|\mathcal{F}_{t-1})\mathbb{1}_{\{T\ge t\}}. \tag{19}$$

Now

$$\begin{aligned}\mathbb{P}(S=t, S\le T) &= \mathbb{P}(B_t, S\ge t, T\ge t)\\ &= \mathbb{E}\left[\mathbb{P}(B_t|\mathcal{F}_{t-1})\mathbb{1}_{\{T\ge t\}}\mathbb{1}_{\{S\ge t\}}\right]\\ &\le \delta\mathbb{E}\left[\mathbb{P}(A_t|\mathcal{F}_{t-1})\mathbb{1}_{\{T\ge t\}}\mathbb{1}_{\{S\ge t\}}\right] \text{ by (19)}\\ &= \delta\mathbb{P}(A_t, S\wedge T\ge t).\end{aligned}$$

Summing over $t\ge 1$ gives the result. □

We will now fit together the results above to obtain a result stated as Proposition 9 below. The statement involves substantial notation, and it is less cumbersome to develop the notation as we proceed.

Fix even $j\ge 4$ and $s>j$ and $y>0$. Start the process with $\mathbf{X}(0)=\mathbf{x}$ arbitrary. Consider events, for $1\le t<s$,

$A_t=\{W(t)>s-t\}$
$C_t=\{W(t)>s-t+j\}$
$B_t=C_t\cap\{(\xi_{s-t+1}(t),\dots,\xi_{s-t+j})$ is (j,y)-good $\}$.

In words:

$A_t=$ "customer s moves at time t"
$C_t=$ "a long wave at time t"
$B_t=$ "the long wave at time t is good".

Note that we are measuring wave length relative to the current position of customer s. We study what happens between the times of good long waves, that is the times

$$T_1 = \min\{t\ge 1:\ B_t \text{ occurs }\} \tag{20}$$

$$T_u = \min\{t> T_{u-1}:\ B_t \text{ occurs }\}, u\ge 2. \tag{21}$$

Note T_u may be infinite (if no such event occurs); this will not be an issue because we will be conditioning on $\{T_u=t_0\}$ and will truncate T_{u+1}.

Fix $u\ge 1$ and $1\le t_0<j/2$. Recall we previously fixed $s>j\ge 4$ and $y>0$; all these variables are integers except y.

By definition, B_{t_0} is the event that the realization of $(\xi_{s-t_0+i}(t_0) = X_{s-t_0+i}(t_0)-X_{s-t_0+i-1}(t_0),\ 1\le i\le j)$ is (j,y)-good. For each $1\le t<s-t_0$, Lemma 7 and the fact that $q_{j,y}(\cdot)$ is monotone increasing in y imply

$$\begin{aligned}&\mathbb{P}(C_{t_0+t}|\mathbf{X}^*_{[s]}(t_0+t),\mathcal{F}(t_0+t-1))\le 2q(j,y)\\ &\text{on } A_{t_0+t}\cap B_{t_0}\cap\{X^*_s(t_0)-X^*_s(t_0+t)\le y\}\cap\{T_u=t_0\}.\end{aligned} \tag{22}$$

Now write

$$\widehat{T}=\min\{t\ge 1: X^*_s(t_0)-X^*_s(t_0+t)>y\}$$

so that the penultimate event in the intersections in (22) is the event $\{\widehat{T} > t_0 + t\}$. Because $B_{t_0+t} \subseteq C_{t_0+t}$ we now have

$$\mathbb{P}(B_{t_0+t}|\mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1)) \le 2q(j,y)$$
$$\text{on } A_{t_0+t} \cap B_{t_0} \cap \{\widehat{T} > t_0+t\} \cap \{T_u = t_0\}.$$

Because $B_{t_0} \cap \{\widehat{T} > t_0+t\} \in \mathcal{F}(t_0+t-1)$ and $A_{t_0+t} \in \sigma(\mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1))$ we can condition down from $\sigma(\mathbf{X}^*_{[s]}(t_0+t), \mathcal{F}(t_0+t-1))$ to $\sigma(A_{t_0+t}, \mathcal{F}(t_0+t-1))$ and write

$$\mathbb{P}(B_{t_0+t}|A_{t_0+t}, \mathcal{F}(t_0+t-1)) \le 2q(j,y)$$
$$\text{on } A_{t_0+t} \cap B_{t_0} \cap \{\widehat{T} > t_0+t\} \cap \{T_u = t_0\}. \tag{23}$$

Define

$$N_A(\tau) = \sum_{t=1}^{\tau} \mathbb{1}_{A_t}$$

and similarly for $N_B(\tau)$ and $N_C(\tau)$.

We can now apply Lemma 8, restated for times $t_0 < t < \infty$. Condition on $\sigma(\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0))$ and restrict to the event $\{T_u = t_0\}$. In the hypotheses of Lemma 8 take $T = \widehat{T} \wedge j/2, S = T_{u+1}, \delta = 2q(j,y)$. So inequality (18) holds by (23). The conclusion of that lemma is now

> conditional on $\sigma(\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0))$ and restricted to the event $\{T_u = t_0\}$, we have;
> $\mathbb{P}(\widehat{T} \wedge j/2 \le T_{u+1}) \le 2q(j,y)\mathbb{E}(N_A(T_{u+1} \wedge \widehat{T} \wedge j/2) - N_A(T_u))$.

Rewriting this with all conditioning made explicit:

$$\begin{aligned}&\mathbb{P}(\widehat{T} \wedge j/2 \le T_{u+1}|\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0)) \\ &\quad \le 2q(j,y)\mathbb{E}(N_A(T_{u+1} \wedge \widehat{T} \wedge j/2) - N_A(T_u)|\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0)) \\ &\qquad \text{on } \{T_u = t_0\}.\end{aligned}$$

The event $\{\widehat{T} \wedge j/2 \le T_{u+1}\}$ is the event $\{T_{u+1} \le j/2, X^*_s(T_u) - X^*_s(T_{u+1}) \le y\} \cup \{T_{u+1} > j/2\}$, and because $N_A(T_{u+1} \wedge \hat{T} \wedge j/2) \le N_A(T_{u+1} \wedge j/2)$ we have

$$\begin{aligned}&\mathbb{P}(X^*_s(T_u) - X^*_s(T_{u+1}) \le y, T_{u+1} \le j/2|\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0)) \\ &\quad \le 2q(j,y)\mathbb{E}(N_A(T_{u+1} \wedge j/2) - N_A(T_u)|\{T_u = t_0\}, \mathcal{F}(t_0-1), X^*_s(t_0)) \\ &\qquad \text{on } \{T_u = t_0\}.\end{aligned}$$

The notation has become rather cumbersome; we will make it more concise by defining the σ-field $\mathcal{F}^-(T_u)$ as follows.

For each t_0, the restriction of $\mathcal{F}^-(T_u)$ to $\{T_u = t_0\}$ is the restriction of $\sigma(\{T_u = t_0\}, \mathcal{F}(t_0 - 1), X^*_s(t_0))$ to $\{T_u = t_0\}$.

Here $\mathcal{F}^-(T_u)$ is a sub-σ-field of the usual pre-T_u σ-field $\mathcal{F}(T_u)$. In words, when $T_u = t_0$ the usual $\mathcal{F}(T_u)$ tells us all about $\mathbf{X}(t_0)$ whereas $\mathcal{F}^-(T_u)$ only tells us $X^*_s(t_0)$ and the fact that a good long wave has just occured. Note that T_u is indeed $\mathcal{F}^-(T_u)$-measurable, as is $X^*_s(T_u)$. Now the inequality above can be rewritten more concisely as

PROPOSITION 9. *For $u \geq 1$, even $j \geq 4, s > j, y > 0$*

$$\begin{aligned}&\mathbb{P}(X^*_s(T_u) - X^*_s(T_{u+1}) \leq y, T_{u+1} \leq j/2 | \mathcal{F}^-(T_u))\\ &\quad\leq 2q(j,y)\mathbb{E}(N_A(T_{u+1} \wedge j/2) - N_A(T_u)|\mathcal{F}^-(T_u))\\ &\qquad \textit{on } \{T_u < j/2\}.\end{aligned}$$

In the next section we will use this result with $y = o(j^{1/2})$, implying $q(j,y)$ is small. Note that Proposition 9 formalizes the property stated at the start of this section

after one long wave, and before individual s moves a distance $o(j^{1/2})$:

either (i) there is unlikely to be another long wave;
or (ii) there are likely to be many waves that move individual s.

except that the Proposition refers to *good* long waves. However, Lemma 6 says that each long wave has probability $\geq 1/2$ to be good, independently of the past, implying that the mean number of long waves between successive good long waves is ≤ 2, and more precisely

LEMMA 10. *For even $j \geq 4$, $s > j$*

$$\mathbb{E}N_C(j/2) \leq 2\mathbb{E}N_B(j/2).$$

2.2. *The upper bound.* Write

$$N_j(\tau) = \sum_{t=1}^{\tau} \mathbb{1}_{\{W(t) > j\}} \tag{24}$$

for the number of waves of length $> j$ up to time τ, and then write

$$\rho^+(j) = \limsup_{\tau\to\infty} \tau^{-1}\mathbb{E}N_j(\tau).$$

PROPOSITION 11. *$\rho^+(j) = O(j^{-1/2})$ as $j \to \infty$.*

Outline of proof. We track the initial rank-j customer for time $j/2$. If the number of good long waves during this time is $O(j^{1/2})$ then we are done, because by Lemma 10 the number of waves of length $> 2j$ must also be $O(j^{1/2})$ and hence their rate is $O(j^{-1/2})$. So suppose the number N_j^* of good long waves is larger than $O(j^{1/2})$. Then we can choose $y_j = o(j^{1/2})$ such that $N_j^* y_j$ grows faster than $O(j)$. Then for most of these N_j^* waves the probability of $\{X_s^*(T_u) - X_s^*(T_{u+1}) \leq y_j\}$ must be $> 1/2$ and so by Proposition 9, because $q(j, y_j) = o(1)$, the mean number of waves of length $> j/2$ is a large multiple of the mean number of waves of length $> 2j$. In other words, writing W for typical wave length,

$$\text{if } \mathbb{P}(W > j) \neq O(j^{-1/2}) \text{ then } \mathbb{P}(W > 2j)/\mathbb{P}(W > j/2) \text{ is small}$$

and this easily implies that in fact $\mathbb{P}(W > j) = O(j^{-1/2})$.

Details of proof. As in the previous section we track customer s, but we now set $s = j$ and track the customer for time $j/2$. Fix y. Take expectation in Proposition 9 over $\{T_u < j/2\}$ to get

$$\begin{aligned} &\mathbb{P}(X_s^*(T_u) - X_s^*(T_{u+1}) \leq y,\ T_{u+1} \leq j/2) \\ &\quad \leq 2q(j, y)\mathbb{E}(N_A(T_{u+1} \wedge j/2) - N_A(T_u))\mathbb{1}_{(T_u < j/2)}. \end{aligned} \tag{25}$$

Note that the number U_j of (j, y)-good long waves before time $j/2$ can be represented as

$$U_j = \min\{u : T_{u+1} > j/2\}.$$

Summing (25) over $u \geq 1$ gives

$$\begin{aligned} &\mathbb{E}\ |\{1 \leq u \leq U_j - 1 :\ X_s^*(T_u) - X_s^*(T_{u+1}) \leq y_j\}| \\ &\quad \leq 2q(j, y)\ \mathbb{E}(N_A(j/2) - N_A(T_1)). \end{aligned} \tag{26}$$

Because the initial position of customer j is at most $c^+ j$ we have

$$\sum_{u=1}^{U_j - 1} (X_s^*(T_u) - X_s^*(T_{u+1})) \leq c^+ j$$

and hence

$$|\{1 \leq u \leq U_j - 1 :\ X_s^*(T_u) - X_s^*(T_{u+1}) > y\}| \leq c^+ j/y$$

implying

$$|\{1 \leq u \leq U_j - 1 :\ X_s^*(T_u) - X_s^*(T_{u+1}) \leq y\}| \geq U_j - 1 - c^+ j/y.$$

Taking expectation and combining with (26), and noting $U_j = N_B(j/2)$, we see

$$\mathbb{E}N_B(j/2) - 1 - c^+ j/y \leq 2q(j, y)\ \mathbb{E}N_A(j/2).$$

Now our customer s has rank decreasing from j to $j/2$, so a wave that moves this customer must have length at least $j/2$:

$$N_A(j/2) \le N_{j/2}(j/2)$$

in the notation at (24). Similarly, a wave reaching rank $2j$ is a long wave, so $N_C(j/2) \ge N_{2j}(j/2)$, and combining with Lemma 10 shows

$$\mathbb{E}N_B(j/2) \ge \tfrac{1}{2}\mathbb{E}N_{2j}(j/2).$$

So now we have

$$\mathbb{E}N_{2j}(j/2) \le 2 + 2c^+ j/y + 4q(j,y)\mathbb{E}N_{j/2}(j/2).$$

The initial configuration is arbitrary, so by averaging over consecutive time-intervals of length $j/2$ and taking time-asymptotics,

$$\tfrac{1}{4}\rho^+(2j) \le \tfrac{1}{j} + \tfrac{c^+}{y} + q(j,y)\rho^+(j/2).$$

Now set

$$\psi(j) = j^{1/2}\rho^+(j).$$

Using the Lemma 5 bound $q(j,y) \le Cyj^{-1/2} + \delta_j$, the previous inequality becomes

$$\tfrac{1}{4}\psi(2j) \le \tfrac{2^{1/2}}{j^{1/2}} + \tfrac{2^{1/2}j^{1/2}c^+}{y} + \tfrac{2Cy}{j^{1/2}}\psi(j/2) + 2\delta_j\psi(j/2).$$

Optimizing over choice of $y = y_j$ gives

$$\tfrac{1}{4}\psi(2j) \le \tfrac{2^{1/2}}{j^{1/2}} + 2^{7/4}(c^+C)^{1/2}\psi^{1/2}(j/2) + 2\delta_j\psi(j/2).$$

From this it is clear that

$$\limsup_{i\to\infty}\psi(4^i) < \infty$$

and then because $\rho^+(j)$ is decreasing we have $\limsup_{i\to\infty}\psi(j) < \infty$. $\square$

3. Lower bounding the rate of long waves. In this section we will show that waves of length $> k$ occur at rate no em less than order $k^{-1/2}$. Some of the estimates we give could likely be improved (see remark below Lemma 16, for instance) but because the previous upper bound shows we end up with the correct order of magnitude, we have not attempted such improvements of intermediate results.

3.1. *The blocks argument.* Here we will employ the "blocks" argument mentioned in the introduction. Fix $t_0 > 1$ and $i \geq 0$ and consider individual $s = t_0 + i$, who is at position $X_i(t_0) = X_s^*(t_0)$ at time t_0. Looking backwards in time, the time since the last move of that individual is

$$L_s(t_0) = \max\{\ell \geq 0 : X_s^*(t_0) = X_s^*(t_0 - \ell)\}$$

where we set $L_s(t_0) = t_0$ if s has never moved, that is if $X_s^*(t_0) = X_s^*(0)$. Recall the notation

$$\mathbf{X}_{[s]}(t) = (X_i(t),\ 0 \leq i \leq s - t)$$

for the positions of individual s and the customers ahead of that individual in the queue. Consider the event

$$A_{s+1}(t_0) := \{L_s(t_0) = L_{s+1}(t_0) < t_0\} \tag{27}$$

that individuals s and $s+1$ are in the same "block" of customers whose last move before t_0 was at the same time. Define

$$\mathcal{G}_s(t_0) = \sigma(\mathbf{X}_{[s]}(t), 0 \leq t \leq t_0; A_{s+1}(t_0)).$$

LEMMA 12. *For $t_0 > 1$ and $i \geq 0$ and $s = t_0 + i$,*

$$\mathbb{P}(X_{s+1}^*(t_0) - X_s^*(t_0) \in \cdot | \mathcal{G}_s(t_0)) = \mu(\cdot) \text{ on } A_{s+1}(t_0).$$

So by conditioning down and changing notation to $X_i(t_0) = X_s^*(t_0)$ we have

$$\mathbb{P}(X_{i+1}(t_0) - X_i(t_0) \in \cdot | X_j(t_0), 1 \leq j \leq i, A_{t_0+i+1}(t_0)) = \mu(\cdot) \text{ on } A_{t_0+i+1}(t_0). \tag{28}$$

We will later show that $1 - \mathbb{P}(A_{t_0+i+1}(t_0))$ becomes small for large i, and then Lemma 12 provides one formalization of the idea, mentioned earlier, that the inter-customer distances within the queue are roughly i.i.d. (μ).

PROOF OF LEMMA 12. The history of the individual s, that is $(X_s^*(t), 0 \leq t \leq t_0)$, tells us the value $\ell = L_s(t_0)$. Consider that history and the history of individual $s+1$ up to time $t_0 - \ell - 1$. Individual $s+1$ moved at time $t_0 - \ell$ if and only if the event

$$B_\ell := \{X_s^*(t_0 - \ell) < X_{s+1}^*(t_0 - \ell - 1) - c^+\}$$

happened. On the event B_ℓ, by construction (1–3) $X_{s+1}^*(t_0 - \ell) - X_s^*(t_0 - \ell) = \xi_{i+\ell+1}(t_0 - \ell)$. But time $t_0 - \ell$ was the time of the last move of individual

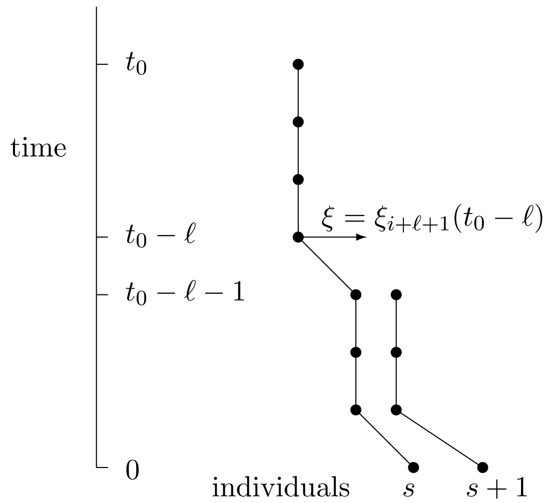

FIG 5. *Individual $s+1$ moves at time $t_0-\ell$ on event B_ℓ.*

s, so individual $s+1$ cannot subsequently move; the spacing is unchanged. See Figure 5.

So

$$X^*_{s+1}(t_0)-X^*_s(t_0)=\xi_{i+\ell+1}(t_0-\ell)\text{ on }B_\ell\cap\{L_s(t_0)=\ell\}.$$

Now the event $B_\ell\cap\{L_s(t_0)=\ell\}$ and the random variables $(\mathbf{X}_{[s]}(t),0\le t\le t_0,X_{s+1}(t),0\le t\le t_0-\ell-1)$ are determined by members of the constructing family $(\xi_j(t),j,t\ge 1)$ not including $\xi_{i+\ell+1}(t_0-\ell)$. So by independence of the constructing family,

$$\begin{aligned}\mathbb{P}(X^*_{s+1}(t_0)-X^*_s(t_0)\in\cdot|\mathbf{X}_{[s]}(t),0\le t\le t_0,X_{s+1}(t),0\le t\le t_0-\ell-1)&\\=\mu(\cdot)\text{ on }B_\ell\cap\{L_s(t_0)=\ell\}.&\end{aligned}$$

Now $A_{s+1}(t_0)$ is the disjoint union $\cup_{0\le\ell<t_0}[B_\ell\cap\{L_s(t_0)=\ell\}]$ and the result follows easily. $\square$

REMARK. Using Lemma 12 and (28) to deduce properties of the distribution of $\mathbf{X}(t_0)=(X_i(t_0),i\ge 0)$ requires some care, because there is some complicated dependence between the event $A_{t_0+i+1}(t_0)$ and the tail sequence $(X_j(t_0),j\ge i+1)$ which we are unable to analyze. This is the obstacle to a "coupling from the past" proof that for large t_0 the configuration $\mathbf{X}(t_0)$ can be defined in terms of the past ξ's with vanishing dependence on the initial configuration $\mathbf{X}(0)$, and thereby proving weak convergence to a unique stationary distribution.

The following argument is unaffected by such dependence. Fix t_0 and inductively on $i \geq 0$ construct $\xi^*_{i+1}(t_0)$ by

on $A_{t_0+i+1}(t_0)$ let $\xi^*_{i+1}(t_0) = X_{i+1}(t_0) - X_i(t_0)$

on $A^c_{t_0+i+1}(t_0)$ take $\xi^*_{i+1}(t_0)$ to have distribution μ independent of all previously-defined random variables.

Then using (28) the sequence $(\xi^*_i(t_0), i \geq 1)$ is i.i.d.(μ). Write $S_k(t_0) = \sum_{i=1}^k \xi^*_i(t_0)$. By construction we immediately have

LEMMA 13.

$$|X_k(t_0) - S_k(t_0)| \leq (c^+ - c_-) \sum_{i=1}^k \mathbb{1}_{A^c_{t_0+i}(t_0)}.$$

3.2. *The lower bound.* We can now outline the proof that waves of length $>k$ occur at rate no less than order $k^{-1/2}$. By showing that the bound in Lemma 13 is sufficiently small (Lemma 16), we will see that the rank-k individual at time t_0, for typical large k and t_0, is at position $k \pm O(k^{1/2})$. By the considering the same estimate for the same individual a time $Dk^{1/2}$ later (for large D), the individual must likely have moved during that time interval. In other words, there is likely to have been a wave of length $> k$ during this $O(k^{1/2})$ time interval.

In section 3.3 we will prove the bound in the following form. Let U_τ have uniform distribution on $\{1, 2, \ldots, \tau\}$,

PROPOSITION 14.

$$\limsup_{\tau\to\infty} \mathbb{P}\left(\max_{U_\tau < t \leq U_\tau + j} W(s) \leq k\right) \leq Bk^{1/2}/j, \quad j < k$$

for constant B not depending on j, k.

To translate this into the same format as Proposition 11, write

$$\rho_-(j) = \liminf_{\tau\to\infty} \tau^{-1}\mathbb{E}N_j(\tau)$$

where as before

$$N_j(\tau) = \sum_{t=1}^{\tau} \mathbb{1}_{\{W(t)>j\}}$$

is the number of waves of length $> j$ up to time τ.

COROLLARY 15. $\rho_-(k) \geq \frac{1}{2(1+Bk^{1/2})}$, *for the constant B in Proposition 14.*

Proof of Corollary 15. The event in Proposition 14 is the event $\{N_k(U_\tau + j) = N_k(U_\tau)\}$, so choosing $j = \lceil 2Bk^{1/2} \rceil$ we have

$$\liminf_{\tau\to\infty} \mathbb{P}(N_k(U_\tau + j) - N_k(U_\tau) \geq 1) \geq \tfrac{1}{2}$$

implying

$$\liminf_{\tau\to\infty} \mathbb{E}(N_k(U_\tau + j) - N_k(U_\tau)) \geq \tfrac{1}{2}.$$

For $1 \leq t \leq \tau + j$ we have $\mathbb{P}(U_\tau < t \leq U_\tau + j) \leq j/\tau$; applying this to the times t counted by $N_k(\tau + j)$ gives

$$\mathbb{E}(N_k(U_\tau + j) - N_k(U_\tau)) \leq \tfrac{j}{\tau}\mathbb{E}N_k(\tau + j).$$

So

$$\liminf_{\tau\to\infty} \tfrac{1}{\tau}\mathbb{E}N_k(\tau + j) \geq \tfrac{1}{2j}$$

establishing the Corollary. □

3.3. *Proof of Proposition 14.* In order to apply Lemma 13 we need to upper bound the probability of the complement of events $A_{s+1}(t_0)$ of the form (27). From the definition

$$A^c_{s+1}(t_0) = \{L_s(t_0) = \ell < L_{s+1}(t_0) \text{ for some } 0 \leq \ell < t_0\} \cup \{L_s(t_0) = t_0\}.$$

In order that event $\{L_s(t_0) = \ell < L_{s+1}(t_0)\}$ occurs it is necessary (but not sufficient – see remark below Lemma 16) that the wave at time $t_0 - \ell$ moves individual s but not individual $s + 1$, which is saying that $W(t_0 - \ell) = s - t_0 + \ell + 1$. And the event $\{L_s(t_0) = t_0\}$ can be rewritten as the event $\{M(t_0) \leq s\}$ for

$$M(t_0) := \min\{s : X^*_s(t_0) = X^*_s(0)\}.$$

So

$$\mathbb{P}(A^c_{s+1}(t_0)) \leq \mathbb{P}(M(t_0) \leq s) + \sum_{\ell=0}^{t_0-1} \mathbb{P}(W(t_0 - \ell) = s - t_0 + \ell + 1).$$

Setting $s = t_0 + i - 1$ for $i \geq 1$ and $\ell = t_0 - j$, this becomes

$$\mathbb{P}(A^c_{t_0+i}(t_0)) \leq \mathbb{P}(M(t_0) \leq t_0 + i - 1) + \sum_{j=1}^{t_0} \mathbb{P}(W(j) = t_0 + i - j).$$

Now note that if individual s is at the same position x at times 0 and t_0 we must have

$$c_- s \leq x \leq c^+(s - t_0)$$

implying that $s \geq \frac{t_0 c^+}{c^+ - c_-}$. So there is a deterministic bound $M(t) \geq \frac{t_0 c^+}{c^+ - c_-}$ which implies that, in the limit we will take at (29), the contribution from the term $\mathbb{P}(M(t_0) \leq t_0 + i - 1)$ is negligible, so we ignore that term in the next calculation. We now average over $1 \leq t_0 \leq \tau$:

$$\begin{aligned}
\tau^{-1} \sum_{t_0=1}^{\tau} \mathbb{P}(A^c_{t_0+i}(t_0)) &\leq \tau^{-1} \sum_{t_0=1}^{\tau} \sum_{j=1}^{t_0} \mathbb{P}(W(j) = t_0 + i - j) \\
&= \tau^{-1} \sum_{j=1}^{\tau} \mathbb{P}(i \leq W(j) \leq \tau + i - j) \\
&\leq \tau^{-1} \sum_{j=1}^{\tau} \mathbb{P}(W(j) \geq i)
\end{aligned}$$

and take limits

$$\limsup_{\tau\to\infty} \tau^{-1} \sum_{t_0=1}^{\tau} \mathbb{P}(A^c_{t_0+i}(t_0)) \leq \limsup_{\tau\to\infty} \tau^{-1} \sum_{j=1}^{\tau} \mathbb{P}(W(j) \geq i). \tag{29}$$

Proposition 11 bounds the right side as $O(i^{-1/2})$. Now we fix k, sum over $1 \leq i \leq k$ and apply Lemma 13 to deduce

LEMMA 16. *There exists a constant C such that, for each $k \geq 1$,*

$$\limsup_{\tau\to\infty} \mathbb{E}|X_k(U_\tau) - S_k(U_\tau)| \leq C k^{1/2}. \tag{30}$$

Recall that U_τ has uniform distribution on $\{1, 2, \ldots, \tau\}$, and that $S_k(t_0)$, and hence $S_k(U_\tau)$, is distributed as the sum of k i.i.d.(μ) random variables.

REMARK. Heuristics suggest that in fact $\mathbb{P}(A^c_{t_0+i}(t_0))$ decreases as order $i^{-1} \log i$. Our order $i^{-1/2}$ bound is crude because, in order for event $A_{s+1}(t_0)$ to occur, we need not only that $L_s(t_0) = \ell < L_{s+1}(t_0)$ for some $0 \leq \ell < t_0$, but also that no subsequent wave during time $(t_0 - \ell, t_0]$ moves individual $s+1$, and we ignored the latter condition. However a better estimate would not help our subsequent arguments, because we use (30) to deduce an $O(k^{1/2})$ bound on the spread of $X_k(U_\tau)$ from the same order bound on the spread of $S_k(U_\tau)$, and this deduced bound would not be improved by a better bound on the difference.

We can now continue with the argument outlined at the start of section 3.2. We already fixed k, and now fix $j < k$. Consider the individual of rank k at time t_0, and therefore of rank $k+j$ at time $t_0 - j$. We have

$$\mathbb{P}(\max_{t_0 - j < t \leq t_0} W(t) \leq k) \leq \mathbb{P}(X_k(t_0) = X_{k+j}(t_0 - j)) \tag{31}$$

because if there was no wave of length $> k$ during the time interval $(t_0 - j, t_0]$ then the individual under consideration does not move. Write

$$\begin{aligned} X_{k+j}(t_0 - j) - j - X_k(t_0) &= S_{k+j}(t_0 - j) - (k+j) \\ &- (S_k(t_0) - k) \\ &+ (X_{k+j}(t_0 - j) - S_{k+j}(t_0 - j)) \\ &- (X_k(t_0) - S_k(t_0)). \end{aligned}$$

Now $S_k(t_0)$ has mean k and variance $k\sigma^2$, and $S_{k+j}(t_0 - j)$ has mean $k+j$ and variance $(k+j)\sigma^2$, so by bounding the expectation of absolute value of each of the four terms in the sum above,

$$\mathbb{E}\left|X_{k+j}(t_0 - j) - j - X_k(t_0)\right| \leq$$

$$(k^{1/2} + (k+j)^{1/2})\sigma + \mathbb{E}\left|X_{k+j}(t_0 - j) - S_{k+j}(t_0 - j)\right| + \mathbb{E}\left|X_k(t_0) - S_k(t_0)\right|.$$

The limit assertion in (30) remains true if U_τ is replaced therein by $U_\tau + j$, and so substituting $U_\tau + j$ for t_0 in the inequality above and applying (30)

$$\limsup_{\tau \to \infty} \mathbb{E}\left|X_{k+j}(U_\tau) - j - X_k(U_\tau + j)\right| \leq 2(\sigma + C)(k+j)^{1/2}.$$

Then by Markov's inequality,

$$\limsup_{\tau \to \infty} \mathbb{P}(X_{k+j}(U_\tau) = X_k(U_\tau + j)) \leq \frac{2(\sigma + C)(k+j)^{1/2}}{j}.$$

Now substitute $U_\tau + j$ for t_0 in (31) and combine with the inequality above:

$$\limsup_{\tau \to \infty} \mathbb{P}\left(\max_{U_\tau < t \leq U_\tau + j} W(t) \leq k\right) \leq \frac{2(\sigma + C)(k+j)^{1/2}}{j}$$

establishing Proposition 14.

4. The CBM limit. Here we develop the idea, outlined in section 1.2, that the representation of the spatial queue via the process $G(t,x)$ in Figure 3, suitably rescaled, converges in some sense to CBM. Given the results and techniques from sections 2 and 3, the proof of Theorem 17 (section 4.1) is conceptually straightforward, and we will not provide all the details of checking various intuitively clear assertions, for instance those involving the topology of our space $\mathbb{F}$. A technical discussion of topologies in more general contexts of this kind can be found in [9], but here we set out a bare minimum required to formulate the result. Finally in section 4.2 we outline how the results concerning customers' space-time trajectories follow from the CBM limit.

Notation and statement of result.

- $C[1,\infty)$ is the space of continuous functions $f : [1,\infty) \to \mathbb{R}$, and d_0 is a metrization of the usual topology of uniform convergence on compact intervals.

- $\mathbf{f} = \{f \in \mathcal{I}(\mathbf{f})\}$ denotes a "coalescing family" of functions $f \in C[1,\infty)$ with some index set $\mathcal{I}(\mathbf{f})$, satisfying the conditions (i, ii) following.

(i) $\mathrm{init}(\mathbf{f}) := \{f(1),\ f \in \mathcal{I}(\mathbf{f})\}$ is a distinct locally finite set;
so if $f \in \mathcal{I}(\mathbf{f})$ is such that $f(1)$ is not maximal in $\mathrm{init}(\mathbf{f})$, then there is a "next" $f^* \in \mathcal{I}(\mathbf{f})$, the function such that $f^*(1)$ is the next largest in $\mathrm{init}(\mathbf{f})$. For such a pair (f, f^*)

(ii) $f^*(t) \geq f(t)\ \forall t \geq 1$, and there is a "coalescing time" $\gamma(f, f^*) < \infty$ such that $f^*(t) = f(t)\ \forall t \geq \gamma(f, f^*)$.
Note this implies that each pair (f_1, f_2) of functions in $\mathbf{f}$ will eventually coalesce at some time $\gamma(f_1, f_2)$.

- $\mathbb{F}$ is the space of such families $\mathbf{f}$, equipped with the "natural" topology in which $\mathbf{f}^{(n)} \to \mathbf{f}$ means the following:

for each finite interval $I \subset \mathbb{R}$ such that $\mathrm{init}(\mathbf{f})$ contains neither endpoint of I, for all large n there is an "onto" map $\iota^{(n)}$ from $\{f^{(n)} \in \mathbf{f}^{(n)} : f^{(n)}(1) \in I\}$ to $\{f \in \mathbf{f} : f(1) \in I\}$ such that, for all choices of $f^{(n)} \in \mathbf{f}^{(n)}$, we have that $f^{(n)} - \iota^{(n)}(f^{(n)})$ converges to the zero function in $C[1,\infty)$; and also there is convergence of the coalescing times, in that $\gamma^{(n)}(f_1^{(n)}, f_2^{(n)}) - \gamma(\iota^{(n)}(f_1^{(n)}),\ \iota^{(n)}(f_2^{(n)})) \to 0$.

(Note that this definition is designed to allow there to be two functions in $\mathbf{f}^{(n)}$ converging to the same function in $\mathbf{f}$, provided their coalescence time $\to 1$.)

- For $\mathbf{f} \in \mathbb{F}$ and $t_0 > 1$, write $\mathbf{f}_{[t_0]}$ for the time-shifted and rescaled family in which each $f \in \mathbf{f}$ is replaced by f' defined by

$$f'(t) = t_0^{-1/2} f(t_0 t),\ t \geq 1. \tag{32}$$

- Write $\mathbf{F}$ for a random family, that is a random element of $\mathbb{F}$, and $\mathbf{F}_{[t_0]}$ for the time-shifted and rescaled family.

- Write $\mathbf{CB}$ for the specific random family consisting of the standard CBM process observed from time 1 onwards. That is, the family of distinct functions $(t \to B(y,t), t \geq 1)$ as y varies. The scale-invariance property of standard CBM implies that $\mathbf{CB}_{[t_0]}$ has the same distribution as $\mathbf{CB}$.

Recall the discussion surrounding the Figure 3 graphic. When this graphic represents the spatial queue, the horizontal axis is "space" x and the vertical axis is "time" t. To compare the spatial queue process with CBM we need to reconsider the horizontal axis as "time" t and the vertical axis as "space" y. That is, we rewrite the function $G(t,x)$ at (6) as $\tilde{G}(y,t)$, so we have

$$\tilde{G}(y,t) = G(y,t) = y + F_y(t), \quad \text{where } F_y(t) = \max\{k : X_k(y) \leq t\}$$

where $y = 0, 1, 2, \ldots$ and $0 \leq t < \infty$. To study this process for large y we rescale as follows: for each n define

$$\widetilde{H}^{(n)}(y,t) = \frac{\tilde{G}(\sigma n^{1/2} y + U_{\tau_n}, nt) \ - \ U_{\tau_n}}{n^{1/2}}, \quad 1 \leq t < \infty. \tag{33}$$

This is defined for y such that $\sigma n^{1/2} y + U_{\tau_n} \in \mathbb{Z}^+$; also $\tau_n \to \infty$ is specified below and U_{τ_n} is uniform random on $\{1, \ldots, \tau_n\}$. We now define the random family $\mathbf{F}^{(n)}$ to consist of the distinct functions

$$t \to \widetilde{H}^{(n)}(y,t), \quad 1 \leq t < \infty$$

as y varies. This rescaling construction is illustrated in Figure 6.

We can now state the convergence theorem, in which we are considering $\mathbf{F}^{(n)}$ and $\mathbf{CB}$ as random elements of the space $\mathbb{F}$.

THEOREM 17. *If $\tau_n \to \infty$ sufficiently fast, then* $\mathbf{F}^{(n)} \xrightarrow{d} \mathbf{CB}$.

We conjecture this remains true without time-averaging, that is if we replace U_{τ_n} by τ_n in the definition of $\mathbf{F}^{(n)}$.

4.1. *Outline proof of Theorem 17.* **Step 1.** Propositions 11 and 14 refer to wave lengths in terms of rank j instead of position (which we are now calling t). But for any individual at any time these are related by $t/j \in [c_-, c^+]$, so the "order of magnitude" bounds in those Propositions remain true when we measure wave length by position. By choosing $\tau_n \to \infty$ sufficiently fast, those Propositions provide information about the point processes $(\mathrm{init}(\mathbf{F}^{(n)}), n \geq 1)$, as follows.

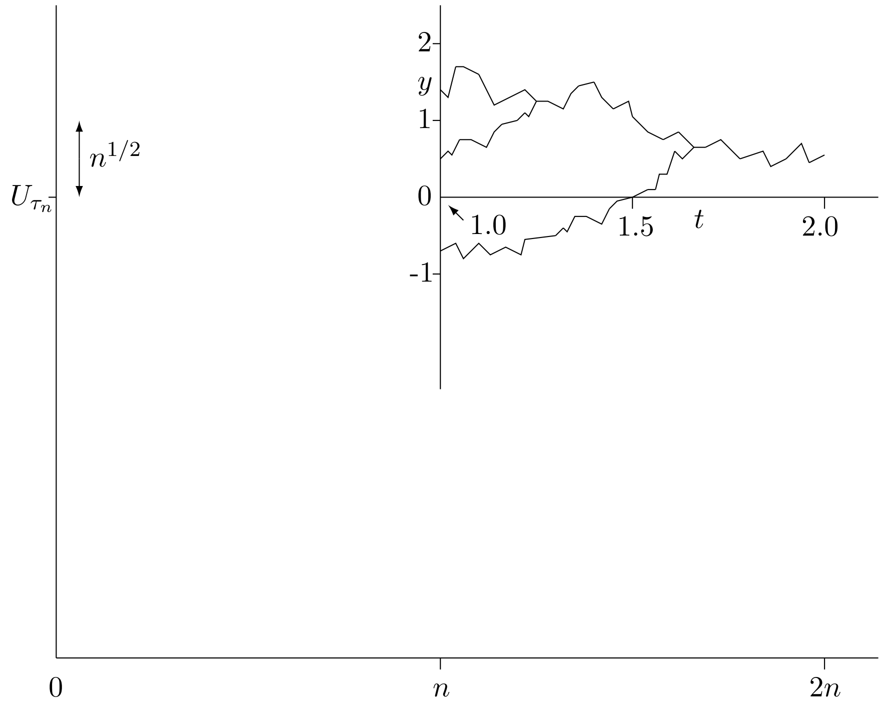

FIG 6. *The rescaling that defines $\mathbf{F}^{(n)}$ (small axes) in terms of $G(t,x)$ (large axes), assuming $\sigma = 1$.*

COROLLARY 18. *There exist constants β_1, β_2 and $\zeta_n \to -\infty$ such that for all sufficiently large n*

$$\mathbb{P}(\mathrm{init}(\mathbf{F}^{(n)}) \cap [a,b] = \emptyset) \le \beta_1/(b-a), \quad \zeta_n \le a < b < \infty. \tag{34}$$

$$\mathbb{E}|\mathrm{init}(\mathbf{F}^{(n)}) \cap [a,b]| \le \beta_2(b-a), \quad -\infty < a < b < \infty. \tag{35}$$

Step 2. Modify the usual topology on simple point processes on $\mathbb{R}$ to allow two distinct points in the sequence to converge to the same point in the limit. In this topology, (35) is a sufficient condition for tightness of the sequence $(\mathrm{init}(\mathbf{F}^{(n)}), n \ge 1)$, and so by passing to a subsequence we may assume

$$\mathrm{init}(\mathbf{F}^{(n)}) \stackrel{d}{\to} \text{ some } \tilde{\eta}^0 \tag{36}$$

where the limit point process $\tilde{\eta}^0$ inherits properties (34, 35):

$$\mathbb{P}(\tilde{\eta}^0 \cap [a,b] = \emptyset) \le \beta_1/(b-a), \ -\infty < a < b < \infty.$$

$$\mathbb{E}|\tilde{\eta}^0 \cap [a,b]| \le \beta_2(b-a), \quad -\infty < a < b < \infty. \tag{37}$$

Now $\tilde{\eta}^0$ has translation-invariant distribution because we use the uniform random variable U_{τ_n} with $\tau_n \to \infty$ to define $\mathbf{F}^{(n)}$. The central part of the proof of Theorem 17 is now to show

PROPOSITION 19. *If $\tau_n \to \infty$ sufficiently fast then*

$$\mathbf{F}^{(n)} \xrightarrow{d} \mathbf{F}^{(\infty)} := \ CBM(\tilde{\eta}^0, 0) \tag{38}$$

where the limit is the coalescing Brownian motion process started with particle positions distributed as $\tilde{\eta}^0$.

Granted (38), we can fix $t_0 > 0$ and apply our "time-shift and rescale" operation $\mathbf{f} \to \mathbf{f}_{[t_0]}$ at (32) to deduce

$$\mathbf{F}^{(n)}_{[t_0]} \xrightarrow{d} \mathbf{F}^{(\infty)}_{[t_0]} \text{ as } n \to \infty. \tag{39}$$

Now $\mathbf{F}^{(n)}$ depends on τ_n, so to make that dependence explicit let us rewrite (39) as

$$\mathbf{F}^{(n,\tau_n)}_{[t_0]} \xrightarrow{d} \mathbf{F}^{(\infty)}_{[t_0]} \text{ as } n \to \infty.$$

But $\mathbf{F}^{(n,\tau_n)}_{[t_0]}$ is just $\mathbf{F}^{(nt_0)}$ with a different shift; precisely

$$\mathbf{F}^{(n,\tau_n)}_{[t_0]} = \mathbf{F}^{(nt_0,\tau_{nt_0})}.$$

So we have

$$\mathbf{F}^{(nt_0,\tau_{nt_0})} \xrightarrow{d} \mathbf{F}^{(\infty)}_{[t_0]} \text{ as } n \to \infty.$$

But the scaling and embedding properties of CBM imply that

$$\mathbf{F}^{(\infty)}_{[t_0]} \xrightarrow{d} \mathbf{CB} \text{ as } t_0 \to \infty.$$

So by taking $t_0 = t_0(n) \to \infty$ sufficiently slowly and setting $m = nt_0(n)$,

$$\mathbf{F}^{(m,.\tau'_m)} \xrightarrow{d} \mathbf{CB} \text{ as } m \to \infty$$

for some $\tau'_m \to \infty$, and this remains true for any larger τ''_m. This is the assertion of Theorem 17.

Step 3. To start the proof of (38), fix Y_0 and T_0, and define a modified process $\text{modi}_{Y_0}(\mathbf{F}^{(\infty)})$ as CBM started with particles at positions $\tilde{\eta}^0 \cap [-Y_0, Y_0]$ only. Define $\text{modi}_{Y_0}(\mathbf{F}^{(n)})$ as follows. Consider the construction (1–3) of the spatial queue process. There is a smallest t (in the notation of (1–3)) such that the corresponding value y in $\mathrm{init}(\mathbf{F}^{(n)})$, that is

$$\sigma n^{1/2}(y + U_{\tau_n}) = t,$$

satisfies $y \geq -Y_0$. We now modify (in Step 4 we will argue that this modification has no effect in our asymptotic regime) the construction (1–3) of the spatial queue process by saying that, for this particular t, we set

$$X_i(t) = \xi_1(t) + \ldots + \xi_i(t),\ 1 \leq i < \infty.$$

That is, we replace all the inter-customer distances to the right of position n by an i.i.d. sequence. For subsequent times $t+1, t+2, \ldots$ we continue the inductive construction (1–3). The effect of this change on the behavior of the queue process to the left of position n, in particular the effect on the process $\mathrm{init}(\mathbf{F}^{(n)})$, is negligible for our purposes in this argument. Now by (37) there are a bounded (in expectation) number of functions in $\mathrm{modi}_{Y_0}(\mathbf{F}^{(n)})$; each function behaves essentially as the rescaled renewal process associated with i.i.d.(μ) summands, until the wave ends, which happens within $O(1)$ steps when the function approaches the previous function. We can apply the classical invariance principle for renewal processes ([2] Theorem 17.3) to show that (as $n \to \infty$) these functions behave as Brownian motion

This argument is sufficient to imply

$$\mathrm{modi}_{Y_0}(\mathbf{F}^{(n)}) \xrightarrow{d} \mathrm{modi}_{Y_0}(\mathbf{F}^{(\infty)}). \tag{40}$$

Step 4. To deduce (38) from (40) we need to show

$$\mathrm{modi}_{Y_0}(\mathbf{F}^{(\infty)}) \xrightarrow{d} \mathbf{F}^{(\infty)} \text{ as } Y_0 \to \infty \tag{41}$$

and the following analogous result for the sequence $(\mathbf{F}^{(n)})$. Take $0 < Y_1 < Y_0 < \infty$ and $T_0 > 0$. Define an event $A(n, Y_1, Y_0, T_0)$ as

> the random family $\mathrm{modi}_{Y_0}(\mathbf{F}^{(n)})$ and the random family $\mathbf{F}^{(n)}$ do not coincide as regards functions f with $f(0) \in [-Y_1, Y_1]$ considered only on $0 \leq t \leq T_0$.

We then need to show: for each Y_1, T_0

$$\lim_{Y_0 \to \infty} \limsup_{n \to \infty} \mathbb{P}(A(n, Y_1, Y_0, T_0)) = 0. \tag{42}$$

It is easy to prove (41) directly from basic properties of standard CBM. Fix Y_1, T_0. As in Figure 4, the particle positions of standard CBM at time T_0 determine a point process $(t_u, u \in \mathbb{Z})$ on $\mathbb{R}$, each time-T_0 particle being the coalescence of all the initially in the interval (t_{u-1}, t_u). So there are some random $-\infty < u_L < u_R < \infty$ such that $[-Y_1, Y_1] \subset [t_{u_L}, t_{u_R}]$, and then on the event $[t_{u_L}, t_{u_R}] \subset [-Y_0, Y_0]$ we have, using the natural coupling of CBM processes, that the behavior of the particles of CBM$(\tilde{\eta}^0, 0)$ started within $[-Y_1, Y_1]$ is unaffected by the behavior of the particles started outside $[-Y_0, Y_0]$. This is enough to prove (41).

Unfortunately the simple argument above does not work for the analogous result (42). But there is a more crude argument which does work; for notational simplicity we will say this argument in the context of CBM but it then readily extends to the family $\mathbf{F}^{(n)}$. We can construct CBM started with particles at positions $\tilde{\eta}^0$ by first taking independent Brownian motions over $0 \le t < \infty$ started at each point of $\tilde{\eta}^0$ (call these virtual paths). Then, when two particles meet, we implement the coalescent by saying that the future path follows the virtual path of the particle with lower starting position. In words, we want to show the following.

> (*) Fix T_0. Consider CBM started with particles at positions $\tilde{\eta}^0 \cap [0, \infty]$ only. Then, in the natural coupling (induced by the construction above) with CBM started with particles at all positions $\tilde{\eta}^0$, the two process coincide, over $0 \le t \le T_0$, for the particles started on some random interval $[Z, \infty)$.

The argument is illustrated in Figure 7. The top left panel indicates the "virtual paths" from which we can construct CBM started with particles at all positions $\tilde{\eta}^0$ on $(-\infty, \infty)$ – we show a finite interval and a sample of initial positions $\{a, b, \ldots, j\}$ in that interval. The top right panel shows the coalescing paths from these initial positions. However if we restricted the process to the half-line $[0, \infty)$ then, as shown in the bottom left panel, some of the coalescing paths may be affected, because (in the figure) the particle started at d no longer coalesces with the particle started at c. In the figure, and over the time interval shown, the affected paths are those of particles d, e, f, g, h but not subsequent particles.

We can upper bound the range of "affecteds" as follows. Consider again the "virtual paths", as shown in the top left panel, and draw path segments according to the following rule, illustrated in the bottom right panel. Start drawing the virtual path of the first particle (d, in the figure) in $[0, \infty)$; when that path meets another virtual path, also start drawing that virtual path too; whenever a drawn virtual path meets another virtual path, start drawing that path also. This produces a branching network, and at time T_0 the paths associated with this set of particles have been drawn. It is easy to check that for the lowest time-T_0 particle in the top right panel whose constituent time-0 particles are distinct from those in the branching network, (the particle arising from i and j in the figure), this time-T_0

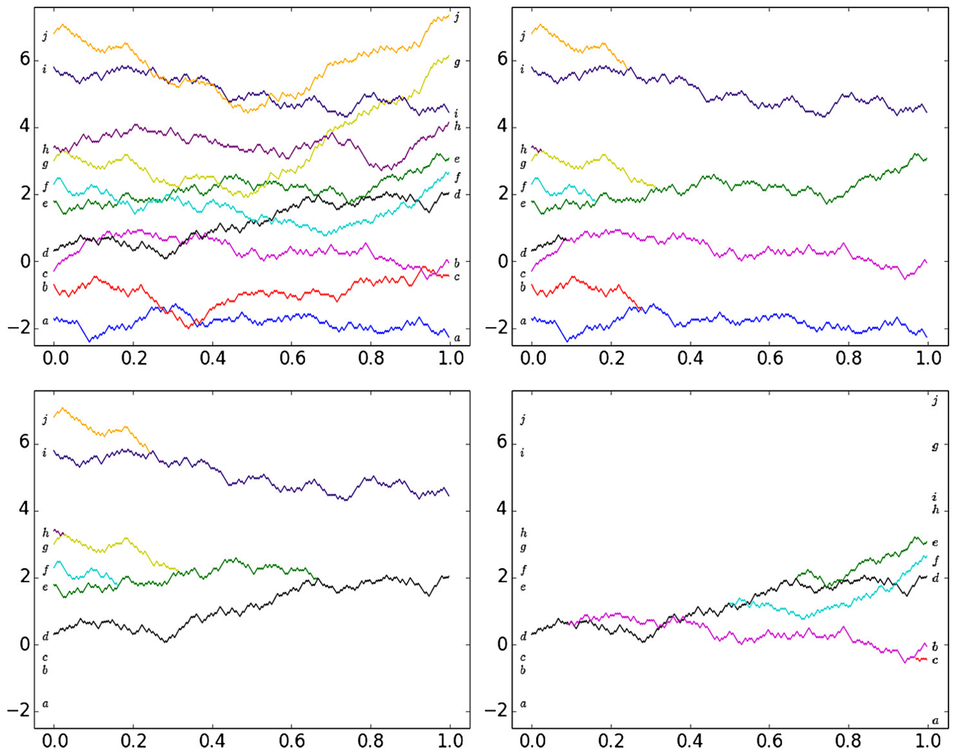

FIG 7. *The construction in Step 4.*

position is unaffected by the change from working on $(-\infty, \infty)$ to working on $[0, \infty)$.

One can now check that, for fixed T_0, only a finite mean number of particles will be affected (note this will include some particles started below 0, for instance b, c in the figure) and then that (*) holds with $\mathbb{E}Z < \infty$.

The point is that the calculation of a bound on $\mathbb{E}Z$ carries over to the rescaled family $(\mathbf{F}^{(n)})$ because the "virtual paths" are just the centered renewal processes, which behave like Brownian motion; and we can use (35) to show that only a bounded (in expectation) number of functions enter the picture.

4.2. *Consequences for wave lengths and space-time trajectories.* Here we show what Theorem 17 implies about the queue process, outlining proofs of Theorem 3 and Corollary 4 stated in section 1.3.

Wave lengths. First let us switch from measuring wave length by rank to measuring it by position. That is, the length of the wave that creates $\mathbf{X}(t)$

from $\mathbf{X}(t-1)$ was defined implicitly at (3) as the rank

$$W(t) := \min\{i : X_i(t) = X_{i-1}(t-1)\}$$

of the first unmoved customer, whereas now we want to consider it as the position of that customer:

$$L(t) := X_{W(t)}(t).$$

Directly from the section 1.1 discussion we see that the length of wave at time t is the position x where $G(t, \cdot)$ coalesces with $G(t-1, \cdot)$, that is where they first make the same jump. Define a point process $\zeta^{(n)}$ to be the set of points

$$\left\{\tfrac{t-U_{\tau_n}}{\sigma n^{1/2}} \ : \ L(t) > n\right\}.$$

Then, if $\tau_n \to \infty$ sufficiently fast, we have from Theorem 17

COROLLARY 20. *As $n \to \infty$ the point processes $\zeta^{(n)}$ converge in distribution to the spatial point process $(B(y,1), y \in \mathbb{R})$ of time-*1 *positions of particles in the standard CBM process.*

Now Lemma 16 tells us that at a typical time in the queue process, the rank-n customer is at position $n \pm O(n^{1/2})$. Combining that result with the fact that, between waves of that length, the customer's position does not change and their rank decreases only by $O(n^{1/2})$, it is not hard to deduce the same result for wave lengths measured by rank: if $\tau_n \to \infty$ sufficiently fast, then

COROLLARY 21. *As $n \to \infty$ the point processes $\left\{\frac{t-U_{\tau_n}}{\sigma n^{1/2}} \ : \ W(t) > n\right\}$ converge in distribution to the spatial point process $\{B(y,1), y \in \mathbb{R}\}$ of time-*1 *positions of particles in the standard CBM process.*

This easily implies the "rate" result stated as Theorem 3.

Space-time trajectories. As defined, $\mathbf{F}^{(n)}$ consists of the functions

$$t \to \widetilde{H}^{(n)}(y,t), \quad 1 \le t < \infty$$

as y varies. The same such function $f \in \mathbf{F}^{(n)}$ arises from all y in some interval, say the interval $[y_-(f), y_+(f)]$. Similarly, a function $f \in \mathbf{CB}$ is the function $t \to B(y,t),\ t \ge 1$ for all y in some interval, which again we may write as the interval $[y_-(f), y_+(f)]$. Now note that the definition (33) of $\widetilde{H}^{(n)}(y,t)$ makes sense for $0 \le t \le 1$ also. For any fixed small $\delta > 0$, Theorem 17 extends to showing convergence in distribution of $(t \to \widetilde{H}^{(n)}(y,t),\ \delta \le t < \infty)$

to standard CBM over the interval $[\delta, \infty)$. This procedure enables us, by taking $\delta \downarrow 0$ ultimately, to show that the Theorem 17 convergence $\mathbf{F}^{(n)} \xrightarrow{d} \mathbf{CB}$ extends to the "augmented" setting where the functions are marked by the intervals $[y_-(f), y_+(f)]$. (At the technical level, this argument allows us to work within the space $\mathbb{F}$ of locally finite collections of functions, instead of some more complicated space needed to handle standard CBM on $(0, \infty)$).

We can now add some details to the verbal argument in section 1.3 for Corollary 4. Regarding the first component, the central point is as follows. Take lines ℓ_n through $(1, 0)$ of slope $-n^{1/2}$. Write $(f_u^{(n)},\ u \in \mathbb{Z})$ for the distinct functions in $\mathbf{F}^{(n)}$ intercepted by line ℓ_n, at some points (x_u, s_u), and write $(y_-(f_u^{(n)}), y_+(f_u^{(n)}))$ for the interval of initial values of functions which coalesce with $f_u^{(n)}$ before x_u. These represent the normalized move times of an individual chosen as being at position n at time U_{τ_n} We know $\mathbf{F}^{(n)} \xrightarrow{d} \mathbf{CB}$ in the "augmented" sense above. This implies that the associated interval endpoints $(y_+(f_u^{(n)}), u \in \mathbb{Z})$ (that is, the normalized move times above) converge in distribution as point processes to $(w_u,\ u \in \mathbb{Z})$, the time-1 positions of particles in CBM. That is the assertion of Corollary 4, as regards the first component.

Regarding the second component, the verbal argument in section 1.3 can be formalized to show that for the process $(w_u,\ u \in \mathbb{Z})$ of positions of time-1 particles in CBM, the inter-particle distances are the rescaled limit of the numbers of customers who cross the starting position of the individual under consideration when that individual moves. To then derive the stated result, the central issue is to identify "numbers of customers who cross" with "distance the distinguished individual moves", which will follow from the lemma below. Consider a typical wave which extends past position n at some large time t. For each position x in the spatial interval $[n \pm Bn^{1/2}]$ there is some number $M^{(n)}(x)$ of customers who cross position x in the wave, and each individual ι in the spatial interval moves some distance $D^{(n)}(\iota)$. From the CBM limit we know these quantities are order $n^{1/2}$. It is enough to show that all these customers move the same distance (to first order), in the following sense.

Lemma 22. *There exist random variables $M^{(n)}$ such that, for each fixed B,*

$$\max_x |n^{-1/2} M^{(n)}(x) - M^{(n)}| \to_p 0,$$
$$\max_\iota |n^{-1/2} D^{(n)}(\iota) - M^{(n)}| \to_p 0 \text{ as } n \to \infty$$

where the maxima are over the spatial interval $[n \pm Bn^{1/2}]$.

Consider the inter-customer distances $(\xi_i(t-1), \xi_i(t), i \geq 1)$ immediately before and after the wave at time t. If, over the interval $[n \pm Bn^{1/2}]$ under consideration, these were i.i.d.(μ) then the assertion of the lemma would be clear, using (for instance) large deviation bounds for i.i.d. sums. Analyzing wave lengths in sections 2 and 3 was complicated because we did not know orders of magnitude. Now we know the power law $\mathbb{P}(W > w) \sim cw^{-1/2}$ for wave length, so conditional on a wave entering the interval $[n \pm Bn^{1/2}]$ the probability it goes only $o(n)$ further is $o(1)$. From the CBM limit we know that the first customer in the interval moves a distance $M^{(n)}$ of order $n^{1/2}$. Given that distance, the subsequent inter-customer distances $\xi_i(t)$ over the interval are i.i.d.(μ) conditioned on a event of probability $1-o(1)$. The same was true for the previous wave that created the inter-customer distances $\xi_i(t-1)$. So the conclusion of the lemma follows from the i.i.d. case.

5. Final remarks.

5.1. *An informal data set.* Spending 17 minutes in line at security at Oakland airport enabled the author to collect the data in Figure 8, whose points show position (of author) and time after each move, the time increments measured by watch and the position increments being estimated by eye. The approximate straight line reflects the fact that people are being served at an approximately constant rate. This is much too little data to

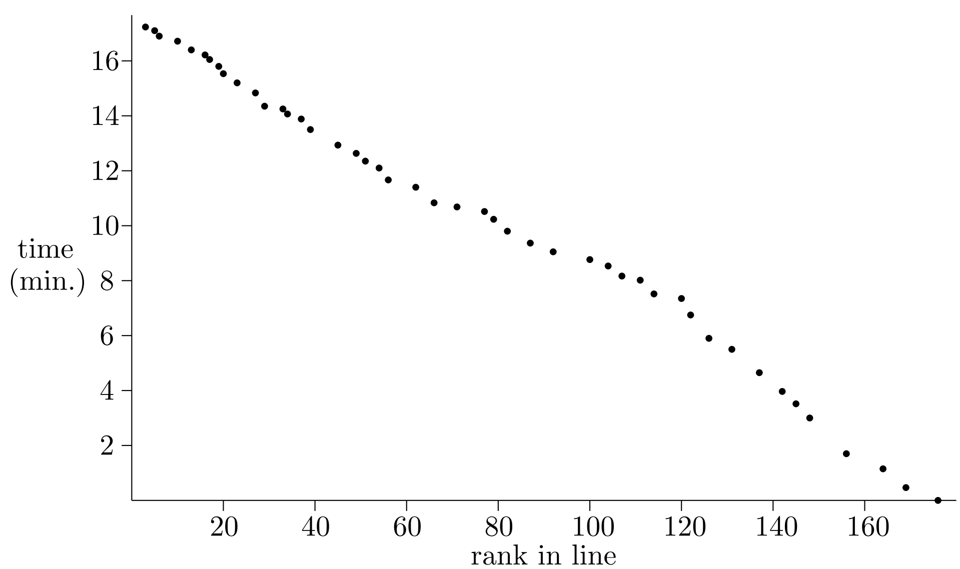

Fig 8. *Data: progress in a long queue.*

assess whether the "square root law" prediction of our model is realistic. But it does show that the underlying qualitative effect – that times between moves and distances moved increase with rank within queue – is real rather than some psychological effect. This appears in Figure 8 as the trend of growing gaps between points as rank in line increases.

5.2. *Stationarity.* As mentioned earlier, it is possible that continuing the argument for Lemma 12 would give a "coupling from the past" proof of convergence to a stationary distribution (Conjecture 1). Assuming existence of a stationary distribution, a stronger version of Lemma 12 would be

Open Problem 23. *Suppose μ has a density that is bounded below on $[c_-, c^+]$ and suppose the stationary distribution $\mathbf{X}(\infty)$ exists. Is it true that the joint distribution of all inter-customer distances $(X_i(\infty) - X_{i-1}(\infty), i \geq 1)$ is absolutely continuous with respect to infinite product measure μ^∞.*

Heuristics based on the the Kakutani equivalence theorem suggest the answer is "no".

Other questions involve mixing or "burn in": how long does it take, from an arbitrary start, for the distribution of the process restricted to the first k positions be close to the stationary distribution? From the initial configuration with all inter-customer distances equal to c_-, the time must be at least order k: is this the correct worst-case-start order?

5.3. *Finite queues.* A natural alternative to our infinite-queue setup would be to take Poisson arrivals with (subcritical) rate $\rho < 1$. Because such a $M/D/1$ queue is empty for a non-vanishing proportion of time, a stationary distribution clearly exists. However, to state the analog of Theorem 17 for the stationary process one still needs a double limit (rank $j \to \infty$ and traffic intensity $\rho \uparrow 1$); the infinite-queue setup should, if one can prove Conjecture 1, allow a single limit statement of form (8).

5.4. *Other possible proof techniques.* As mentioned earlier we suspect there may be simpler proofs using other techniques. Here are brief comments on possible proof techniques we did think about but did not use.

Forwards coupling. The classical notion of (forwards) Markov chain coupling – showing that versions of the process from two different initial states can be coupled so that they eventually coincide in some useful sense – does not seem to work for our process.

Feller property. A Feller process $(X(t))$ on a compact space always has at least one stationary distribution, by considering subsequential weak limits of $X(U_\tau)$. Our spatial queue process does have a compact state space $\mathbb{X}$; it is not precisely Feller, because of the discontinuity in the construction (1–3), but Feller continuity holds a.e. with respect to product uniform measure, so under mild conditions (e.g. that μ has a density component) the same argument applies to prove existence of at least one stationary distribution. However, such "soft" arguments cannot establish uniqueness of stationary distribution or convergence thereto. One could rephrase our limit results in terms of stationary processes with such subsequential stationary distributions, but that hardly seems a simpler reformulation.

CBM as a space-indexed Markov process. A conceptually different starting point for an analysis of our spatial queue model is to observe that the standard CBM process $(B_y(s), 0 \leq s < \infty, y \in \mathbb{R})$ with $B_y(0) = y$ of section 1.2, with "time s" and "space" y, can in fact be viewed in the opposite way. Define

$$X_y = (B_y(s) - y,\ 0 \leq s < \infty)$$

so that X_y takes values in the space $C_0(\mathbb{R}^+)$ of continuous functions f with $f(0) = 0$. Now the process $(X_y, -\infty < y < \infty)$ with "time" y is a continuous-time $C_0(\mathbb{R}^+)$-valued Markov process. Apparently CBM has not been studied explicitly in this way. In principle one could determine its generator and seek to apply general techniques for weak convergence of discrete-time Markov chains to continuous-time limits. But we have not attempted this approach.

5.5. *Generalizations of CBM.* Finally we mention that the extensions of CBM to the *Brownian web* and *Brownian net* have been shown to arise as scaling limits of various one-dimensional models [9]. But our queue model seems to be a novel addition to this collection of CBM-related models.

Acknowledgments. I thank Weijian Han and Dan Lanoue for earlier simulations of the process, a referee and Editors for suggestions for improving the exposition, and Jon Warren for the reference to [12, 11].

REFERENCES

[1] R. A. Arratia. *Coalescing Brownian motions on the line*. PhD thesis, University of Wisconsin–Madison, 1979. MR2630231

[2] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968. MR0233396

[3] M. Blank. Metric properties of discrete time exclusion type processes in continuum. *J. Stat. Phys.*, 140(1):170–197, 2010. MR2651444

[4] A. Borodin, P. L. Ferrari, M. Prähofer, and Tomohiro Sasamoto. Fluctuation properties of the TASEP with periodic initial configuration. *J. Stat. Phys.*, 129(5–6):1055–1080, 2007. MR2363389

[5] D. Bullock, R. Haseman, J. Wasson, and R. Spitler. Automated measurement of wait times at airport security. *Transportation Research Record: Journal of the Transportation Research Board*, 2177:60–68, 2010.

[6] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II*. Probability and its Applications (New York). Springer, New York, second edition, 2008. MR2371524

[7] L. Gray and D. Griffeath. The ergodic theory of traffic jams. *J. Statist. Phys.*, 105(3–4):413–452, 2001. MR1871652

[8] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley-Interscience, 1975. MR1383820

[9] E. Schertzer, R. Sun, and J. M. Swart. The Brownian web, the Brownian net, and their universality. *ArXiv 1506.00724*, June 2015. MR2712322

[10] B. Tóth and W. Werner. The true self-repelling motion. *Probab. Theory Related Fields*, 111(3):375–452, 1998. MR1640799

[11] R. Tribe, S. K. Yip, and O. Zaboronski. One dimensional annihilating and coalescing particle systems as extended Pfaffian point processes. *Electron. Commun. Probab.*, 17(40):7, 2012. MR2981896

[12] R. Tribe and O. Zaboronski. Pfaffian formulae for one dimensional coalescing and annihilating systems. *Electron. J. Probab.*, 16(76):2080–2103, 2011. MR2851057

David Aldous
Department of Statistics
367 Evans Hall # 3860
U.C. Berkeley CA 94720
USA
E-mail: aldous@stat.berkeley.edu