ALEA, Lat. Am. J. Probab. Math. Stat. **14**, 947–965 (2017)
DOI: 10.30757/ALEA.v14-41

Inventory Accumulation with Multiple Products

Cheng Mao and Tianyou Zhou

Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA, 02139, United States.
E-mail address: { maocheng , tianyou } @mit.edu

Abstract. Sheffield (2016) proposed an inventory accumulation model with two types of products which encodes the critical Fortuin-Kasteleyn model on a random planar map, and showed that a two-dimensional inventory accumulation trajectory in the discrete model scales to a correlated planar Brownian motion. In this work, we generalize the inventory model to k types of products for any integer $k \geq 2$, and prove that the corresponding trajectory scales to a k-dimensional Brownian motion with a certain covariance matrix. We also discuss implications of the scaling limit result in inventory theory and beyond.

1. Introduction

In a *last-in-first-out* (LIFO) inventory system (Cohen and Pekelman, 1977/78; Prastacos, 1979), most recently produced products are issued to fulfill customers' orders. The LIFO discipline is usually suitable for inventories of perishable products (Nahmias, 1982; Karaesmen et al., 2011), e.g., foods, blood units and chemical products. Moreover, the LIFO method of accounting for inventories is widely used to reduce tax liability in times of inflation (Cohen and Halperin, 1980). In some applications such as blood bank management, supplies and demands are stochastic and thus not under control of the inventory manager (Kaspi and Perry, 1983; Keilson and Seidmann, 1990; Parlar et al., 2011), so it is important to understand how inventories evolve with time. One effective way to study the long-term behavior of an inventory is through its continuous limit under the rescaling of time and the inventory size (Perry, 1997). See also Abate and Whitt (1997); Limic (2001) for examples of scaling limits of LIFO queues among the vast literature in queueing theory.

Received by the editors September 1st, 2016; accepted November 29th, 2017.
2010 *Mathematics Subject Classification.* Primary 60F17; secondary 60G50.
Key words and phrases. Inventory accumulation, first-in-last-out models, scaling limits, Brownian motion, random walks.

1.1. *A stochastic LIFO model, applications and a generalization.* In a recent paper (Sheffield, 2016), Sheffield introduced a simple stochastic inventory model at a LIFO retailer with two types of products, called *hamburgers* and *cheeseburgers*. In this inventory model, production of a hamburger, production of a cheeseburger, consumption of a hamburger, consumption of a cheeseburger and consumption of the *freshest* burger happen with respective probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1-p}{4}$, $\frac{1-p}{4}$ and $\frac{p}{2}$ at each discrete time point, where the freshest burger means the most recently produced burger regardless of type. It was established that the evolution of the two-burger inventory from time $-\infty$ to ∞ scales to a two-dimensional Brownian motion with covariance depending on p, which cleanly describes the long-term behavior of the inventory. An interesting phase transition happens at $p = 1/2$. In particular, when $p \geq 1/2$, the burger inventory remains balanced in the long run, i.e., the discrepancy between the two types of burgers remains small.

Balancing inventories of multiple products (resp. lengths of queues) is critical in inventory (resp. queueing) theory, because such a balance helps reduce backordering and holding costs (resp. service time and server load). Balancing policies have often been studied via scaling limits of the systems (Wein, 1992; Dimakis and Walrand, 2006; Mukherjee et al., 2016). For the model of Sheffield (2016), it may be easier to interpret the results if we reverse the time and switch the roles of production and consumption. Namely, in addition to production and consumption of two types of burgers, we assume that with probability $p/2$ the retailer produces a *flexible* burger which caters to both types of demands. The scaling limit result clearly remains true. If the retailer aims to balance the two types of demands, it suffices to tune up the proportion p of flexible burgers, and $p = 1/2$ is the critical threshold above which the two types are macroscopically balanced.

Along a different line of literature in operations research, from earlier results on auctions (Kruk, 2003) to a recent surge of research on limit order books (Cont and de Larrard, 2013, 2012; Guo et al., 2015; Paulsen, 2016), functional limit theorems have been crucial for studying price dynamics, especially in high frequency settings. Another way of understanding the stochastic inventory model of Sheffield (2016) is to view production of burgers as sell orders driving prices down, and view consumption of burgers as buy orders driving prices up in a limit order market. With probability p, buyers are willing to accept the lowest price regardless of the type of the product. The scaling limit result therefore characterizes the price dynamics of different products under this model. Particularly, the more similar the products are, the more likely buyers will accept whichever with the lowest price (corresponding to a higher p value in the model) and the higher correlation the prices of the products will have, as suggested by the covariance of the limiting Brownian motion. Therefore, beyond Markovian models of limit order books (Cont and de Larrard, 2013; Abergel and Jedidi, 2013; Gao and Deng, 2016), the non-Markovian inventory model yields interesting results on the correlation between price dynamics of different products.

As most inventory systems (or limit order books) consist of more than two types of products, a generalization to multiple types of products is desired to make the model of Sheffield (2016) more adaptive to further applications. This motivates us to study the stochastic LIFO inventory model with k types of products for any integer $k \geq 2$. In particular, we prove that the corresponding k-dimensional inventory trajectory scales to a k-dimensional Brownian motion and identify its

covariance matrix. A phase transition occurs at the critical probability $p = 1 - 1/k$ which generalizes the two-dimensional result.

On the technical side, the high-level strategy of our proof for the k-dimensional scaling limit result is the same as that for dimension two. In short, various renewal times and excursions of the discrete process between renewal times are introduced to describe the non-Markovian inventory trajectory. Integrability of these renewal times sits at the heart of the phase transition between the case where the well-balanced inventory scales to a standard Brownian motion, and the case where the imbalanced inventory scales to a Brownian motion with a nontrivial covariance structure.

It is worth noting that the k-dimensional scaling limit result is far from being obvious and many nontrivial adjustments need to be made provided the proof for dimension two, because interactions between multiple types of products are more complicated in higher dimensions. Originality is required for some proofs. For example, monotonicity properties of inventories used in Sheffield (2016, Section 3.4) do not hold in higher dimensions, so instead we make use of the property that if two inventory stacks are close in an appropriate sense, then they still stay close after adding the same product or order to each of them.

1.2. *Random planar maps and beyond.* Interestingly, the original motivation of Sheffield (2016) to study this inventory model was rather distant from inventory theory. *Planar maps* are connected planar graphs embedded into the two-dimensional sphere defined up to homeomorphisms of the sphere (Tutte, 1963), and the *critical Fortuin-Kasteleyn (FK) cluster model* is a statistical physics model defined on a random planar map (Fortuin and Kasteleyn, 1972; Grimmett, 2006). Perhaps surprisingly, the inventory model bijectively encodes a random planar map weighted by the partition function of the critical FK cluster model (Bernardi, 2008; Sheffield, 2016). In particular, the probability p in the inventory model is fundamental because it is connected to the parameter q of the FK model via the relation $q = 4p^2/(1-p)^2$. Thus the inventory model serves as a new tool to study scaling limits of the FK model. See also Le Gall and Miermont (2011, 2012) for recent developments on scaling limits of random planar maps.

More recently, efforts have been made to understand variations of the inventory accumulation model and their connections to models on random planar maps. Several conditional versions of the hamburger-cheeseburger model were studied in Gwynne et al. (2015); Berestycki et al. (2017); Gwynne and Sun (2017, 2015). For example, in Gwynne et al. (2015) orders after time 0 are conditioned to consume burgers produced after time 0, in which case the inventory trajectory scales to a Brownian motion conditioned to stay in the positive quadrant; in Gwynne and Sun (2015), the inventory is conditioned to be empty after a finite period, in which case the trajectory scales to a Brownian motion conditioned to return to the origin. Furthermore, the hamburger-cheeseburger model finds applications in the abelian sandpile model and the uniform spanning unicycle model (Sun and Wilson, 2016), and a generalized version of the inventory model with more rules of burger production and consumption was introduced to study scaling limits of the active spanning tree model (Gwynne et al., 2016).

So far, little research has been devoted to higher-dimensional analogues of random planar maps, partly due to the difficulty of enumeration and lack of bijective representations. See Benjamini and Curien (2011) for an interesting higher-dimensional result among the few. We hope that our generalized model and results can be used to construct potentially interesting higher-dimensional objects, possibly as follows.

When there are only two types of burgers, the number of hamburgers and the number of cheeseburgers after n steps can be interpreted as two walks on $\mathbb{Z}$. As explained in Sheffield (2016), each of these walks separately encodes a tree (via a standard bijection between walks and trees) along with a path tracing the boundary of the tree. Furthermore, one can form a larger graph by starting with these two trees and then adding an edge between a vertex on the first tree and a vertex on the second tree if those vertices are both visited at the same time by the traversing paths. This construction can be used to give a bijection between burger-order sequences and FK planar maps; see Sheffield (2016, Section 4) for more details.

It is straightforward to generalize this construction to our setting to obtain k trees along with extra edges joining vertices on different trees. We are not aware of any natural physical interpretation of the random graph obtained this way when $k > 2$, but we feel that this might be an interesting avenue for future research.

The rest of the paper is organized as follows. In Section 2, we describe the inventory model in detail and state the main scaling limit theorem. Section 3 is devoted to computing the covariance matrix of the limiting Brownian motion. We prove various technical estimates in Section 4 and 5, and finish the proof of the main theorem in Section 6.

2. Model setup and the main theorem

We consider a last-in-first-out retailer with k types of products, to which we refer as *burger 1*, ..., *burger k*. To adapt the two-dimensional model introduced in Sheffield (2016), we define an alphabet of symbols

$$\Theta = \{\textcircled{1}, \textcircled{2}, \ldots, \textcircled{k}, \boxed{1}, \boxed{2}, \ldots, \boxed{k}, \boxed{\mathrm{F}}\}$$

which represent the k types of burgers, the corresponding k types of orders each of which consumes the most recently produced burger of the same type, and the "flexible" order which consumes the most recently produced burger regardless of type in the remaining burger stack.

A word in the alphabet Θ is a concatenation of symbols in Θ that describes a series of production or consumption activities at the retailer. For example, if $W = \textcircled{2}\textcircled{3}\boxed{3}\textcircled{1}\boxed{2}\boxed{\mathrm{F}}$, then the word W describes the following sequence of activities: a burger 2 is produced, a burger 3 is produced, a burger 3 is ordered, a burger 1 is produced, a burger 2 is ordered and the freshest burger is ordered, which is burger 1 in this case.

To describe the evolution of burger inventory mathematically, we consider the collection $\mathcal{G}$ of (reduced) words in the alphabet Θ modulo the following relations

$$\textcircled{i}\boxed{i} = \textcircled{i}\boxed{\mathrm{F}} = \varnothing \quad \text{and} \quad \textcircled{i}\boxed{j} = \boxed{j}\textcircled{i} \tag{2.1}$$

where $1 \le i, j \le k$ and $i \ne j$. Intuitively, the first relation means that an order $\boxed{i}$ or $\boxed{\mathrm{F}}$ consumes a preceding burger $\textcircled{i}$, and the second means that we move an order

one position to the left if it does not consume the immediately preceding burger. For example,

$$W = ②③\boxed{3}①\boxed{2}\boxed{\text{F}} = ②①\boxed{2}\boxed{\text{F}} = ②\boxed{2}①\boxed{\text{F}} = \varnothing,$$

where $\boxed{3}$ consumes ③, $\boxed{2}$ consumes ② and $\boxed{\text{F}}$ consumes ①. By the same argument as in the proof of Sheffield (2016, Proposition 2.1), we see that $\mathcal{G}$ is a semigroup with $\varnothing$ as the identity and concatenation as the binary operation.

Let $X(n)$ be i.i.d. random variables indexed by $\mathbb{Z}$ (i.e. time), each of which takes its value in Θ with respective probabilities

$$\left\{\frac{1}{2k}, \frac{1}{2k}, \ldots, \frac{1}{2k}, \frac{1-p}{2k}, \frac{1-p}{2k}, \ldots, \frac{1-p}{2k}, \frac{p}{2}\right\}.$$

Let μ denote the corresponding probability measure on the space Ω of maps from $\mathbb{Z}$ to Θ. In this paper, we follow the convention that probabilities and expectations are with respect to μ unless otherwise mentioned. For $m \le n$, we write

$$X(m,n) := \overline{X(m)X(m+1)\cdots X(n)}$$

where $\overline{\cdot}$ means that a word is *reduced* modulo the relations (2.1). Then $X(m,n)$ describes the remaining orders and burgers (after all consumptions) between time m and time n at the retailer.

If a burger is added at time m and consumed at time n, we define $\phi(m) = n$ and $\phi(n) = m$. Otherwise, if a burger at m has no corresponding order, then $\phi(m) = \infty$, or if an order at n has no corresponding burger, then $\phi(n) = -\infty$. Proposition 2.2 in Sheffield (2016) remains valid in this k-burger setting:

Proposition 2.1. *It is μ-almost sure that for every $m \in \mathbb{Z}$, $\phi(m)$ is finite.*

Since a slight modification of the original proof will work here, we only describe the ideas. Let E_i be the event that every burger of type i is ultimately consumed. It can be shown that the union of E_i's has probability one, and since E_i's are translation-invariant, the zero-one law implies that each of them occurs almost surely. A similar argument works for orders, so each $X(m)$ has a correspondence, which is the statement of Proposition 2.1.

Hence for each $n \in \mathbb{Z}$, we can define a letter

$$Y(n) := \begin{cases} X(n) & X(n) \neq \boxed{\text{F}}, \\ \boxed{\text{i}} & X(n) = \boxed{\text{F}}, X(\phi(n)) = ⓘ. \end{cases}$$

Namely, $Y(n)$ is the same as $X(n)$ except that when $X(n) = \boxed{\text{F}}$, $Y(n)$ is the order type corresponding to the burger it actually consumes. Moreover, we define the *semi-infinite burger stack* $X(-\infty, n)$ to be the sequence of $X(m)$ where $m \le n$ and $\phi(m) > n$. It contains no orders almost surely since each order consumes an earlier burger at a finite time due to Proposition 2.1. It is almost surely infinite, because otherwise the number of burgers minus the number of orders in $X(-\infty, n)$ is a simple random walk in n and will visit -1 at a finite time almost surely, but an order added at or before that time will consume no burger which contradicts Proposition 2.1.

Next, we give definitions of several important discrete processes that will be shown to scale to Brownian motions.

Definition 2.2. For a word W in the alphabet Θ, we define $\mathcal{C}^i(W)$ to be the net burger *count* of type i, i.e., the number of $\textcircled{i}$ minus the number of $\boxed{i}$. Also, we define $\mathcal{C}(W)$ to be the total burger count, i.e.,

$$\mathcal{C}(W) := \sum_{i=1}^{k} \mathcal{C}^i(W).$$

If W has no $\boxed{F}$, then for $1 \le i \ne j \le k$, we define $\mathcal{D}^{ij}(W)$ to be the net *discrepancy* of burger i over burger j, i.e.,

$$\mathcal{D}^{ij}(W) := \mathcal{C}^i(W) - \mathcal{C}^j(W).$$

Definition 2.3. Given the infinite $X(n)$ sequence, let $\mathcal{C}_n^i$ be the integer-valued process defined by $\mathcal{C}_0^i = 0$ and $\mathcal{C}_n^i - \mathcal{C}_{n-1}^i = \mathcal{C}^i(Y(n))$ for all n[1]. Let $\mathcal{C}_n := \sum_{i=1}^k \mathcal{C}_n^i$ and $\mathcal{D}_n^{ij} := \mathcal{C}_n^i - \mathcal{C}_n^j$.

For any integer n, we define two vector-valued processes A_n and $\widetilde{A}_n$ by

$$A_n := (\mathcal{D}_n^{12}, \mathcal{D}_n^{23}, \dots, \mathcal{D}_n^{k-1,k}, \mathcal{C}_n) \quad \text{and} \quad \widetilde{A}_n := (\mathcal{C}_n^1, \mathcal{C}_n^2, \dots, \mathcal{C}_n^k).$$

We extend these definitions to real numbers by piecewise linear interpolation so that $t \mapsto A_t$ and $t \mapsto \widetilde{A}_t$ are infinite continuous paths.

When $n > 0$, we have $\mathcal{C}_n^i = \mathcal{C}^i(Y(1,n))$; when $n < 0$, we have $\mathcal{C}_n^i = \mathcal{C}^i(Y(n+1,0))$; similarly for $\mathcal{C}_n$ and $\mathcal{D}_n^{ij}$. As shorthand notation, we write

$$\mathcal{C}^i(m) = \mathcal{C}^i(Y(m)) \quad \text{and} \quad \mathcal{C}^i(m,n) = \mathcal{C}^i(Y(m,n))$$

for $m \le n$, and we let $\mathcal{C}(m), \mathcal{D}^{ij}(m), \mathcal{C}(m,n)$ and $\mathcal{D}^{ij}(m,n)$ be defined similarly.

Note that the two processes A_n and $\widetilde{A}_n$ actually code the same information about the evolution of the sequence $Y(n)$. Specifically, if we view A_n and $\widetilde{A}_n$ as column vectors, then it follows from Definition 2.3 that $A_n = M\widetilde{A}_n$ where M is a $k \times k$ invertible matrix defined by

$$M_{ij} = \begin{cases} 1 & i = j, 1 \le i \le k-1, \\ -1 & i+1 = j, 1 \le i \le k-1, \\ 1 & i = k, \\ 0 & \text{otherwise.} \end{cases}$$

It is more natural to describe the evolution of $Y(1,n)$ by $\widetilde{A}_n$ as its i-th coordinate corresponds to the burger count of type i. However, A_n gives another interesting perspective to view the stack $Y(1,n)$. Consider the line $\mathscr{L}$ through $(0,\dots,0)$ and $(1,\dots,1)$ in $\mathbb{R}^k$. Since $\mathcal{C}_n$ is a simple random walk along $\mathscr{L}$ and is independent of the other $k-1$ coordinates of A_n, we may view A_n as the Cartesian product of a one-dimensional simple random walk and an independent walk on the perpendicular $(k-1)$-dimensional hyperplane. The idea of separating the net burger count from the net burger discrepancies is inherited from the two-dimensional case.

With the linear relation established between A_n and $\widetilde{A}_n$, we are ready to state two equivalent versions of the main scaling limit theorem.

[1]In this work, we use subscripts n to index *cumulated* stochastic processes, such as $\mathcal{C}_n^i$, $\mathcal{D}_n^{ij}$ and A_n, while we use bracketed n to index random variables at discrete time points, such as $X(n)$, $Y(n)$ and $\mathcal{C}^i(n)$.

Theorem 2.4 (Main theorem, version 1). *As $\varepsilon \to 0$, the random variables $\varepsilon A_{t/\varepsilon^2}$ converge in law (with respect to the L^∞ metric on compact intervals) to*

$$(\mathbf{B}^1_{\alpha t}, B^2_t),$$

where $\mathbf{B}^1_t = (W^1_t, \dots, W^{k-1}_t)$ is a $(k-1)$-dimensional Brownian motion with covariance

$$\operatorname{Cov}(W^i_t, W^j_t) = \begin{cases} t & i = j, \\ -\frac{t}{2} & |i-j| = 1, \\ 0 & \text{otherwise}, \end{cases}$$

B^2_t is a standard one-dimensional Brownian motion independent of $\mathbf{B}^1_t$ and $\alpha := \max\{\frac{2}{k} - \frac{2p}{k-1}, 0\}$.

Theorem 2.5 (Main theorem, version 2). *As $\varepsilon \to 0$, the random variables $\varepsilon \widetilde{A}_{t/\varepsilon^2}$ converge in law (with respect to the L^∞ metric on compact intervals) to a k-dimensional Brownian motion*

$$\widetilde{\mathbf{B}}_t = (V^1_t, \dots, V^k_t)$$

with covariance

$$\operatorname{Cov}(V^i_t, V^j_t) = \begin{cases} (\frac{1}{k^2} - \frac{\alpha}{2k} + \frac{\alpha}{2})t & i = j, \\ (\frac{1}{k^2} - \frac{\alpha}{2k})t & i \neq j, \end{cases}$$

where $\alpha := \max\{\frac{2}{k} - \frac{2p}{k-1}, 0\}$.

It can be verified that $(\mathbf{B}^1_{\alpha t}, B^2_t) = M\widetilde{\mathbf{B}}_t$ in distribution, so it is not hard to see that the two theorems are indeed equivalent.

Theorem 2.4 is a direct generalization of Sheffield (2016, Theorem 2.5). We will focus on proving this version in later sections. We noted that $\mathcal{C}_n$ is a simple random walk independent of $\mathcal{D}^{ij}_n$, so it scales to B^2_t which is independent of $\mathbf{B}^1_t$ as in the theorem. Moreover, the value of α suggests that a phase transition happens at $p = 1 - \frac{1}{k}$, so that when p gets larger than this value, the process $\widetilde{A}_n$ looks like a 1-dimensional brownian motion when viewed from a large scale). It will be further explained in the next section.

To see that the limit in Theorem 2.5 is reasonable, we consider the special case $p = 0$, i.e., there are no "flexible" orders. In this case, $\widetilde{A}_n$ is a simple random walk on $\mathbb{Z}^k$, so we expect the limit to be a standard k-dimensional Brownian motion. Indeed, if $p = 0$, then $\alpha = 2/k$ and

$$\operatorname{Cov}(V^i_t, V^j_t) = \begin{cases} \frac{1}{k} & i = j, \\ 0 & i \neq j. \end{cases}$$

3. Computation of the covariance matrix and the critical value

In this section, we calculate the covariance matrix $[\operatorname{Cov}(\mathcal{D}^{i,i+1}_n, \mathcal{D}^{j,j+1}_n)]_{ij}$ where $1 \le i, j \le k-1$. It determines the value of α, the critical value of p and the covariance matrix of the limiting Brownian motion as in Theorem 2.4.

3.1. *First calculations.* Following the argument in Sheffield (2016, Section 3.1), we let J be the smallest positive integer for which the reduced word $X(-J,-1)$ contains exactly one burger (which is the rightmost burger in the semi-infinite stack $X(-\infty,-1)$). We use $|W|$ to denote the length of a word W and let $\chi = \chi(p) := \mathbb{E}[|X(-J,-1)|]$.

The orders in the reduced word $X(-J,-1)$ are of types different from the one burger in $X(-J,-1)$. In particular, we have that

$$|\mathcal{D}^{ij}(-J,-1)| \le |X(-J,-1)| = -\mathcal{C}(-J,-1)+2. \tag{3.1}$$

Since $\mathcal{C}(-n,-1)$ is a martingale in n, for a fixed n the optional stopping theorem applied to the stopping time $J\wedge n$ implies that

$$0 = \mathbb{E}[\mathcal{C}(-1,-1)] = \mathbb{E}[\mathcal{C}(-J,-1)\mathbb{1}_{J\le n}] + \mathbb{E}[\mathcal{C}(-n,-1)\mathbb{1}_{J>n}]. \tag{3.2}$$

In the case $J>n$, $\mathcal{C}(-n,-1)\le 0$, so $\mathbb{E}[\mathcal{C}(-J,-1)\mathbb{1}_{J\le n}]\ge 0$. Letting $n\to\infty$, we see that $\mathbb{E}[\mathcal{C}(-J,-1)]\ge 0$. On the other hand, $\mathbb{E}[\mathcal{C}(-J,-1)]\le 1$, so by (3.1),

$$\chi = \mathbb{E}[|X(-J,-1)|] \in [1,2]. \tag{3.3}$$

Note that $\chi=2$ if and only if $\mathbb{E}[\mathcal{C}(-J,-1)]=0$. Therefore, as $n\to\infty$ in (3.2), we deduce that

$$\chi = 2 \text{ if and only if } \lim_{n\to\infty} \mathbb{E}[\mathcal{C}(-n,-1)\mathbb{1}_{J>n}] = 0. \tag{3.4}$$

By (3.1), (3.3) and symmetry, $\mathbb{E}[\mathcal{D}^{ij}(-J,-1)]$ exists and equals zero. Moreover, since $|\mathcal{D}^{ij}(-n,-1)| \le -\mathcal{C}(-n,-1)$ for $n<J$, by (3.4),

$$\chi = 2 \text{ implies that } \lim_{n\to\infty} \mathbb{E}[|\mathcal{D}^{ij}(-n,-1)|\mathbb{1}_{J>n}] = 0. \tag{3.5}$$

It turns out that there is a dichotomy between $\chi=2$ and $1\le\chi<2$, which corresponds exactly to the phase transition at $p=1-1/k$. In this section, we focus on the case $\chi=2$ and show that $p\le 1-1/k$. We leave the case $1\le\chi<2$ to the following sections.

3.2. *Computation of* $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)]$. In preparation for computing the covariance $\mathrm{Cov}(\mathcal{D}_n^{ij},\mathcal{D}_n^{lm}) = \mathbb{E}[\mathcal{D}_n^{ij}\mathcal{D}_n^{lm}]$ for any $i\ne j$ and $l\ne m$, we first calculate $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)]$. Note that on the event $X(0)\ne\boxed{\mathrm{F}}$, $\mathcal{D}^{ij}(0)$ is determined by $X(0)$ independently of $\mathcal{D}^{lm}(-J,-1)$, so by symmetry between i and j we have that

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)\,|\,X(0)\ne\boxed{\mathrm{F}}] = 0.$$

In the sequel, we split the analysis on the event $X(0)=\boxed{\mathrm{F}}$ into three cases.

First, suppose that the indices i,j,l and m are distinct. On the event that $X(0)=\boxed{\mathrm{F}}$ and $Y(0)\ne\boxed{\mathrm{i}}$ or $\boxed{\mathrm{j}}$, we have $\mathcal{D}^{ij}(0)=0$, so

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)\,|\,X(0)=\boxed{\mathrm{F}},\ Y(0)\ne\boxed{\mathrm{i}}\text{ or }\boxed{\mathrm{j}}] = 0.$$

On the event that $X(0)=\boxed{\mathrm{F}}$ and $Y(0)=\boxed{\mathrm{i}}$ or $\boxed{\mathrm{j}}$, we have that $\mathcal{D}^{ij}(0)=-1$ or 1, and thus by symmetry between l and m,

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)\,|\,X(0)=\boxed{\mathrm{F}},\ Y(0)=\boxed{\mathrm{i}}\text{ or }\boxed{\mathrm{j}}] = 0.$$

Therefore, we conclude that

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)] = 0. \tag{3.6}$$

Next, we evaluate $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{ij}(-J,-1)]$ for $i\ne j$. On the event $X(0)=\boxed{\mathrm{F}}$, we have $\phi(0)=-J$. Suppose that $Y(0)=\boxed{\mathrm{i}}$. Then for any $l\ne i$, $\mathcal{D}^{il}(0)=-1$, and for

any $l, m \neq i$, $\mathcal{D}^{lm}(0) = 0$. Because $X(-J,-1)$ contains a burger i and (possibly) orders of types other than i, it follows that

$$\begin{aligned}|X(-J,-1)| + k - 2 &= \sum_{l\neq i} \mathcal{D}^{il}(-J,-1) \\ &= -\sum_{l\neq i} \mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1) = -\frac{1}{2}\sum_{l\neq m} \mathcal{D}^{lm}(0)\mathcal{D}^{lm}(-J,-1),\end{aligned}$$

where in the last sum indices l and m range over $\{1,\dots,k\}$. The above equation was proved on the event $X(0) = \boxed{\text{F}}$ and $Y(0) = \boxed{\text{i}}$, yet the quantity does not depend on i. Hence taking the expectation yields that

$$\begin{aligned}\chi + k - 2 &= \mathbb{E}\big[|X(-J,-1)| \,\big|\, X(0) = \boxed{\text{F}}\big] + k - 2 \\ &= -\frac{1}{2}\sum_{l\neq m} \mathbb{E}[\mathcal{D}^{lm}(0)\mathcal{D}^{lm}(-J,-1) \,|\, X(0) = \boxed{\text{F}}] \\ &= -\frac{k(k-1)}{2}\mathbb{E}[\mathcal{D}^{lm}(0)\mathcal{D}^{lm}(-J,-1) \,|\, X(0) = \boxed{\text{F}}], \qquad (3.7)\end{aligned}$$

where the first equality holds because $|X(-J,-1)|$ is independent of $X(0)$. Together with the case $X(0) \neq \boxed{\text{F}}$, the equation (3.7) (with dummy indices (l,m) replaced by (i,j)) implies that for any $i \neq j$,

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{ij}(-J,-1)] = -\frac{p(\chi+k-2)}{k(k-1)}, \qquad (3.8)$$

since $X(0) = \boxed{\text{F}}$ with probability $p/2$.

It remains to compute $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)]$ for distinct i, j and l. On the event $X(0) = \boxed{\text{F}}$ and $Y(0) \neq \boxed{\text{i}}$ or $\boxed{\text{j}}$, we have $\mathcal{D}^{ij}(0) = 0$, so $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)] = 0$. On the event $X(0) = \boxed{\text{F}}$ and $Y(0) = \boxed{\text{j}}$, we have that $\mathcal{D}^{ij}(0) = 1$ and thus $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)] = 0$. Finally, on the event $X(0) = \boxed{\text{F}}$ and $Y(0) = \boxed{\text{i}}$, we observe that $\mathcal{D}^{ij}(0) = \mathcal{D}^{il}(0) = -1$, so $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)] = \mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)]$. Summarizing the cases above, we obtain that

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)] = \mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)\mathbb{1}_{X(0)=\boxed{\text{F}},Y(0)=\boxed{\text{i}}}]. \qquad (3.9)$$

Since $\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1) = \mathcal{D}^{li}(0)\mathcal{D}^{li}(-J,-1)$ and $\mathcal{D}^{il}(0) = 0$ if $Y(0) \neq \boxed{\text{i}}$ or $\boxed{\text{l}}$,

$$\begin{aligned}\mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)\mathbb{1}_{X(0)=\boxed{\text{F}},Y(0)=\boxed{\text{i}}}] &= \mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)\mathbb{1}_{X(0)=\boxed{\text{F}},Y(0)=\boxed{\text{l}}}] \\ &= \frac{1}{2}\mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)\mathbb{1}_{X(0)=\boxed{\text{F}}}] \\ &= \frac{1}{2}\mathbb{E}[\mathcal{D}^{il}(0)\mathcal{D}^{il}(-J,-1)].\end{aligned}$$

Together with (3.9) and (3.8), this implies that

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{il}(-J,-1)] = \frac{1}{2}\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{ij}(-J,-1)] = -\frac{p(\chi+k-2)}{2k(k-1)}. \qquad (3.10)$$

3.3. *The covariance matrix and the phase transition.* Conditional on the event $J < n$, $\mathcal{D}^{lm}(-n,-J-1)$ is independent of $\mathcal{D}^{ij}(0)$ because even if $X(0)$ were $\boxed{\text{F}}$ it would

consume a burger after time $-J$. Therefore we have that $\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-n,-J-1)\mathbb{1}_{J<n}]=0$ and it is not hard to see that

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-n,-1)]=\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)\mathbb{1}_{J\le n}]+\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-n,-1)\mathbb{1}_{J>n}] \tag{3.11}$$

where the rightmost term tends to zero as $n\to\infty$ if $\chi=2$ because of (3.5). Therefore, summarizing (3.6), (3.8) and (3.10), we see that for $i\neq j$, $l\neq m$,

$$\chi=2 \text{ implies } \lim_{n\to\infty}\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-n,-1)]=\begin{cases}-\frac{p}{k-1} & i=l, j=m,\\ -\frac{p}{2(k-1)} & i=l, j\neq m,\\ 0 & i,j,l,m \text{ distinct.}\end{cases} \tag{3.12}$$

Moreover, $\mathcal{D}^{ij}(0)^2=1$ if $Y(0)$ is of type i or j, and $\mathcal{D}^{ij}(0)\mathcal{D}^{il}(0)=1$ if $Y(0)$ is of type i, so

$$\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(0)]=\begin{cases}\frac{2}{k} & i=l, j=m,\\ \frac{1}{k} & i=l, j\neq m,\\ 0 & i,j,l,m \text{ distinct.}\end{cases} \tag{3.13}$$

Now we evaluate $\operatorname{Cov}(\mathcal{D}_n^{ij},\mathcal{D}_n^{lm})=\mathbb{E}[\mathcal{D}_n^{ij}\mathcal{D}_n^{lm}]$. Using

$$\mathcal{D}_r^{ij}\mathcal{D}_r^{lm}=\mathcal{D}^{ij}(r)\mathcal{D}^{lm}(r)+\mathcal{D}^{ij}(r)\mathcal{D}_{r-1}^{lm}+\mathcal{D}_{r-1}^{ij}\mathcal{D}^{lm}(r)+\mathcal{D}_{r-1}^{ij}\mathcal{D}_{r-1}^{lm}$$

recursively for $2\le r\le n$ and applying the translation invariance of the law of Y_m, we deduce that when $\chi=2$,

$$\begin{aligned}&\operatorname{Cov}(\mathcal{D}_n^{ij},\mathcal{D}_n^{lm})\\ &=\sum_{r=1}^{n}\mathbb{E}[\mathcal{D}^{ij}(r)\mathcal{D}^{lm}(r)]+\sum_{r=2}^{n}\mathbb{E}[\mathcal{D}^{ij}(r)\mathcal{D}_{r-1}^{lm}+\mathcal{D}_{r-1}^{ij}\mathcal{D}^{lm}(r)]\\ &=n\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(0)]+\sum_{r=2}^{n}\big(\mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(1-r,-1)]+\mathbb{E}[\mathcal{D}^{lm}(0)\mathcal{D}^{ij}(1-r,-1)]\big)\\ &=\begin{cases}\frac{2n}{k}-\frac{2np}{k-1}+o(n) & i=l, j=m,\\ \frac{n}{k}-\frac{np}{k-1}+o(n) & i=l, j\neq m,\\ o(n) & i,j,l,m \text{ distinct},\end{cases}\end{aligned} \tag{3.14}$$

where the last equation follows from (3.12) and (3.13).

For $i=l$ and $j=m$, the variance is nonnegative, so

$$\chi=2 \text{ implies } p\le 1-\frac{1}{k}. \tag{3.15}$$

We remark that (3.14) and (3.15) suggest that the phase transition happens at the critical value $p=1-\frac{1}{k}$. Let $\alpha=\max\{\frac{2}{k}-\frac{2p}{k-1},0\}$. When $\chi=2$ and $p\le 1-\frac{1}{k}$, it follows immediately from (3.14) that

$$\operatorname{Cov}(\mathcal{D}_n^{i,i+1},\mathcal{D}_n^{j,j+1})=\begin{cases}\alpha n+o(n) & i=j,\\ -\frac{\alpha n}{2}+o(n) & |i-j|=1,\\ o(n) & \text{otherwise.}\end{cases} \tag{3.16}$$

This explains why the limiting Brownian motion should have the covariance matrix as in Theorem 2.4. In the following sections, we will take care of the case $\chi<2$ and prove that the convergence indeed happens.

4. Excursion words revisited

This section generalizes the discussion of excursion words in Sheffield (2016, Section 3.3) to the k-burger case. The proof structure and most arguments are largely based on those in the original paper. Since adaptation is required throughout the proof, we include most details for completeness.

First, we quote two results (Sheffield, 2016, Lemma 3.3 and 3.4) directly:

Lemma 4.1. *Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. random variables on some measure space and ψ a measurable function on that space such that $\mathbb{E}[\psi(Z_1)] < \infty$. Let T be stopping time of the process $Z_1, Z_2, \ldots$ and $\mathbb{E}[T] < \infty$. Then $\mathbb{E}[\sum_{j=1}^{T} \psi(Z_j)] < \infty$.*

Lemma 4.2. *Let $Z_1, Z_2, \ldots$ be i.i.d. random variables on some measure space and let $\mathcal{Z}_n$ be a non-negative integer-valued process adapted to the filtration of the Z_n (i.e., each $\mathcal{Z}_n$ is a function of $Z_1, Z_2, \ldots, Z_n$) that has the following properties:*

(1) Bounded initial expectation: $\mathbb{E}[\mathcal{Z}_1] < \infty$.
(2) Positive chance to hit zero when close to zero: *For each $k > 0$ there exists a positive chance p_k such that conditioned on any choice of $Z_1, Z_2, \ldots, Z_n$ for which $\mathcal{Z}_n = k$, the conditional probability that $\mathcal{Z}_{n+1} = 0$ is at least p_k.*
(3) Uniformly negative drift when far from zero: *There exist positive constants C and c such that if we condition on any choice of $Z_1, Z_2, \ldots, Z_n$ for which $\mathcal{Z}_n \geq C$, the conditional expectation of $\mathcal{Z}_{n+1} - \mathcal{Z}_n$ is less than $-c$.*
(4) Bounded expectation when near zero: *There further exists a constant b such that if we condition on any choice of $Z_1, Z_2, \ldots, Z_n$ for which $\mathcal{Z}_n < C$, then the conditional expectation of $\mathcal{Z}_{n+1}$ is less than b.*

Then $\mathbb{E}[\min\{n : \mathcal{Z}_n = 0\}] < \infty$.

Let E denote the reduced word $X(1, K-1)$ where K is the smallest integer such that $\mathcal{C}_K < 0$, and call E an *excursion word*. Next we introduce a few quantities representing successive excursion words and their boundaries.

Definition 4.3. Let $V_0 := X(0)$ and $K_0 := 0$. For any positive integer i, let V_i be the symbol corresponding to the i-th record minimum of $\mathcal{C}_n$, counting forward from zero. More formally, with K_{i-1} determined, we let $V_i := X(K_i)$ where K_i is the smallest integer larger than K_{i-1} such that $\mathcal{C}_{K_i} < \mathcal{C}_{K_{i-1}}$. Similarly, for any negative integer i, let V_i be the $-i$-th record minimum of $\mathcal{C}_n$, counting backward from zero. Moreover, let E_i be the reduced word in between V_{i-1} and V_i, i.e.,

$$E_i := \overline{X(K_{i-1}+1)\cdots X(K_i-1)}.$$

Note that in particular $K = K_1$ and $E = E_1$. It is easy to check that E almost surely contains no $\boxed{\text{F}}$ symbols and there are always as many burgers as orders in the word E. In addition, E_i's and E are i.i.d. excursion words. The following lemma parallel to Sheffield (2016, Lemma 3.5) also holds:

Lemma 4.4. *If p is such that $\chi < 2$, then the expected word length $\mathbb{E}[|E|]$ is finite, and hence the expected number of symbols in E of each type in $\{\textcircled{1}, \ldots, \textcircled{k}, \boxed{1}, \ldots, \boxed{\text{k}}\}$ is $\mathbb{E}[|E|]/(2k)$.*

Since E is balanced between burgers and orders, the second statement follows from the first immediately by symmetry. For the first statement, it suffices to prove that the expected number of burgers in E_{-1} is finite, since E and E_{-1} have the same distribution. The original proof still works, so we omit it. The idea is to introduce

a sequence of record times similar to the quantity J defined in Section 3.1, and use them to construct a martingale running backward in time. Then applying the optional stopping theorem with an appropriate choice of the stopping time leads to the conclusion.

Next, we consider the following sequences:

(1) *m-th empty order stack*: let O_m be the m-th smallest value of $j \geq 0$ with the property that $X(-j, 0)$ has an empty order stack.
(2) *m-th empty burger stack*: B_m is the m-th smallest value of $j \geq 1$ with the property that $X(1, j)$ has an empty burger stack.
(3) *m-th left record minimum*: $L_m = L_m^0$ is the smallest value of $j \geq 0$ such that $\mathcal{C}(-j, 0) = m$. Thus, $X(-L_m, 0) = \overline{V_{-m}E_{-m}\dots V_{-1}E_{-1}}$.
(4) *m-th left minimum with no orders of type $1, 2, \dots, i$*: for $1 \leq i \leq k$, L_m^i is the m-th smallest value of $j \geq 0$ with the property that $j = L_{m'}$ for some m' and $X(-j, 0)$ has no orders of type $1, 2, \dots, i$.

We observe that all these record sequences have the property that the words between two consecutive records are i.i.d.. Moreover, for $1 \leq i \leq k$, each L_m^i is equal to $L_{m'}^{i-1}$ for some m' by definition. Thus we can write each $X(-L_m^i, -L_{m-1}^i - 1)$ as a product of consecutive words of the form $X(-L_{m'}^{i-1}, -L_{m'-1}^{i-1} - 1)$. We have the following lemma:

Lemma 4.5. *Consider the following statements:*

(1) $\mathbb{E}[|E|] < \infty$;
(2) $\mathbb{E}[|X(-L_1^i, 0)|] < \infty$ *where* $0 \leq i \leq k$;
(3) $\mathbb{E}[|X(-O_1, 0)|] < \infty$;
(4) $\mathbb{E}[|X(1, B_1)|] < \infty$.

We have that (1), (2) and (3) are equivalent, and (4) implies (1).

Proof: *(1) implies (2):* Note that for $i = 0$, $L_1^0 = L_1$ and $X(-L_1^0, 0) = \overline{V_{-1}E_{-1}}$. Since E_{-1} and E have the same law, (2) follows immediate from (1) when $i = 0$. To prove (2) for $1 \leq i \leq k$, we use induction.

Assume (2) holds for $i - 1$. Let $H(m)$ be the number of orders of type i in $X(-L_m^{i-1}, 0)$. If we can apply Lemma 4.2 with $Z_m = X(-L_m^{i-1}, -L_{m-1}^{i-1} - 1)$ and $\mathcal{Z}_m = H(m)$, then $\mathbb{E}[\min\{m : H(m) = 0\}] < \infty$. That means the expected number of $X(-L_m^{i-1}, -L_{m-1}^{i-1} - 1)$ concatenated to produce $X(-L_1^i, 0)$ is finite. Since $X(-L_m^{i-1}, -L_{m-1}^{i-1} - 1)$ are identically distributed as $X(-L_1^{i-1}, 0)$ which has finite expected length by inductive hypothesis, Lemma 4.1 implies that $X(-L_1^i, 0)$ also has finite expected length.

Therefore it remains to check the four assumptions of Lemma 4.2. It is easy to see that Assumption (1), (2) and (4) follow from the construction of the sequence and the inductive hypothesis, so we focus on the negative drift assumption. For any $m > 1$,

$$H(m) = \max\{H(m-1) - h_m, 0\} + o_m,$$

where h_m is the number of burger i in $X(-L_m^{i-1}, -L_{m-1}^{i-1} - 1)$ and o_m is the number of order i in it. The expected number of burger i equals the expected number of order i in E_{-m} by Lemma 4.4, while the expected number of burger i in V_{-m} is $1/k$, which has no orders. Hence $\mathbb{E}[h_m] \geq \mathbb{E}[o_m] + 1/k$ since $X(-L_m^{i-1}, -L_{m-1}^{i-1} - 1)$ is a concatenation of at least one $\overline{V_{-m'}E_{-m'}}$. Note that

$$H(m) - H(m-1) = o_m - h_m + (h_m - H(m-1))\mathbb{1}_{\{H(m-1)-h_m<0\}}$$

and $\mathbb{E}[(h_m - j)\mathbb{1}_{h_m>j}] \le \mathbb{E}[h_m \mathbb{1}_{h_m>j}] \to 0$ as $j \to \infty$ by assumption. Thus there is $C > 0$ such that $\mathbb{E}[H(m) - H(m-1) \mid H(m-1) = j] \le -1/(2k)$ for $j > C$, so the negative drift assumption is verified.

(2) implies (3): By definition, $X(-O_1, 0)$ corresponds to the first time that the stack contains only burgers, while $X(-L_1^k, 0)$ corresponds to the first time that the stack contains only burgers and increases in length, it follows easily that $|X(-O_1, 0)| \le |X(-L_1^k, 0)|$, so the expectation is finite.

(3) implies (1): The number of burgers in $X(-O_1, 0)$ is at least the number of burgers in E_{-1}, which accounts for half of its length, so $\mathbb{E}[|E_{-1}|] < \infty$. Thus the same holds for E.

(4) implies (1): Analogous to the previous implication. □

The next lemma on the asymptotic fractions of burgers and orders is key to the proof of the main theorem.

Lemma 4.6. *If $\mathbb{E}[|E|] < \infty$, then as $n \to \infty$ the fraction of* ⓘ *symbols among the rightmost n elements of $X(-\infty, 0)$ tends to $1/k$ almost surely for any i. Also, as $n \to \infty$ the fraction of* i *or* F *symbols among the leftmost n elements of $X(1, \infty)$ tends to some positive constant almost surely.*

On the other hand if $\mathbb{E}[|E|] = \infty$, then as $n \to \infty$ the fraction of F *symbols among the leftmost n elements of $X(1, \infty)$ tends to zero almost surely.*

Proof: If $\mathbb{E}[|E|] < \infty$, then the words $X(-O_m, -O_{m-1} - 1)$ are i.i.d. with finite expectations by Lemma 4.5. Hence $X(-\infty, 0)$ is a concatenation of i.i.d. words $X(-O_m, -O_{m-1} - 1)$. The law of large numbers implies that the number of each type of burgers in $X(-O_m, 0)$ is given by $Cm + o(m)$ almost surely for some constant C. By symmetry, these constants are all equal to $\mathbb{E}[|X(-O_1, 0)|]/k$. The first statement then follows, and the second is proved analogously.

For the last statement, we note that $X(1, \infty)$ is an i.i.d. concatenation of burger-free words $X(B_{m-1}+1, B_m)$, and an F symbol can be added only when the burger stack is empty. Hence the number of F symbols in $X(1, B_m)$ grows like a constant times m. If $\mathbb{E}[|E|] = \infty$, Lemma 4.5 implies that $\mathbb{E}[|X(1, B_1)|] = \infty$. Thus the number of orders in $X(1, B_m)$ grows faster than any constant multiple of m almost surely, so the fraction of F symbols tends to zero almost surely. □

5. Bounded increments and tail estimates

We fix a semi-infinite stack $S_0 = X(-\infty, 0)$ and let $X(1), X(2), \ldots$ be chosen according to μ. An analogy of Sheffield (2016, Lemma 3.10) still holds in this case, but it requires a different proof as we will see.

Lemma 5.1. *For $N > 0$, $\mathbb{E}[\mathcal{D}_N^{ij} | X(l) : 1 \le l \le n]$ and $\mathbb{E}[\mathcal{D}_N^{ij} | X(l) : 1 \le l \le n, \mathcal{C}_l : l \le N]$ are both martingales in n with increments of magnitude at most two.*

Instead of monotonicity properties of stacks used in Sheffield (2016) which do not generalize to higher dimensions, we introduce the notion of *neighbor stacks* which allows us to prove a similar result.

Definition 5.2. Two semi-infinite stacks S_0 and S_1 are called neighbors if S_1 can be achieved from S_0 by removing an arbitrary burger from S_0, or vice versa.

For example, $S_0 = \cdots ②①①③②②③$ and $S_1 = \cdots ②①①②②③$ are neighbors, because one can get S_1 from S_0 by removing the fourth burger from the right.

Lemma 5.3. *If S_0 and S_1 are neighbors, then for any word W, $\overline{S_0W}$ and $\overline{S_1W}$ are still neighbors.*

Proof: Assume that we get S_1 from S_0 by deleting a ⓙ. By induction, we may also assume that W contains a single element.

If W is a burger, then for $\sigma = 1, 2$, $\overline{S_\sigma W}$ is achieved by adding W onto S_σ. If $W = \boxed{\text{F}}$, then $\overline{S_\sigma W}$ is achieved by deleting the rightmost burger from S_σ. If $W = \boxed{\text{i}}$, then $\overline{S_\sigma W}$ is achieved by deleting the rightmost ⓘ from S_σ. Hence in these three cases, it is easily seen that the resulting two stacks are still neighbors.

If $W = \boxed{\text{j}}$ and there is a ⓙ in S_0 to the right of the ⓙ which we deleted to get S_1, then $\overline{S_\sigma W}$ is achieved by deleting the rightmost ⓙ from S_σ. Hence the resulting two stacks are neighbors. Otherwise, the ⓙ deleted to get S_1 is the rightmost ⓙ in S_0, so $\overline{S_0W} = S_1$. Hence $\overline{S_0W}$ and $\overline{S_1W}$ are neighbors. □

Proof of Lemma 5.1: Since the two conditional expectations are clearly martingales in n, we only need to prove that the increments are bounded. To this end, it suffices to show that changing $X(l)$ for a single $1 \le l \le N$ only changes $\mathcal{D}_N^{ij}$ by at most two.

Suppose that $X(l)$ is changed to $X(l)'$. Here we make the convention that a product of words is always reduced. It is easy to see that $X(-\infty, l)$ and $X(-\infty, l-1)X(l)'$ have a common neighbor $X(-\infty, l-1)$. Lemma 5.3 then implies that $X(-\infty, N)$ and $X(-\infty, l-1)X(l)'X(l+1, N)$ have a common neighbor $X(-\infty, l-1)X(l+1, N)$. Since the ij-discrepancy differs by at most one between neighbors, we see that $\mathcal{D}_N^{ij}$ changes by at most two if we change a single $X(l)$. □

The following tail estimates are adapted from Sheffield (2016, Lemma 3.12 and 3.13).

Lemma 5.4. *Fix any $p \in [0, 1]$ and a semi-infinite stack $S_0 = X(-\infty, 0)$. There exist positive constants C_1 and C_2 such that for any choice of S_0, $a > 0$, $n > 1$ and any i, j,*

$$\mathbb{P}(\max_{1\le l\le n} |\mathcal{C}_l| > a\sqrt{n}) \le C_1 e^{-C_2 a} \quad \text{and} \quad \mathbb{P}(\max_{1\le l\le n} |\mathcal{D}_l^{ij}| > a\sqrt{n}) \le C_1 e^{-C_2 a}.$$

The original proof carries over almost verbatim. The idea is that Lemma 5.1 gives bounded increments of the martingales, so we can apply a pre-established tail estimate of martingales with bounded increments. We remark that it is an important technique to estimate the tails of martingales with bounded jumps. See Dembo (1996) for more interesting results.

Lemma 5.5. *Fix any $p \in [0, 1]$. There exist positive constants C_1 and C_2 such that for any $a \ge k \log k$ and $n > 1$,*

$$\mathbb{P}(|X(1, n)| > a\sqrt{n}) \le C_1 e^{-C_2 a/k}.$$

Proof: Let the semi-infinite stack S_0 be rotating among ⓙ, …, ⓚ. According to Lemma 5.4, we have the following control on the fluctuations of $\mathcal{C}_l$ and $\mathcal{D}_l^{ij}$:

$$\max_{1\le l\le n} \{|\mathcal{C}_l| \vee \max_{1\le i,j\le k} |\mathcal{D}_l^{ij}|\} \le a\sqrt{n}/(4k-1)$$

with probability at least $1 - C_1 k^2 e^{-2C_2 a/k}$ where C_1 and C_2 are positive constants. Note that $C_1 k^2 e^{-2C_2 a/k} \le C_1 e^{-C_2 a/k}$ when $a \ge k \log k$ and $C_2 \ge 2$.

Conditional on this high probability event, we claim that no burger in S_0 except the rightmost $a\sqrt{n}(2k-1)/(4k-1)$ burgers will be consumed in the first n steps. Assume the opposite. If the first such burger is consumed at step l and is an (m), then at this moment all burgers to the right are of types different from (m). Since $\mathcal{C}_l \ge -a\sqrt{n}/(4k-1)$, there are at least $a\sqrt{n}(2k-2)/(4k-1)$ burgers above the (m). Among them there are at least $2a\sqrt{n}/(4k-1)$ burgers of some type $m' \ne m$. Hence $|\mathcal{D}_l^{mm'}| > a\sqrt{n}/(4k-1)$, which is a contradiction.

It follows from the claim that there are at most $a\sqrt{n}(2k-1)/(4k-1)$ orders in $X(1,n)$. Since $\mathcal{C}_l$ fluctuates by at most $a\sqrt{n}/(4k-1)$, there are at most $2ka\sqrt{n}/(4k-1)$ burgers in $X(1,n)$. Therefore, $|X(1,n)| \le a\sqrt{n}$. □

6. Proof of the main theorem

The proof parallels that in Sheffield (2016, Section 3.5 and 3.6).

6.1. *The case* $\chi < 2$. In this subsection, we will resolve the remaining case from Section 3, i.e., the case $\chi < 2$. We will use the results from Section 4 and 5 to prove that when $\chi < 2$, the scaling limit of A_n on a compact interval has the law of a one-dimensional Brownian motion. This means that the total burger count $\mathcal{C}_n$ dominates. As we remarked after the statement of Theorem 2.4, $\mathcal{C}_n$ is a simple random walk and thus scales to a Brownian motion, so it suffices to show that $\mathcal{D}_n^{ij}$ scales to 0 in law on compact intervals.

In addition to the statement above, we will show that $\chi < 2$ implies that $p > 1 - 1/k$. Together with (3.15), this gives the dichotomy mentioned in Section 3.1, namely,

$$\chi < 2 \iff p > 1 - 1/k \quad \text{and} \quad \chi = 2 \iff p \le 1 - 1/k. \tag{6.1}$$

Thus this subsection proves Theorem 2.4 in the case $p > 1 - 1/k$. We divide the proof into three lemmas.

Lemma 6.1. *If* $\mathbb{E}[|E|] < \infty$ *(which holds when* $\chi < 2$*), then* $\mathrm{Var}[\mathcal{D}_n^{ij}] = o(n)$ *for all pairs* (i,j).

Proof: First, we prove that the random variables $n^{-1/2}\mathcal{D}_n^{ij}$ converge to 0 in probability. To do this, we consider the following events:

(1) $|X(1,n)| < a\sqrt{n}$;
(2) The top $2ka\sqrt{n}$ burgers in stack $X(-\infty, 0)$ are well balanced across all burger types with error at most $\varepsilon\sqrt{n}$, i.e., the number of burgers of any type is between $(2a-\varepsilon)\sqrt{n}$ and $(2a+\varepsilon)\sqrt{n}$;
(3) The top $n + 2ka\sqrt{n}$ burgers in the stack $X(-\infty, n)$ are well balanced across all burger types with error at most $\varepsilon\sqrt{n}$.

We assert that if all three events happen, then $|n^{-1/2}\mathcal{D}_n^{ij}| < 4\varepsilon$. First, (1) and (2) together imply that all the orders in $X(1,n)$ are fulfilled by the top $2ka\sqrt{n}$ burgers in $X(-\infty, 0)$, so the burgers below height $-2ka\sqrt{n}$ in $X(-\infty, 0)$ are not affected by $X(1,n)$. Hence the stacks $X(-\infty, 0)$ and $X(-\infty, n)$ are identical below height $-2ka\sqrt{n}$. On the other hand, $|X(1,n)| < a\sqrt{n}$ implies that $|\mathcal{C}_n| < a\sqrt{n}$, so the number of burgers in $X(-\infty, n)$ above height $-2ka\sqrt{n}$ is at least $(2k-1)a\sqrt{n}$. By

(2) and (3), the discrepancies between pairs of burger types above height $-2ka\sqrt{n}$ are less than $2\varepsilon\sqrt{n}$ for both stacks, so $|\mathcal{D}_n^{ij}|$ is at most $4\varepsilon\sqrt{n}$, as desired.

Next, we observe that all three events happen with arbitrarily high probability if we choose a and n to be sufficiently large. For any fixed $\varepsilon > 0$, we first choose a (depending on k) large enough so that (1) happens with high probability using Lemma 5.5. Then by Lemma 4.6, we choose n large enough so that (2) and (3) happen with high probability. Thus we conclude that $\lim_{n\to\infty} \mathbb{P}[|n^{-1/2}\mathcal{D}_n^{ij}| > \varepsilon] = 0$ for all $\varepsilon > 0$, i.e., $n^{-1/2}\mathcal{D}_n^{ij}$ converge to 0 in probability.

It remains to check that $\mathrm{Var}[n^{-1/2}\mathcal{D}_n^{ij}] = \mathbb{E}[n^{-1}(\mathcal{D}_n^{ij})^2]$ tends to 0 as $n \to \infty$. This follows from the fact that $n^{-1}(\mathcal{D}_n^{ij})^2$ tends to 0 in probability together with the uniform bounds on the tails given by Lemma 5.4. □

The following two lemmas are proved in exactly the same way as Sheffield (2016, Lemma 3.15 and 3.16), so we omit the proofs.

Lemma 6.2. *If* $\mathrm{Var}[\mathcal{D}_n^{ij}] = o(n)$*, then* $n^{-1/2}\max\{|\mathcal{D}_l^{ij}| : 1 \le l \le nt\}$ *converges to zero in probability as* $n \to \infty$ *for any fixed* $t > 0$.

The trick of the proof is to first divide the time interval into small subintervals, then observe the convergence at the end points, and finally use approximation to complete the proof. Note that by Lemma 6.2, we immediately obtain that A_n converges in law to a one-dimensional Brownian motion on compact intervals.

Lemma 6.3. *If* $\chi < 2$ *and* $\mathrm{Var}[\mathcal{D}_n^{ij}] = o(n)$*, then*

$$\lim_{n\to\infty} \mathbb{E}[|\mathcal{D}^{ij}(-n,-1)|\mathbb{1}_{J>n}] = 0.$$

Interested readers may refer to the proof in the original paper which involves introducing new measures via Radon-Nikodym derivatives and recentering the sequence. The original proof also uses the fact that one-dimensional random walk conditioned to stay positive scales to a three-dimensional Bessel process, which is explained in Pitman (1975).

Letting $n \to \infty$ in (3.11) and using Lemma 6.3 and (3.8), we deduce that

$$\lim_{n\to\infty} \mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-n,-1)] = \mathbb{E}[\mathcal{D}^{ij}(0)\mathcal{D}^{lm}(-J,-1)] = -\frac{p(\chi+k-2)}{k(k-1)}.$$

Following the same computation as in (3.14), we obtain that

$$\mathrm{Var}(\mathcal{D}_n^{ij}) = \frac{2n}{k} - \frac{2np(\chi+k-2)}{k(k-1)} + o(n).$$

By Lemma 6.1, we must have $\frac{2n}{k} = \frac{2np(\chi+k-2)}{k(k-1)}$, i.e., $p = \frac{k-1}{\chi+k-2}$. Hence $\chi < 2$ implies that $p > 1 - 1/k$, which gives us the promised dichotomy (6.1).

6.2. *The case* $\chi = 2$. It finally remains to prove the main theorem in the case $\chi = 2$. First, if $p = 1 - 1/k$, then $\mathrm{Var}[\mathcal{D}_n^{ij}] = o(n)$ by (3.14), so the convergence follows from our argument in Section 6.1.

Next, we may assume $p < 1-1/k$, so that $\mathrm{Var}[\mathcal{D}_n^{ij}] \neq o(n)$. By the contrapositive of Lemma 6.1, we must have $\mathbb{E}[|E|] = \infty$. Then we can apply the second part of Lemma 4.6, which asserts that the number of $\boxed{\mathrm{F}}$ symbols in $X(1,n)$ is small relative to the total number of orders in $X(1,n)$ as n gets large. To be more precise, the number of $\boxed{\mathrm{F}}$ in $X(1,\lfloor tn\rfloor)$ is $o(\sqrt{n})$ with probability tending to one as $n \to \infty$

by Lemma 4.6 and Lemma 5.5. Therefore, for $t_1 + t_2 = t_3$, the laws of $A_{\lfloor t_1 n \rfloor}$ and $A_{\lfloor t_2 n \rfloor}$ add to the law of $A_{\lfloor (t_1+t_2)n \rfloor}$ up to an error of $o(\sqrt{n})$ with high probability.

On the other hand, since the variances of $n^{-1/2}A_{tn}$ converge to constants as $n \to \infty$ for any fixed t, the random variables $n^{-1/2}A_{tn}$ converge subsequentially in law to a limit. Moreover, if we choose a finite collection of t values, namely $0 < t_1 < t_2 < \cdots < t_m < \infty$, the joint law of

$$\left(n^{-1/2}A_{\lfloor t_1 n \rfloor}, n^{-1/2}A_{\lfloor t_2 n \rfloor}, \ldots, n^{-1/2}A_{\lfloor t_m n \rfloor}\right)$$

also converges subsequentially to a limiting law.

Now we combine the two observations above. We have that the law of $n^{-1/2}A_{\lfloor tn \rfloor}$ is equal to the law of the sum of l independent copies of $n^{-1/2}A_{\lfloor tn/l \rfloor}$ plus a term which is $o(1)$ with high probability (since there is a multiplicative factor $n^{-1/2}$). Hence, the subsequential weak limit of $n^{-1/2}A_{\lfloor tn \rfloor}$ must equal the sum of l i.i.d. random variables. In particular, since l is arbitrary, the limiting law has to be infinitely divisible. Note that the process $n^{-1/2}A_{\lfloor tn \rfloor}$ is almost surely continuous in t, so we conclude that the subsequential limit discussed above has to be a Gaussian with mean zero. We refer to Bertoin (1996) for more background on infinitely divisible processes, Lévy processes and Gaussian processes.

The covariance matrix of $n^{-1/2}A_n$ is already given by our computation in Section 3, and Lemma 5.4 guarantees that $n^{-1/2}A_{\lfloor tn \rfloor}$ are tight, so the subsequential limit has the correct covariance matrix. We conclude that the limit indeed has the Gaussian distribution given in Theorem 2.4. Moreover, our argument implies that any subsequence of $n^{-1/2}A_{tn}$ has a further subsequence converging in law to this Gaussian distribution, so the whole sequence converges to this law.

The same is true if we choose a finite collection of t_i's, so the finite-dimensional joint law of

$$\left(n^{-1/2}A_{\lfloor t_1 n \rfloor}, n^{-1/2}A_{\lfloor t_2 n \rfloor}, \ldots, n^{-1/2}A_{\lfloor t_m n \rfloor}\right)$$

converges to a limiting law, which is exactly the law of $(\mathbf{W}_{t_1}, \mathbf{W}_{t_2}, \ldots, \mathbf{W}_{t_m})$, where $\mathbf{W}_t$ is the k-dimensional Brownian motion $(\mathbf{B}^1_{\alpha t}, B^2_t)$ described in Theorem 2.4.

The transition from a discrete collection of t_i's to a compact interval follows similarly as in the proof of Lemma 6.2. As the maximum gap between t_i's gets smaller, the probability that (the norm of) the fluctuation in some interval $[t_i, t_{i+1}]$ exceeds ε tends to zero as $n \to \infty$ for both $n^{-1/2}A_{\lfloor tn \rfloor}$ and $\mathbf{W}_t$ where $t \in [0, t_m]$. Hence the two processes are uniformly close on the interval $[0, t_m]$ with probability tending to one as $n \to \infty$. Therefore, Theorem 2.4 is fully proved.

Acknowledgements

We thank Scott Sheffield and Xin Sun for suggesting this project and commenting on earlier versions of the paper. This work could not be completed without their helpful ideas and suggestions. We thank the 2014 UROP+ program at Massachusetts Institute of Technology during which part of this work was completed, and thank Pavel Etingof and Slava Gerovitch for directing the program.

References

J. Abate and W. Whitt. Limits and approximations for the $M/G/1$ LIFO waiting-time distribution. *Oper. Res. Lett.* **20** (5), 199–206 (1997). MR1455070.

F. Abergel and A. Jedidi. A mathematical approach to order book modeling. *Int. J. Theor. Appl. Finance* **16** (5), 1350025, 40 (2013). MR3085985.

I. Benjamini and N. Curien. On limits of graphs sphere packed in Euclidean space and applications. *European J. Combin.* **32** (7), 975–984 (2011). MR2825530.

N. Berestycki, B. Laslier and G. Ray. Critical exponents on Fortuin-Kasteleyn weighted planar maps. *Comm. Math. Phys.* **355** (2), 427–462 (2017). MR3681382.

O. Bernardi. Tutte polynomial, subgraphs, orientations and sandpile model: new connections via embeddings. *Electron. J. Combin.* **15** (1), Research Paper 109, 53 (2008). MR2438581.

J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge (1996). ISBN 0-521-56243-0. MR1406564.

M. A. Cohen and R. Halperin. Optimal inventory order policy for a firm using the LIFO inventory costing method. *Journal of Accounting Research* **18** (2), 375–389 (1980).

M. A. Cohen and D. Pekelman. LIFO inventory systems. *Management Sci.* **24** (11), 1150–1162 (1977/78). MR521504.

R. Cont and A. de Larrard. Order book dynamics in liquid markets: limit theorems and diffusion approximations (2012). DOI: 10.2139/ssrn.1757861.

R. Cont and A. de Larrard. Price dynamics in a Markovian limit order market. *SIAM J. Financial Math.* **4** (1), 1–25 (2013). MR3032934.

A. Dembo. Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab.* **1**, no. 3, 11–17 (1996). MR1386290.

A. Dimakis and J. Walrand. Sufficient conditions for stability of longest-queue-first scheduling: second-order properties using fluid limits. *Adv. in Appl. Probab.* **38** (2), 505–521 (2006). MR2264955.

C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica* **57**, 536–564 (1972). MR0359655.

X. Gao and S. J. Deng. Hydrodynamic limit of order-book dynamics. *Probability in the Engineering and Informational Sciences* pages 1–30 (2016). DOI: 10.1017/S0269964816000413.

G. Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (2006). ISBN 978-3-540-32890-2; 3-540-32890-4. MR2243761.

X. Guo, Z. Ruan and L. Zhu. Dynamics of order positions and related queues in a limit order book (2015). DOI: 10.2139/ssrn.2607702.

E. Gwynne, A. Kassel, J. Miller and D. B. Wilson. Active spanning trees with bending energy on planar maps and sle-decorated liouville quantum gravity for $\kappa > 8$. *ArXiv Mathematics e-prints* (2016). arXiv: 1603.09722.

E. Gwynne, C. Mao and X. Sun. Scaling limits for the critical Fortuin-Kasteleyn model on a random planar map I: cone times. *ArXiv Mathematics e-prints* (2015). arXiv: 1502.00546.

E. Gwynne and X. Sun. Scaling limits for the critical Fortuin-Kastelyn model on a random planar map III: finite volume case. *ArXiv Mathematics e-prints* (2015). arXiv: 1510.06346.

E. Gwynne and X. Sun. Scaling limits for the critical Fortuin-Kasteleyn model on a random planar map II: local estimates and empty reduced word exponent. *Electron. J. Probab.* **22**, Paper No. 45, 56 (2017). MR3661659.

I. Z. Karaesmen, A. Scheller-Wolf and B. Deniz. *Managing Perishable and Aging Inventories: Review and Future Research Directions*, pages 393–436. Springer US (2011). ISBN 978-1-4419-6485-4. DOI: 10.1007/978-1-4419-6485-4_15.

H. Kaspi and D. Perry. Inventory systems of perishable commodities. *Adv. in Appl. Probab.* **15** (3), 674–685 (1983). MR706623.

J. Keilson and A. Seidmann. Product selection policies for perishable inventory systems (1990). Available on-line at https://dspace.mit.edu/handle/1721.1/5372#files-area.

Ł. Kruk. Functional limit theorems for a simple auction. *Math. Oper. Res.* **28** (4), 716–751 (2003). MR2015910.

J.-F. Le Gall and G. Miermont. Scaling limits of random planar maps with large faces. *Ann. Probab.* **39** (1), 1–69 (2011). MR2778796.

J.-F. Le Gall and G. Miermont. Scaling limits of random trees and planar maps. In *Probability and statistical physics in two and more dimensions*, volume 15 of *Clay Math. Proc.*, pages 155–211. Amer. Math. Soc., Providence, RI (2012). MR3025391.

V. Limic. A LIFO queue in heavy traffic. *Ann. Appl. Probab.* **11** (2), 301–331 (2001). MR1843048.

D. Mukherjee, S. C. Borst, J. S. H. van Leeuwaarden and P. A. Whiting. Universality of load balancing schemes on the diffusion scale. *J. Appl. Probab.* **53** (4), 1111–1124 (2016). MR3581245.

S. Nahmias. Perishable inventory theory: A review. *Operations research* **30** (4), 680–708 (1982).

M. Parlar, D. Perry and W. Stadje. FIFO versus LIFO issuing policies for stochastic perishable inventory systems. *Methodol. Comput. Appl. Probab.* **13** (2), 405–417 (2011). MR2788865.

W. Paulsen. *Abstract algebra*. Textbooks in Mathematics. CRC Press, Boca Raton, FL (2016). ISBN 978-1-4987-1976-6. MR3468791.

D. Perry. A double band control policy of a Brownian perishable inventory system. *Probab. Engrg. Inform. Sci.* **11** (3), 361–373 (1997). MR1457239.

J. W. Pitman. One-dimensional Brownian motion and the three-dimensional Bessel process. *Advances in Appl. Probability* **7** (3), 511–526 (1975). MR0375485.

G. P. Prastacos. LIFO distribution systems. *J. Oper. Res. Soc.* **30** (6), 539–546 (1979). MR533239.

S. Sheffield. Quantum gravity and inventory accumulation. *Ann. Probab.* **44** (6), 3804–3848 (2016). MR3572324.

X. Sun and D. B. Wilson. Sandpiles and unicycles on random planar maps. *Electron. Commun. Probab.* **21**, Paper No. 57, 12 (2016). MR3548769.

W. T. Tutte. A census of planar maps. *Canad. J. Math.* **15**, 249–271 (1963). MR0146823.

L. M. Wein. Dynamic scheduling of a multiclass make-to-stock queue. *Oper. Res.* **40** (4), 724–735 (1992). MR1179800.