Title

Diffusion approximation for a heavily loaded multi-user wireless communication system with cooperation

Journal

Queueing Systems: Theory and Applications, 62(4)

ISSN

1572-9443

Authors

Bhardwaj, S.
Williams, R. J.

Publication Date

2009-08-01

DOI

10.1007/s11134-009-9119-8

Peer reviewed

Queueing Syst (2009) 62: 345–382
DOI 10.1007/s11134-009-9119-8

Diffusion approximation for a heavily loaded multi-user wireless communication system with cooperation

S. Bhardwaj · R.J. Williams

Received: 23 May 2008 / Revised: 20 April 2009 / Published online: 12 June 2009

Abstract A cellular wireless communication system in which data is transmitted to multiple users over a common channel is considered. When the base stations in this system can cooperate with each other, the link from the base stations to the users can be considered a multi-user multiple-input multiple-output (MIMO) downlink system. For such a system, it is known from information theory that the total rate of transmission can be enhanced by cooperation. The channel is assumed to be fixed for all transmissions over the period of interest and the ratio of anticipated average arrival rates for the users, also known as the relative traffic rate, is fixed. A packet-based model is considered where data for each user is queued at the transmit end. We consider a simple policy which, under Markovian assumptions, is known to be throughput-optimal for this coupled queueing system. Since an exact expression for the performance of this policy is not available, as a measure of performance, we establish a heavy traffic diffusion approximation. To arrive at this diffusion approximation, we use two key properties of the policy; we posit the first property as a reasonable manifestation of cooperation, and the second property follows from coordinate convexity of the capacity region. The diffusion process is a semimartingale reflecting Brownian motion (SRBM) living in the positive orthant of N-dimensional space (where N is the number of users). This SRBM has one direction of reflection associated with each of the $2^N - 1$ boundary faces, but show that, in fact, only those directions associated with

The research of S. Bhardwaj was supported by a UCSD ECE departmental dissertation fellowship for 2007–08.
The research of R.J. Williams was supported in part by NSF grant DMS-0604537.

S. Bhardwaj (✉)
Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA
e-mail: bhardwajs@alumni.ucsd.edu

R.J. Williams
Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA
e-mail: williams@math.ucsd.edu

the $(N-1)$-dimensional boundary faces matter for the heavy traffic limit. The latter is likely of independent theoretical interest.

Keywords Cooperative queueing systems · Cellular wireless systems · Heavy traffic · Multi-input multi-output (MIMO) · Semimartingale reflecting Brownian motion (SRBM) · Reduction of directions of reflection · Multi-dimensional diffusion

Mathematics Subject Classification (2000) Primary 60J60 · 60K25 · 90B18 · 90B22

1 Introduction

Current cellular wireless systems consider each base station as a separate entity with no cooperation among base stations. Infrastructure cooperation, that is, cooperation among base stations, has been proposed as a means of achieving higher throughput (see, e.g. [6, 17, 22]) where the main idea is to consider the base stations as one end of a multiple-input multiple-output (MIMO) system. For such a system, it is known from the information-theoretic literature that the rate of transmission can be enhanced by cooperation at the transmit end, that is, among the base stations.

In this paper, we consider a MIMO downlink system where data is buffered at the transmit end and the channel is assumed to be fixed for all transmissions over the period of interest (one might view this as one period for a quasi-static channel). The N-user (where N is an arbitrary positive integer) MIMO downlink system can be seen as a model of a cellular system with N users and multiple cooperating base station antennas. The latter might consist of multiple cooperating base stations, each with a single antenna, or a single-cell cellular system with a multi-antenna base station or a combination thereof.

This communication system has a corresponding queueing system formulation where, even in the simple case of Poisson arrivals, independently for each user, it is not known how to minimize the average delay for a given load. Furthermore, closed-form expressions for average delay are unavailable for many simple policies; usually, this means that any meaningful comparison has to be done via simulations. However, when the ratio of the average arrival rates (also known as the relative traffic rate) is specified in advance, the maximum possible throughput can be computed, and a simple policy can be shown to be throughput-optimal[1] under Markovian assumptions. An exact expression for the performance of this policy is not available. In this paper, as a measure of performance, we prove a limit theorem justifying a diffusion approximation for the queueing system when heavily loaded and operating under this policy (see Theorem 6.1). For this, we use two key properties of the policy: (9) and (10); we posit the first property as a reasonable manifestation of cooperation and the second property follows from coordinate convexity of the capacity region. The approximating diffusion is an N-dimensional semimartingale reflecting Brownian motion

[1] For a Markovian system, throughput-optimal means that the long run average departure rate exists and equals the long run average arrival rate whenever the nominal load lies inside the capacity region, cf. [9, p. 26].

(SRBM) living in the positive N-dimensional orthant. Our limit theorem has general distributional assumptions on the arrivals and packet lengths. In particular, we do not require Markovian assumptions.

We are not aware of analyses of other policies that have been shown to be throughput-optimal for a general convex (rather than a convex polyhedral) capacity region. However, scheduling policies for certain heavily loaded wireless systems with convex polyhedral capacity regions have been studied in [21, 23] (also see references therein). In [23], Stolyar considered a generalized switch. He showed that under MaxWeight scheduling and certain restrictive conditions, including a resource pooling condition, in heavy traffic there is state space collapse (SSC), the workload process converges to a one-dimensional reflecting Brownian motion (RBM), and MaxWeight asymptotically minimizes the workload. Shakkotai et al. [21] studied a throughput-optimal scheduling rule, which they called an exponential scheduling rule, and showed that under a resource pooling condition this policy is asymptotically pathwise optimal in the sense that there is SSC, the workload process is asymptotically minimized and converges to a one-dimensional RBM. In the following, we point out some of the differences between our assumptions and those in [21, 23]. The Maxweight policy [23] is designed for the case when the capacity region is a convex polyhedron while the policy we consider is designed for more general convex capacity regions. Moreover, a complete resource pooling (CRP) condition is assumed in [23]. Whereas for the convex capacity region considered here, the analogue of the CRP condition typically does not hold (see Sect. V.B of [3] for further explanation of this point). The arrival process in [23] is assumed to be an ergodic Markov process while we assume that the arrival process is a renewal process. In [21], the capacity region is a convex polyhedron and a CRP condition similar to that in [23] is assumed; however, service is given to only one queue at a time while here we can serve more than one queue at the same time, which leads to an enhanced transmission rate. A significant difference between our work and that in [21, 23] is that we do not assume complete resource pooling and accordingly our diffusion approximation is in general multi-dimensional rather than one-dimensional.

The rest of this paper is organized as follows. In Sect. 1.1, we explain the notation used in this paper and present some mathematical preliminaries. We describe the communication system of interest in Sect. 2 and develop a queueing analogue for it in Sect. 3. The stochastic assumptions for the model are specified in Sect. 3 and the workload process is introduced as our performance process of interest. The service policy and its key properties are described in Sect. 3.4. We formally define the heavy traffic conditions in Sect. 4. In Sect. 5, we define scaling, present standard functional limit theorems for the stochastic primitives, and define some parameters for the limit process. In Sect. 6, we first define an SRBM (Definition 6.1), and then present the main result of this paper (Theorem 6.1) which states that the sequence of diffusion-scaled workload processes converges in distribution to an SRBM as described in Definition 6.1. We provide our proof of the main result in Sect. 7. The first step in the proof is to show that the sequence of diffusion-scaled workload processes is C-tight for which we use some recent results of Kang and Williams [14]. A key result for our proof of Theorem 6.1 is Theorem 7.7. The limit SRBM has one direction of reflection associated with each of the $2^N - 1$ boundary faces. In Theorem 7.7,

we show that, in fact, only those directions associated with the $(N-1)$-dimensional boundary faces matter for the heavy traffic limit. Appendix A contains the proof of an auxiliary lemma.

For two-user systems, a result similar to Theorem 6.1 was proved in Theorem VIII.3 of Bhardwaj, Williams, and Acampora [3]. The result here is considerably more general and the proof is different. Indeed, the result in [3] is only for a two-user system, whereas the result presented in this paper is for an arbitrary number of users. Moreover, the approximation result in [3] is for diffusion-scaled queue length, while our main theorem here is for diffusion-scaled workload. In [3], the SRBM data was simplified because in two dimensions the nominal vector of reflection at the origin can be written as a convex combination of the directions on the two sides of the quadrant. In higher dimensions, this is usually not possible. Indeed, there is one direction of reflection for each of the 2^N-1 boundary faces. A key element of the proof presented here is to show that the pushing at boundary faces of dimension $N-2$ or less is inconsequential (see Theorem 7.7).

1.1 Notation and preliminaries

We will use the following notation throughout the paper. We will use $\mathcal{N}$ to denote the set $\{1,2,\ldots,N\}$ where N is a finite positive integer, $\mathcal{K}$ to denote an arbitrary subset of $\mathcal{N}$, and $\mathcal{K}^c$ to denote the complement of $\mathcal{K}$ in $\mathcal{N}$. We will use $\mathcal{P}(\mathcal{A})$ to indicate the power set of an arbitrary set $\mathcal{A}$. We will use $|\mathcal{A}|$ to denote the cardinality of the set $\mathcal{A}$. The symbol $1_{\mathcal{A}}$ denotes the indicator function of a set $\mathcal{A}$, i.e., $1_{\mathcal{A}}(x)=1$ if $x\in\mathcal{A}$ and $1_{\mathcal{A}}(x)=0$ if $x\notin\mathcal{A}$.

Let $\mathbb{Z}$ denote the set of all integers, $\mathbb{Z}_+$ the set of all non-negative integers, $\mathbb{R}$ denote the set of real numbers, and $\mathbb{R}_+$ denote the set of non-negative real numbers, which is also denoted by $[0,\infty)$. The symbol $\mathbb{R}^N$ will denote N-dimensional Euclidean space, and the positive orthant in this space will be denoted by $\mathbb{R}^N_+=\{x\in\mathbb{R}^N : x_i\geq 0 \text{ for all } i\in\mathcal{N}\}$. All vectors and matrices in this paper are assumed to have real-valued entries. Let $0=(0,0,\ldots,0)\in\mathbb{R}^N_+$. We denote the inner product on $\mathbb{R}^N$ by $\langle\cdot,\cdot\rangle$, i.e., $\langle x,y\rangle=\sum_{i=1}^N x_i y_i$, for $x,y\in\mathbb{R}^N$. The usual Euclidean norm on $\mathbb{R}^N$ will be denoted by $\|\cdot\|$ so that $\|x\|=\sqrt{\langle x,x\rangle}=(\sum_{i=1}^N x_i^2)^{1/2}$ for $x\in\mathbb{R}^N$. Let $\mathcal{B}(\mathbb{R}^N)$ denote the σ-algebra of Borel subsets of $\mathbb{R}^N$. For any non-empty set $\mathcal{K}\subseteq\mathcal{N}$ and any $x\in\mathbb{R}^N$, $x_{\mathcal{K}}$ will denote the vector whose components are those of x with indices in $\mathcal{K}$. Let $e_{\mathcal{N}}\in\mathbb{R}^N$ denote the vector whose entries are all 1. For $x,y\in\mathbb{R}^N$, we shall use $x\wedge y$ to denote the vector whose i-th component is the minimum of x_i and y_i for each $i\in\mathcal{N}$. All vector inequalities are understood to hold componentwise. For $a\in\mathbb{R}^N$, we shall use $\mathrm{diag}(a)$ to denote the $N\times N$ diagonal matrix whose diagonal entries are given by the entries in a. We will let $(\cdot)'$ denote transpose. For any set $\emptyset\neq\mathcal{K}\subseteq\mathcal{N}$, we define the face $F_{\mathcal{K}}$ by

$$F_{\mathcal{K}}\triangleq\big\{x\in\mathbb{R}^N_+ : x_i=0 \text{ for all } i\in\mathcal{K}\big\}. \tag{1}$$

For example, $F_{\mathcal{N}}=\{0\}$, the set consisting of the origin in $\mathbb{R}^N$. When $\mathcal{K}=\{i\}$ for $i\in\mathcal{N}$, we write F_i in place of $F_{\{i\}}$ sometimes. We define the index set of any point $x\in\mathbb{R}^N_+$ by

$$\mathcal{K}(x)\triangleq\{i\in\mathcal{N} : x_i=0\} \tag{2}$$

with the convention that $\mathcal{K}(w) = \emptyset$ if $w > 0$. A domain in $\mathbb{R}^N$ is an open connected subset of $\mathbb{R}^N$. For each continuously differentiable real-valued function f defined on some non-empty domain $S \subseteq \mathbb{R}^N$, $\nabla f(x)$ is the gradient of f at $x \in S$:

$$\big(\nabla f(x)\big)_i = \frac{\partial f}{\partial x_i}(x), \quad i = 1, 2, \ldots, N. \tag{3}$$

For any set $S \subseteq \mathbb{R}^N$, we write $\overline{S}$ for the closure of S, S^o for the interior of S, and $\partial S = \overline{S} \setminus S^o$.

All stochastic processes used in this paper will be assumed to have paths that are right continuous with finite left limits (r.c.l.l.). We denote by $\mathbb{D}^N$ the space of r.c.l.l. functions from $[0, \infty)$ into $\mathbb{R}^N$ and we endow this space with the usual Skorokhod J_1-topology (see Ethier and Kurtz [8, Chap. 3, Sect. 5]) which makes it a Polish space. We denote by $\mathbb{C}^N$ the space of continuous functions from $[0, \infty)$ into $\mathbb{R}^N$, also endowed with the Skorokhod J_1-topology under which convergence of elements in $\mathbb{C}^N$ is equivalent to uniform convergence on compact time intervals. We endow $\mathbb{D}^N$ (or $\mathbb{C}^N$) with the Borel σ-algebra induced by the Skorokhod J_1-topology and denote this σ-algebra by $\mathcal{M}^N$. The abbreviation *u.o.c.* will stand for *uniformly on compacts* and will be used to indicate that a sequence of functions in $\mathbb{D}^N$ (or $\mathbb{C}^N$) is converging uniformly on compact time intervals to a limit in $\mathbb{D}^N$ (or $\mathbb{C}^N$). An N-dimensional process is a measurable function from a probability space into $(\mathbb{D}^N, \mathcal{M}^N)$. Consider $W, W^1, W^2, \ldots$, each of which is an N-dimensional process (possibly defined on different probability spaces). The sequence $\{W^n\}_{n=1}^{\infty}$ is said to be *tight* if the probability measures induced by the sequence $\{W^n\}_{n=1}^{\infty}$ on $(\mathbb{D}^N, \mathcal{M}^N)$ form a tight sequence, i.e., they form a weakly relatively compact sequence in the space of probability measures on $(\mathbb{D}^N, \mathcal{M}^N)$. The notation "$W^n \Rightarrow W$" will mean that "$W^n$ converges in distribution to W as $n \to \infty$". The sequence of processes $\{W^n\}_{n=1}^{\infty}$ is called *C-tight* if it is tight and if each weak limit point (obtained as a weak limit along a subsequence) is in $\mathbb{C}^N$ almost surely.

A triple $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\})$ will be called a filtered space if Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and $\{\mathcal{F}_t, t \geq 0\}$ is an increasing family of sub-σ-algebras of $\mathcal{F}$, i.e., a filtration. From now on, we will write a filtration $\{\mathcal{F}_t, t \geq 0\}$ as simply $\{\mathcal{F}_t\}$. If P is a probability measure on $(\Omega, \mathcal{F})$, then $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is called a filtered probability space. An N-dimensional process $X = \{X(t), t \geq 0\}$ defined on $(\Omega, \mathcal{F}, P)$ is called $\{\mathcal{F}_t\}$-adapted if for each $t \geq 0$, $X(t) : \Omega \to \mathbb{R}^N$ is measurable when Ω is endowed with the σ-algebra $\mathcal{F}_t$ and $\mathbb{R}^N$ has the usual Borel σ-algebra $\mathcal{B}(\mathbb{R}^N)$, and X is said to be a continuous process if its sample paths are continuous P-a.s.

2 Communication system model

In this section we specify the communication system under consideration. We consider a cellular wireless network where base stations cooperate over noise-free infinite capacity links. We do not make any distinction between a single-cell cellular system having multiple base-station antennas and the traditional cellular system with cooperating single-antenna base stations. Here by cooperation we mean that the base

stations can perform joint beamforming and/or power control but there is a constraint on the total power that the base stations can share. We do not make any assumptions about the number of receive antennas per user.

The downlink channel for such a system with N users can be modeled as an N-user MIMO Broadcast Channel (BC). We assume that the channel is fixed for all transmissions over the period of interest (some authors refer to this as a quasi-static channel). Moreover, we assume that the transmit end (with the cooperating base stations) has perfect channel state information (CSI).

Weingarten et al. [24] have shown that for such a system, dirty paper coding (DPC), introduced by Costa [7], achieves the capacity. Furthermore, the capacity region can be computed by using the duality of the MIMO multiple access channel (MAC) and the MIMO BC [13] where the BC capacity region is obtained by taking the convex hull of the union over the set of capacity regions of the dual MIMO MACs such that the total MAC power is the same as the power in the BC.

For an N-user system, the capacity region is an N-dimensional closed, bounded convex and coordinate convex set in $\mathbb{R}_+^N$ containing the origin. (Here coordinate convex means that if x is in the region, then for any $y \in \mathbb{R}_+^N$, $x - y$ is in the region whenever $x - y$ is in $\mathbb{R}_+^N$.) For an example of such a capacity region in the two-user case, see Fig. 1 of [3]. Coordinate convexity leads to the property (10) of our policy. We assume that cooperation leads to the property (9), expressed in terms of sums of rates. In the case of two users, the latter simply requires that the bit rates have been normalized so that the two single-user capacities are equal and then the property is known to hold for MIMO systems [3]. For more than two users, (9) is an assumption which we propose as a reasonable generalization of the two-dimensional case. However, there is currently no proof that this property holds for MIMO systems.

At the transmit end, packets arrive for each user and are buffered before transmission. We assume that there is given a nominal average packet arrival rate (e.g., an estimate of the true average arrival rate) and nominal average packet size (measured in bits). The nominal average *bit* arrival rate for each user is then the product of the nominal average packet arrival rate times the nominal average packet size for that user. The ratio of the nominal average bit arrival rate for user i relative to that for user 1 is called the relative traffic rate and is denoted by κ_i (this is assumed to be strictly positive). This nominal relative traffic rate is specified in advance with the assumption that $\kappa_1 = 1$; thus, it is expected that, on average, the i-th user will have κ_i times as much data as user 1. The actual traffic rate may deviate from this nominal average rate due to estimation error and stochastic fluctuations. Naturally, when there is no data for one (or many) of the users to transmit (the corresponding queue for that(those) user(s) is empty), these users do not receive any transmission capacity and the other users can expect an enhanced transmission rate. We formally describe the transmission policy and associated conditions in Sect. 3.

3 Queueing analogue

In this section, we develop a queueing analogue for the system described in Sect. 2. To this end, we describe the physical structure, and the stochastic primitives specifying the packet arrivals and sizes. We formulate dynamic equations satisfied by the

workload process in terms of the stochastic primitives and the policy or service discipline to be used with this system.

3.1 Physical structure

A queueing model describing our communication system has N queues in parallel where each queue buffers packets intended for a given user. We assume that each of the queues has infinite buffer capacity. The queues are served by a single server corresponding to a base station with multiple cooperating antennas.

3.2 Stochastic primitives

We assume that the system starts empty and that there is an N-dimensional packet arrival process $E = \{(E_1(t), E_2(t), \ldots, E_N(t)), t \geq 0\}$ where $E_i(t)$ is the number of packets that have arrived to the i-th queue in $(0, t]$. (Here E is used to indicate that the arrivals are *exogenous*.) For $i \in \mathcal{N}$, $E_i(\cdot)$ is assumed to be a (non-delayed) renewal process defined from a sequence of strictly positive independent and identically distributed (i.i.d.) random variables $\{u_i(k), k = 1, 2, \ldots\}$, where for $k = 1, 2, \ldots$, the random variable $u_i(k)$ denotes the time between the arrival of the $(k-1)$-st and the k-th packet to the i-th queue (where the 0-th arrival occurs at time 0). Each $u_i(k)$, $k = 1, 2, \ldots$, is assumed to have finite mean $1/\lambda_i \in (0, \infty)$ and finite squared coefficient of variation (variance divided by the mean squared) $\alpha_i^2 \in (0, \infty)$. Then

$$E_i(t) = \max\left\{n \geq 0 : \sum_{j=1}^{n} u_i(j) \leq t\right\}, \quad i \in \mathcal{N},\ t \geq 0, \tag{4}$$

where a sum up to $n = 0$ is defined to be zero. The packet lengths (in bits) for the successive arrivals to the i-th queue are given by a sequence of strictly positive i.i.d. random variables $\{v_i(k), k = 1, 2, \ldots\}$ with average packet length $m_i = 1/\mu_i \in (0, \infty)$ and squared coefficient of variation $\beta_i^2 \in (0, \infty)$. We assume that all interarrival and service time processes are mutually independent. For $i \in \mathcal{N}$ and $n \in \mathbb{Z}_+$, we define

$$V_i(n) \triangleq \sum_{j=1}^{n} v_i(j). \tag{5}$$

We refer to the processes $E(\cdot)$ and $V(\cdot)$ as *stochastic primitives* for our system model. For convenience, to avoid the need to consider exceptional null sets, we assume without loss of generality that $E_i(t) < \infty$ for all $t \geq 0$ and $E_i(t) \to \infty$ as $t \to \infty$ for each $i \in \mathcal{N}$, surely.

3.3 Workload process

For $i \in \mathcal{N}$, the workload $W_i(t)$ of the i-th queue at time $t \geq 0$ is given by

$$\begin{aligned} W_i(t) &\triangleq \sum_{j=1}^{E_i(t)} v_i(j) - T_i(t) \\ &= V_i\big(E_i(t)\big) - T_i(t), \end{aligned} \tag{6}$$

where $T_i(t)$ is the cumulative amount of service (measured in bits) given to the i-th queue up to time t. We next describe the service discipline which, in turn, specifies the functional form of $T_i(\cdot)$.

3.4 Service discipline

When service is given to a queue, it goes to the packet at the head of the line, where it is assumed that packets are queued in the order of their arrival with the packet that arrived the longest time ago being at the head of the line. A vector $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ indicates the rates (in bits per second) of serving the N queues, i.e., σ_1 is the rate for queue 1, σ_2 is the rate for queue 2, and so on. The service rate for each queue is a very simple function of the vector of workloads. Given a workload of $w = (w_1, w_2, \ldots, w_N)$, the set of indices for the empty queues is the index set $\mathcal{K}(w)$, as defined by (2). The rates $\sigma = \Lambda(w)$ are given by the function[2] $\Lambda : \mathbb{R}_+^N \to \mathbb{R}_+^N$ defined by

$$\Lambda(w) \triangleq c^{\mathcal{K}(w)}, \tag{7}$$

where $c^{\mathcal{K}}$ is a fixed vector for each $\mathcal{K} \subseteq \mathcal{N}$ with $c_i^{\mathcal{K}} = 0$ if $i \in \mathcal{K}$ (corresponding to the fact that an empty queue should not be served) and $c_i^{\mathcal{K}} > 0$ if $i \notin \mathcal{K}$. The vector of service rates $c^{\mathcal{K}}$ is chosen such that it lies on the boundary of the capacity region and the service rate for each of the users with positive workload is related by the relative traffic rate as described below. Recall, from Sect. 2, $(\kappa_i, i \in \mathcal{N})$ is the given vector of nominal relative traffic rates. For all $\mathcal{K} \subsetneqq \mathcal{N}$, the non-zero entries of the service rate vector $c^{\mathcal{K}}$ are chosen such that

$$\frac{c_i^{\mathcal{K}}}{\kappa_i} = \frac{c_j^{\mathcal{K}}}{\kappa_j} \tag{8}$$

whenever $i, j \in \mathcal{K}^c$, and $\sum_i c_i^{\mathcal{K}}$ is as large as possible while still keeping $c^{\mathcal{K}}$ in the capacity region. (We make the non-degeneracy assumption that the capacity region is such that we can choose $c_i^{\mathcal{K}} > 0$ for all $i \in \mathcal{K}^c$.) When all of the queues are non-empty ($\mathcal{K} = \emptyset$), the service rate vector, $c^{\emptyset}$, lies on the boundary of the capacity region and for all $i \in \mathcal{N}$, $c_i^{\emptyset} = \kappa_i c_1^{\emptyset}$, i.e., $c^{\emptyset}$ is in the direction of the vector κ and is the furthest point along that direction which lies in the capacity region (see Fig. 1 for an example of the capacity region and the service rates for a two-user system).

The following condition is assumed to be satisfied by the $c^{\mathcal{K}}$'s. It requires that the maximum of the sum of the rates is achieved only when all of the queues are non-empty. This form of cooperation is known to hold for two-user MIMO systems [3] and it seems a reasonable generalization for N-user systems, although a formal proof of this property is not known at this time.

$$\sum_{i \in \mathcal{N}} c_i^{\emptyset} > \sum_{i \in \mathcal{N}} c_i^{\mathcal{K}} \quad \text{for all } \emptyset \subsetneqq \mathcal{K} \subseteq \mathcal{N}. \tag{9}$$

[2] We only need $\Lambda(\cdot)$ defined on $\mathbb{Z}_+^N$ for the moment, but we extend the domain of $\Lambda(\cdot)$ to $\mathbb{R}_+^N$ so that later when we rescale the workload process, $\Lambda(\cdot)$ is well defined for the rescaled process.

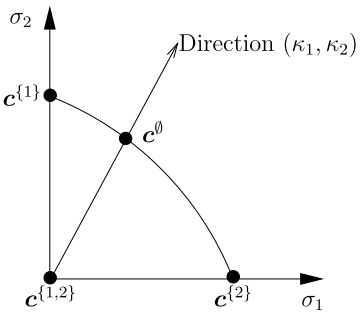

Fig. 1 An example of the capacity region for a two-user system. Service rate $c^{\{1,2\}} = (0,0)$, $c^{\{2\}}$ is along the direction $(\kappa_1, 0)$ and $c^{\{1\}}$ is along the direction $(0, \kappa_2)$

As a result of coordinate convexity of the capacity region, the service rate for a fixed non-empty queue is least when all of the queues are non-empty. Therefore,

$$c_i^{\emptyset} \le c_i^{\mathcal{K}} \quad \text{for all } i \notin \mathcal{K} \text{ and } \mathcal{K} \neq \emptyset. \tag{10}$$

For example, $c_i^{\emptyset} \le c_i^{\{j\}}$ for all $i \neq j$, $i, j \in \mathcal{N}$.

Remark Properties (9) and (10) are used to prove properties of the reflection vectors associated with our diffusion approximation for the workload process. In particular, they are used in proving Lemma 5.3 and the properties in Sect. 7.3.2.

Our model is a single server, N-class queueing system where the N classes correspond to the N queues (users). The following scaling property of $\Lambda(\cdot)$ is a mathematical statement of the property of the scheduling policy that the amount of service given to the queues in any state does not change when all workloads are multiplied by the same positive factor.

Lemma 3.1 *For any $w \in \mathbb{R}_+^N$ and $a > 0$, $\Lambda(aw) = \Lambda(w)$.*

Proof The proof follows easily from the fact that $\Lambda(\cdot)$ depends only on which queues are empty and these are unchanged by the positive scalar factor a. □

For $t \ge 0$, $i \in \mathcal{N}$, we can now give an explicit expression for $T_i(t)$ as

$$\begin{aligned} T_i(t) &\triangleq \int_0^t \Lambda_i\big(W(s)\big)\,ds \\ &= \sum_{\mathcal{K} \subseteq \mathcal{N}} c_i^{\mathcal{K}} \int_0^t 1_{\{\mathcal{K}(W(s))=\mathcal{K}\}}\,ds. \end{aligned} \tag{11}$$

In fact, $c^{\mathcal{N}} = 0$ and so the sum could be reduced to that over $\mathcal{K} \subsetneq \mathcal{N}$, including $\mathcal{K} = \emptyset$.

4 Heavy traffic assumptions

We wish to consider the behavior of the queueing system when it is heavily loaded. (Kelly and Laws [15] have argued that in this regime "important features of good control policies are displayed in sharpest relief".) For this purpose one may regard a given system as a member of a sequence of systems approaching the heavy traffic limit. To obtain a reasonable approximation, the workload process is rescaled using diffusion scaling. This corresponds to viewing the system over long intervals of time of order r^2 (where r will tend to infinity in the asymptotic limit) and regarding a single packet as only having a small contribution to the overall congestion level, where this is quantified to be of order $1/r$. Formally, we consider a sequence of systems indexed by r, where r tends to infinity through a sequence of values in $(0, \infty)$. These systems all have the same basic structure as that described in the last section; however, the arrival rates may vary with r. We assume that the interarrival times for the system indexed by r are given for each $i \in \mathcal{N}$, $k = 1, 2, \ldots$, by

$$u_i^r(k) = \frac{1}{\lambda_i^r} \check{u}_i(k), \tag{12}$$

where the $\check{u}_i(k)$ do not depend on r, have mean one and squared coefficient of variation α_i^2. The packet lengths $\{v_i(k)\}_{k=1}^{\infty}$, $i \in \mathcal{N}$, do not change with r. [The above structure is convenient for allowing the sequence of systems to approach heavy traffic by simply changing arrival rates and keeping the underlying sources of variability $\check{u}_i(k)$ and $v_i(k)$ fixed as r varies. This type of set-up has been used previously by others in treating heavy traffic limits (see, e.g., Peterson [18] and Bell and Williams [1]). For a first pass, the reader may want to simply choose $\lambda_i^r = \lambda_i$ for all r.] All processes and parameters that depend on r will from now on have a superscript of r appended. The nominal relative traffic rate and the service rates $\{c^{\mathcal{K}}, \mathcal{K} \subseteq \mathcal{N}\}$ are assumed fixed throughout and do not vary with r. We define $\lambda_i \triangleq \mu_i c_i^{\emptyset}$ for $i = 1, 2, \ldots, N$.

Assumption 4.1 (Heavy Traffic Assumption) There exists $\theta \in \mathbb{R}_+^N$ such that for each $i \in \mathcal{N}$,

$$r\big(\lambda_i^r - \lambda_i\big)m_i \to \theta_i \quad \text{as } r \to \infty. \tag{13}$$

We may regard $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ as the nominal average packet arrival rate with nominal bit arrival rate $b = (\lambda_i m_i, i \in \mathcal{N})$ and nominal traffic rate given by $\kappa_i = b_i/b_1$ for $i \in \mathcal{N}$. Then under the heavy traffic assumption, the service rates

$$\big(c_1^{\emptyset}, c_2^{\emptyset}, \ldots, c_N^{\emptyset}\big),$$

for the scheduling policy satisfy $b = c^{\emptyset}$ and the r-th system has a perturbed average packet arrival rate λ^r for which the average bit arrival rate $b^r = (\lambda_i^r m_i, i \in \mathcal{N})$ is close to $(c_1^{\emptyset}, c_2^{\emptyset}, \ldots, c_N^{\emptyset})$.

5 Scaling, standard limit theorems, and parameters

5.1 Scaling

Fluid (or functional law of large numbers) scaling is indicated by placing a bar over a process. For $r > 0$, $i \in \mathcal{N}$, and $t \geq 0$, we define

$$\bar{E}_i^r(t) \triangleq r^{-2} E_i^r\left(r^2 t\right), \tag{14}$$

$$\bar{V}_i^r(t) \triangleq r^{-2} V_i^r\left(r^2 t\right), \tag{15}$$

$$\bar{T}_i^r(t) \triangleq r^{-2} T_i^r\left(r^2 t\right), \tag{16}$$

$$\bar{W}_i^r(t) \triangleq r^{-2} W_i^r\left(r^2 t\right). \tag{17}$$

Diffusion (or functional central limit theorem) scaling is indicated by placing a hat over a process. For $r > 0$, $i \in \mathcal{N}$, and $t \geq 0$, we define

$$\hat{W}_i^r(t) \triangleq \frac{W_i^r(r^2 t)}{r}. \tag{18}$$

To apply diffusion-scaling to the primitive stochastic processes $E^r(\cdot)$ and $V(\cdot)$ (note that $V(\cdot)$ does not depend on r), we must center them before scaling. Accordingly, for $r > 0$, $i \in \mathcal{N}$, and $t \geq 0$, we define

$$\hat{E}_i^r(t) \triangleq \frac{1}{r}\left(E_i^r\left(r^2 t\right) - \lambda_i^r r^2 t\right) \tag{19}$$

and

$$\hat{V}_i^r(t) \triangleq \frac{1}{r}\left(V_i\left(r^2 t\right) - m_i r^2 t\right). \tag{20}$$

5.2 Functional limit theorems for stochastic primitives

We will use the following functional central limit theorem (FCLT) for the stochastic primitives in the sequel.

Proposition 5.1 (FCLT) *The diffusion-scaled processes* $(\hat{E}^r(\cdot), \hat{V}^r(\cdot))$ *jointly converge in distribution to* $(B_E(\cdot), B_V(\cdot))$ *as* $r \to \infty$, *i.e.*,

$$\left(\hat{E}^r(\cdot), \hat{V}^r(\cdot)\right) \Rightarrow \left(B_E(\cdot), B_V(\cdot)\right) \quad \text{as } r \to \infty, \tag{21}$$

where $B_E(\cdot)$ *and* $B_V(\cdot)$ *are independent* N*-dimensional driftless Brownian motions starting from the origin with diagonal covariance matrices*

$$\Gamma_E \triangleq \operatorname{diag}\left(\lambda_1 \alpha_1^2, \lambda_2 \alpha_2^2, \ldots, \lambda_N \alpha_N^2\right) \tag{22}$$

and

$$\Gamma_V \triangleq \operatorname{diag}\left(m_1^2 \beta_1^2, m_2^2 \beta_2^2, \ldots, m_N^2 \beta_N^2\right), \tag{23}$$

respectively.

Remark As there is a single source of variability (not depending on r) for each of E_i^r, V_i, $i \in \mathcal{N}$, only the finiteness of the second moments of $\breve{u}_i(k)$ and $v_i(k)$ is required for the FCLT. Furthermore, since a Brownian motion is a continuous process, the weak convergence of $(\hat{E}^r(\cdot), \hat{V}^r(\cdot))$ to a Brownian motion implies C-tightness of the sequence $\{(\hat{E}^r(\cdot), \hat{V}^r(\cdot))\}$.

Proof By results of Iglehart and Whitt [11], an FCLT for the renewal counting process $E^r(\cdot)$ can be inferred from that for the partial sums of $\{u_i^r(k)\}_{k=1}^{\infty}$. FCLTs for the partial sums of $\{u_i^r(k)\}_{k=1}^{\infty}$ and $\{v_i(k)\}_{k=1}^{\infty}$ follow from Theorem 3.1 of Prokhorov [19]. The joint convergence follows from the independence of $E^r(\cdot)$ and $V(\cdot)$. □

As a corollary, we have the following functional law of large numbers (FLLN) for the stochastic primitives. For each $t \geq 0$, let $\lambda(t) \triangleq \lambda t$ and $m(t) \triangleq mt$.

Corollary 5.2 (FLLN) *The fluid-scaled processes* $(\bar{E}^r(\cdot), \bar{V}^r(\cdot))$ *jointly converge in distribution to* $(\lambda(\cdot), m(\cdot))$ *as* $r \to \infty$, *i.e.*,

$$\left(\bar{E}^r(\cdot), \bar{V}^r(\cdot)\right) \Rightarrow \left(\lambda(\cdot), m(\cdot)\right) \quad \text{as } r \to \infty. \tag{24}$$

Remark Here again, the weak convergence of $(\bar{E}^r(\cdot), \bar{V}^r(\cdot))$ to a continuous process implies C-tightness of the sequence $\{(\bar{E}^r(\cdot), \bar{V}^r(\cdot))\}$.

Proof Proposition 5.1 implies that

$$\left(\frac{1}{r}\hat{E}^r(\cdot), \frac{1}{r}\hat{V}^r(\cdot)\right) \Rightarrow (0,0) \quad \text{as } r \to \infty. \tag{25}$$

The desired result follows from this and the fact that $\lambda^r \to \lambda$ as $r \to \infty$ (see (13)). □

5.3 Covariance and reflection matrices

In this subsection, we define two matrices that are part of the data for the heavy traffic limit of the workload process. We first define the *covariance matrix* Γ as the $N \times N$ diagonal matrix whose i-th diagonal entry is

$$\Gamma_{ii} \triangleq \lambda_i m_i^2 \left(\alpha_i^2 + \beta_i^2\right), \quad i \in \mathcal{N}. \tag{26}$$

We define the *reflection matrix* R as the $N \times N$ matrix whose entries are

$$R_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \frac{c_i^{\emptyset} - c_i^{\{j\}}}{c_j^{\emptyset}} & \text{if } i \neq j. \end{cases} \tag{27}$$

For example, when $N = 3$, the reflection matrix R is

$$R = \begin{bmatrix} 1 & \frac{c_1^{\emptyset} - c_1^{\{2\}}}{c_2^{\emptyset}} & \frac{c_1^{\emptyset} - c_1^{\{3\}}}{c_3^{\emptyset}} \\ \frac{c_2^{\emptyset} - c_2^{\{1\}}}{c_1^{\emptyset}} & 1 & \frac{c_2^{\emptyset} - c_2^{\{3\}}}{c_3^{\emptyset}} \\ \frac{c_3^{\emptyset} - c_3^{\{1\}}}{c_1^{\emptyset}} & \frac{c_3^{\emptyset} - c_3^{\{2\}}}{c_2^{\emptyset}} & 1 \end{bmatrix}. \tag{28}$$

The matrix R defined by (27) has a special structure in that it satisfies the Harrison–Reiman (HR) condition [10]. We use this structure in proving the convergence of the diffusion-scaled workload process.

Definition 5.1 (Harrison–Reiman (HR) Condition) An $N \times N$ matrix R satisfies the HR condition if $R = I - Q$, where I is the $N \times N$ identity matrix, and the $N \times N$ matrix Q has zeros along the diagonal, all of the entries of Q are non-negative, and Q has spectral radius strictly less than one.

Remark When $R = I - Q$ where Q has zeros on the diagonal and the entries of Q are non-negative, the HR condition is equivalent to the requirement that R is a non-singular M-matrix. Such matrices are discussed for example in Berman and Plemmons [2, Chap. 6].

Lemma 5.3 *The reflection matrix R satisfies the HR condition.*

Proof It is easy to see that an $N \times N$ matrix R satisfies the HR condition if $R = I - P'$ where I is the $N \times N$ identity matrix, P is an $N \times N$ matrix whose diagonal entries are zero, and whose off-diagonal entries are non-negative and such that each row-sum is strictly less than 1. To show that R has this form, note that the diagonal entries of R are all equal to 1 and from the condition (10), the off-diagonal entries are all non-positive. Therefore, it suffices to show that the sum of each column of R is strictly greater than 0. But the sum of the j-th column of R is

$$1 + \sum_{i \in \mathcal{N} \setminus \{j\}} \frac{c_i^{\emptyset} - c_i^{\{j\}}}{c_j^{\emptyset}} = \frac{1}{c_j^{\emptyset}} \left(\sum_{i \in \mathcal{N}} c_i^{\emptyset} - \sum_{i \in \mathcal{N} \setminus \{j\}} c_i^{\{j\}} \right) \tag{29}$$

which is strictly greater than 0 by (9) with $\mathcal{K} = \{j\}$ and since $c_j^{\{j\}} = 0$. □

6 Diffusion approximation—main theorem

6.1 Definition of an SRBM

Before defining an SRBM, we define an $\{\mathcal{F}_t\}$-adapted Brownian motion. Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, a vector $\theta \in \mathbb{R}^N$, an $N \times N$ symmetric, strictly positive-definite matrix Γ, and a probability distribution ν on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$,

an $\{\mathcal{F}_t\}$-Brownian motion with drift vector θ, covariance matrix Γ, and initial distribution ν, is an N-dimensional $\{\mathcal{F}_t\}$-adapted process, X, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that the following hold under P:

(i) X is an N-dimensional Brownian motion whose sample paths are almost surely continuous and that has initial distribution ν,
(ii) $\{X_i(t) - X_i(0) - \theta_i t, \mathcal{F}_t, t \geq 0\}$ is a martingale for $i \in \mathcal{N}$, and
(iii) $\{(X_i(t) - X_i(0) - \theta_i t)(X_j(t) - X_j(0) - \theta_j t) - \Gamma_{ij} t, \mathcal{F}_t, t \geq 0\}$ is a martingale for $i, j \in \mathcal{N}$.

If $\nu = \delta_x$, the unit mass at $x \in \mathbb{R}^N$, we say that X starts from x.

Now, fix $\theta \in \mathbb{R}^N$, Γ an $N \times N$ symmetric strictly positive-definite covariance matrix, R an $N \times N$ matrix satisfying the HR condition, and ν a probability measure on $(\mathbb{R}^N_+, \mathcal{B}(\mathbb{R}^N_+))$. Recall the definition of F_i, $i \in \mathcal{N}$ from Sect. 1.1.

Definition 6.1 (Semimartingale Reflecting Brownian Motion (SRBM)) A semimartingale reflecting Brownian motion (abbreviated as SRBM) with the data $(\mathbb{R}^N_+, \theta, \Gamma, R, \nu)$ is an $\{\mathcal{F}_t\}$-adapted, N-dimensional process, W, defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that

(i) P-a.s., $W(t) = X(t) + RY(t)$ for all $t \geq 0$,
(ii) P-a.s., W has continuous paths and $W(t) \in \mathbb{R}^N_+$ for all $t \geq 0$,
(iii) under P, X is an N-dimensional $\{\mathcal{F}_t\}$-Brownian motion with drift vector θ, covariance matrix Γ, and initial distribution ν,
(iv) Y is an $\{\mathcal{F}_t\}$-adapted, N-dimensional process such that P-a.s. for each $i \in \mathcal{N}$,
  (a) $Y_i(0) = 0$,
  (b) Y_i is continuous and non-decreasing,
  (c) Y_i can only increase when W is on the face F_i, i.e., for all $t \geq 0$,

$$Y_i(t) = \int_0^t 1_{F_i}\big(W(s)\big)\, dY_i(s). \tag{30}$$

When $\nu = \delta_x$ for $x \in \mathbb{R}^N_+$, we may say that W is an SRBM with the data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ that starts from x.

Remark It is known from the work of Harrison and Reiman [10] that when R satisfies the HR condition, there is strong existence and uniqueness (and hence weak existence and uniqueness) for an SRBM given the data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ and the initial distribution ν.

6.2 Main theorem

In this subsection, we state the main theorem and give an outline of the proof. Recall the parameters θ, Γ, and R defined in (13), (26), and (27).

Theorem 6.1 *The diffusion-scaled workload process $\hat{W}^r(\cdot)$ converges in distribution as $r \to \infty$ to an SRBM with data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ that starts from the origin.*

To prove this theorem, we first show that the sequence of processes $\{\hat{W}^r(\cdot)\}$ is C-tight (Sect. 7.3), i.e., any subsequence has a further subsequence that converges weakly to an almost surely continuous limit process. We then show that any weak limit point of such a subsequence is an SRBM with "extensive" data (Sect. 7.4), a notion that we make precise later (see Definition 7.1). For an SRBM with extensive data, there is a direction of reflection associated with each of the $2^N - 1$ boundary faces and there might be pushing in these directions at those boundary faces. In fact, we show that the pushing at boundary faces of dimension $N-2$ or less is negligible (Sect. 7.5) and consequently, the SRBM with extensive data reduces to the simpler form as described in Theorem 6.1. Finally, we show that such an SRBM is unique in law and when combined with the C-tightness, we conclude that the sequence of diffusion-scaled workload processes converges in distribution to an SRBM with data $(\mathbb{R}_+^N, \theta, \Gamma, R)$ that starts from the origin.

7 Proof of the main theorem

7.1 Pre-limit workload process

Throughout this section θ, Γ, and R are given by (13), (26), and (27) respectively. From (5), (6), (11), and (18), the diffusion-scaled workload process can be written so that for $r > 0$, $i \in \mathcal{N}$, and $t \geq 0$,

$$\hat{W}_i^r(t) = \frac{1}{r} V_i\big(E_i^r\big(r^2 t\big)\big) - \frac{T_i^r(r^2 t)}{r}, \tag{31}$$

where

$$T_i^r(t) \triangleq \int_0^t \Lambda_i\big(W^r(s)\big)\, ds. \tag{32}$$

We can rewrite (31) as

$$\begin{aligned}
\hat{W}_i^r(t) &= \frac{1}{r}\big[V_i\big(E_i^r\big(r^2 t\big)\big) - m_i E_i^r\big(r^2 t\big)\big] + \frac{1}{r}\big[m_i E_i^r\big(r^2 t\big) - m_i \lambda_i^r r^2 t\big] \\
&\quad + \lambda_i^r m_i r t - \frac{1}{r} T_i^r\big(r^2 t\big) \\
&= \hat{V}_i^r\big(\bar{E}_i^r(t)\big) + m_i \hat{E}_i^r(t) + \big(\lambda_i^r - \lambda_i\big) m_i r t + \frac{1}{r}\lambda_i m_i \int_0^{r^2 t} ds \\
&\quad - \frac{1}{r}\int_0^{r^2 t} \Lambda_i\big(W^r(s)\big)\, ds \\
&= \hat{X}_i^r(t) + \sum_{\mathcal{K} \subseteq \mathcal{N}} \big(\lambda_i m_i - c_i^{\mathcal{K}}\big) \hat{U}^{r,\mathcal{K}}(t) \\
&= \hat{X}_i^r(t) + \sum_{\emptyset \neq \mathcal{K} \subseteq \mathcal{N}} \big(c_i^{\emptyset} - c_i^{\mathcal{K}}\big) \hat{U}^{r,\mathcal{K}}(t),
\end{aligned} \tag{33}$$

where

$$\hat{X}_i^r(t) \triangleq \hat{V}_i^r\left(\bar{E}_i^r(t)\right) + m_i \hat{E}_i^r(t) + \left(\lambda_i^r - \lambda_i\right) m_i r t, \tag{34}$$

$$\hat{U}^{r,\mathcal{K}}(t) \triangleq \frac{1}{r}\int_0^{r^2 t} 1_{\{\mathcal{K}(W^r(s))=\mathcal{K}\}}\, ds = r\int_0^t 1_{\{\mathcal{K}(\hat{W}^r(s))=\mathcal{K}\}}\, ds \tag{35}$$

and we have used the facts that for any $w \in \mathbb{R}_+^N$,

$$\sum_{\mathcal{K}\subseteq\mathcal{N}} 1_{\{\mathcal{K}(w)=\mathcal{K}\}} = 1, \tag{36}$$

and $c_i^{\emptyset} = \lambda_i m_i$, $i \in \mathcal{N}$. For notational convenience, we will sometimes write $\hat{U}^r(\cdot)$ in place of $\{\hat{U}^{r,\mathcal{K}}(\cdot), \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ in the sequel.

7.2 Convergence to Brownian motion

Our next result shows that the sequence of processes $\{\hat{X}^r(\cdot)\}$ converges in distribution to a Brownian motion. This result will be used in proving that the sequence of processes $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$ is C-tight (see Sect. 7.3) and that any weak limit point of this sequence defines an SRBM with extensive data (see Sect. 7.4).

Lemma 7.1 *The sequence of processes $\{\hat{X}^r(\cdot)\}$ converges in distribution to an N-dimensional Brownian motion that starts from the origin and has drift θ and covariance matrix Γ.*

Proof For all $t \geq 0$, $r > 0$, define

$$\theta(t) \triangleq \theta t, \tag{37}$$

$$\lambda(t) \triangleq \lambda t, \tag{38}$$

and

$$\hat{\theta}_i^r(t) \triangleq r\left(\lambda_i^r - \lambda_i\right) m_i t \quad \text{for all } i \in \mathcal{N}. \tag{39}$$

By Assumption 4.1, $\hat{\theta}^r(\cdot) \to \theta(\cdot)$ u.o.c. as $r \to \infty$. Combining this result with the standard functional central limit theorem (Proposition 5.1), we conclude that the sequence of processes $\{(\hat{E}^r(\cdot), \hat{V}^r(\cdot), \bar{E}^r(\cdot), \hat{\theta}^r(\cdot))\}$ converges in distribution to $(B_E(\cdot), B_V(\cdot), \lambda(\cdot), \theta(\cdot))$ where $B_E(\cdot)$ and $B_V(\cdot)$ are independent N-dimensional driftless Brownian motions starting from the origin with covariance matrices Γ_E and Γ_V given by (22) and (23) respectively. Then from (34), using the random time change lemma of Billingsley [4, p. 151], we conclude that $\{\hat{X}^r(\cdot)\}$ converges in distribution to $B_V(\lambda(\cdot)) + \mathrm{diag}(m) B_E(\cdot) + \theta(\cdot)$, which is an N-dimensional Brownian motion that starts from the origin, has drift θ, and a diagonal covariance matrix whose i-th diagonal entry is

$$\lambda_i m_i^2 \beta_i^2 + m_i^2 \lambda_i \alpha_i^2 = \lambda_i m_i^2 \left(\alpha_i^2 + \beta_i^2\right) = \Gamma_{ii}, \ i \in \mathcal{N}. \qquad \square \tag{40}$$

7.3 C-tightness

Theorem 7.2 *The sequence of processes* $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$ *is C-tight.*

To prove the C-tightness, we use a result from Kang and Williams [14]. In particular, we show that the Assumptions (A1)–(A5) and the Assumption 4.1 of [14] are satisfied by the geometric data and the sequence of processes, $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$, from which the C-tightness follows by Theorem 4.2 of [14]. This verification is carried out below.

7.3.1 Domain

For each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, define $n^{\mathcal{K}}$ as the N-dimensional vector whose i-th element is $1/\sqrt{|\mathcal{K}|}$ if $i \in \mathcal{K}$ and 0 otherwise, that is, for $i \in \mathcal{N}$,

$$n_i^{\mathcal{K}} = \frac{1}{\sqrt{|\mathcal{K}|}} 1_{\{i \in \mathcal{K}\}}. \tag{41}$$

Then for each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $\|n^{\mathcal{K}}\| = 1$. For each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, define $G^{\mathcal{K}}$ as

$$G^{\mathcal{K}} \triangleq \left\{x \in \mathbb{R}^N : \left\langle n^{\mathcal{K}}, x \right\rangle > 0\right\}. \tag{42}$$

Then for each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $G^{\mathcal{K}}$ is an open half-space of $\mathbb{R}^N$ and, therefore, a non-empty domain in $\mathbb{R}^N$. Define the domain G as

$$G \triangleq \bigcap_{\emptyset \neq \mathcal{K} \subseteq \mathcal{N}} G^{\mathcal{K}}. \tag{43}$$

In fact, $G = \{x \in \mathbb{R}^N : x_i > 0 \text{ for all } i \in \mathcal{N}\}$. Hence, $\overline{G} = \mathbb{R}^N_+$. (While the collection $\{G^{\{i\}}, i = 1, 2, \ldots, N\}$ is sufficient to define G, we include the other domains as well since they will have directions of reflection associated with them.)

Lemma 7.3 *The domain G with the representation* (43) *satisfies Assumptions* (A1)–(A3) *of* [14, *Sect.* 3].

Remark Note that the inward unit normal vector for $G^{\mathcal{K}}$ is $n^{\mathcal{K}}$.

Proof Since G is a finite intersection of half-spaces, $\overline{G}$ is a convex polyhedron. We also note that for all $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $\partial G \cap \partial G^{\mathcal{K}} \neq \emptyset$ since the origin is in $\partial G \cap \partial G^{\mathcal{K}}$. Consequently, by Lemma A.3 of [14], we only need to show that G satisfies Assumption (A1) of [14]. Recall that each $G^{\mathcal{K}}$ is a half-space. Therefore, for each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $G^{\mathcal{K}}$ is a non-empty domain, $G^{\mathcal{K}} \neq \mathbb{R}^N$, and the boundary $\partial G^{\mathcal{K}}$ of $G^{\mathcal{K}}$ is C^1. Therefore, the non-empty domain G satisfies Assumption (A1) and hence, Assumptions (A1)–(A3) of [14] hold. □

7.3.2 Reflection vectors

For each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, define the reflection vector $\gamma^{\mathcal{K}}$ such that

$$\gamma_i^{\mathcal{K}} \triangleq c_i^{\emptyset} - c_i^{\mathcal{K}} \quad \text{for each } i \in \mathcal{N}. \tag{44}$$

By this definition, if $i \in \mathcal{K}$, $c_i^{\mathcal{K}} = 0$ and therefore, $\gamma_i^{\mathcal{K}} = c_i^{\emptyset} > 0$. On the other hand, if $i \in \mathcal{K}^c$, $\gamma_i^{\mathcal{K}} = c_i^{\emptyset} - c_i^{\mathcal{K}} \leq 0$ by (10). With this definition of $\{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$, (33) can be rewritten in vector form as

$$\hat{W}^r(t) = \hat{X}^r(t) + \sum_{\emptyset \neq \mathcal{K} \subseteq \mathcal{N}} \gamma^{\mathcal{K}} \hat{U}^{r,\mathcal{K}}(t). \tag{45}$$

Moreover, it is easy to see that the matrix whose columns are given by $\gamma^{\{1\}}, \ldots, \gamma^{\{N\}}$ is

$$R \operatorname{diag}\bigl(c_1^{\emptyset}, c_2^{\emptyset}, \ldots, c_N^{\emptyset}\bigr), \tag{46}$$

where R is the $N \times N$ reflection matrix defined in (27). To facilitate the use Theorem 4.2 of [14], we define the normalized reflection vectors $\{\tilde{\gamma}^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ by

$$\tilde{\gamma}^{\mathcal{K}} \triangleq \frac{\gamma^{\mathcal{K}}}{\|\gamma^{\mathcal{K}}\|}, \tag{47}$$

so that $\|\tilde{\gamma}^{\mathcal{K}}\| = 1$ for all $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$.

Lemma 7.4 *The reflection vectors* $\{\tilde{\gamma}^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ *satisfy Assumptions* (A4)–(A5) *of* [14, *Sect.* 3].

Proof Since the reflection vectors are constant, it is clear that the uniform Lipschitz continuity property of Assumption (A4) of [14] is satisfied. Also, we have normalized the vectors to be of unit length.

To verify (A5), we need to show that there is a constant $a \in (0, 1)$ such that for each $x \in \partial G$, there are non-negative constants $(b_{\mathcal{L}}(x) : \emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x))$ and $(d_{\mathcal{L}}(x) : \emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x))$ such that

$$\sum_{\emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x)} b_{\mathcal{L}}(x) = 1, \tag{48}$$

$$\min_{\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)} \left\langle \sum_{\emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x)} b_{\mathcal{L}}(x) n^{\mathcal{L}}, \tilde{\gamma}^{\mathcal{M}} \right\rangle \geq a, \tag{49}$$

$$\sum_{\emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x)} d_{\mathcal{L}}(x) = 1, \tag{50}$$

$$\min_{\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)} \left\langle \sum_{\emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x)} d_{\mathcal{L}}(x) \tilde{\gamma}^{\mathcal{L}}, n^{\mathcal{M}} \right\rangle \geq a. \tag{51}$$

To this end, for any $x \in \partial G$ and $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}(x)$, set

$$b_{\mathcal{L}}(x) \triangleq 1_{\{\mathcal{L}=\mathcal{K}(x)\}} \tag{52}$$

and

$$d_{\mathcal{L}}(x) \triangleq 1_{\{\mathcal{L}=\mathcal{K}(x)\}}. \tag{53}$$

Then

$$\sum_{\emptyset\neq\mathcal{L}\subseteq\mathcal{K}(x)} b_{\mathcal{L}}(x) n^{\mathcal{L}} = n^{\mathcal{K}(x)} \tag{54}$$

and

$$\sum_{\emptyset\neq\mathcal{L}\subseteq\mathcal{K}(x)} d_{\mathcal{L}}(x) \tilde{\gamma}^{\mathcal{L}} = \tilde{\gamma}^{\mathcal{K}(x)}. \tag{55}$$

Therefore, to verify that Assumption (A5) of [14] is satisfied, we only need to verify that for each $x \in \partial G$ and $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)$, $\langle n^{\mathcal{K}(x)}, \tilde{\gamma}^{\mathcal{M}} \rangle$ and $\langle \tilde{\gamma}^{\mathcal{K}(x)}, n^{\mathcal{M}} \rangle$ are bounded below by a strictly positive constant not depending on x or $\mathcal{M}$. We first verify that $\langle \tilde{\gamma}^{\mathcal{K}(x)}, n^{\mathcal{M}} \rangle$ has such a lower bound. From (44) and (47), for all $i \in \mathcal{K}(x)$,

$$\tilde{\gamma}_i^{\mathcal{K}(x)} = \frac{c_i^{\emptyset}}{\|\gamma^{\mathcal{K}(x)}\|} > 0. \tag{56}$$

Thus, using (41), for each $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)$,

$$\begin{aligned} \langle \tilde{\gamma}^{\mathcal{K}(x)}, n^{\mathcal{M}} \rangle &= \frac{1}{\sqrt{|\mathcal{M}|}} \sum_{i\in\mathcal{M}} \tilde{\gamma}_i^{\mathcal{K}(x)} \\ &\geq \frac{\min_{i\in\mathcal{M}} c_i^{\emptyset}}{\sqrt{|\mathcal{M}|}\|\gamma^{\mathcal{K}(x)}\|} \\ &\geq \frac{\min_{i\in\mathcal{N}} c_i^{\emptyset}}{\sqrt{N}\max_{\emptyset\neq\mathcal{L}\subseteq\mathcal{N}}\|\gamma^{\mathcal{L}}\|} \\ &> 0, \end{aligned} \tag{57}$$

where the second inequality follows because we are taking the minimum over a larger set in the third line and for all $x \in \partial G$, $|\mathcal{K}(x)| \leq N$. Next, we show that for each $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)$, $\langle n^{\mathcal{K}(x)}, \tilde{\gamma}^{\mathcal{M}} \rangle$ has a uniform strictly positive lower bound. To this

end, we have for $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}(x)$,

$$\begin{aligned}
\langle n^{\mathcal{K}(x)}, \tilde{\gamma}^{\mathcal{M}} \rangle &= \frac{1}{\sqrt{|\mathcal{K}(x)|}} \sum_{i \in \mathcal{K}(x)} \gamma_i^{\mathcal{M}} / \|\gamma^{\mathcal{M}}\| \\
&= \frac{1}{\sqrt{|\mathcal{K}(x)|}} \sum_{i \in \mathcal{K}(x)} \left(c_i^{\emptyset} - c_i^{\mathcal{M}}\right) / \|\gamma^{\mathcal{M}}\| \\
&= \frac{1}{\sqrt{|\mathcal{K}(x)|}} \left[\sum_{i \in \mathcal{N}} \left(c_i^{\emptyset} - c_i^{\mathcal{M}}\right) - \sum_{i \in (\mathcal{K}(x))^c} \left(c_i^{\emptyset} - c_i^{\mathcal{M}}\right) \right] \Big/ \|\gamma^{\mathcal{M}}\| \\
&\geq \frac{1}{\sqrt{|\mathcal{K}(x)|}} \sum_{i \in \mathcal{N}} \left(c_i^{\emptyset} - c_i^{\mathcal{M}}\right) / \|\gamma^{\mathcal{M}}\| \\
&\geq \frac{1}{\sqrt{N}} \min_{\emptyset \neq \mathcal{L} \subseteq \mathcal{N}} \sum_{i \in \mathcal{N}} \left(c_i^{\emptyset} - c_i^{\mathcal{L}}\right) \Big/ \max_{\emptyset \neq \mathcal{L} \subseteq \mathcal{N}} \|\gamma^{\mathcal{L}}\| \\
&> 0, \qquad (58)
\end{aligned}$$

where the first inequality follows from (10) with $\mathcal{M}$ in place of $\mathcal{K}$ and the last inequality follows from (9). □

Remark Properties (9) and (10) were used critically in deriving (58). This property (58) is then used critically in proving Lemmas 7.8 and 7.9 below, which in turn lead to proving our main technical result Theorem 7.7.

Proof of Theorem 7.2 For each $r > 0$, let

$$\hat{Z}^r \triangleq \left(\hat{W}^r, \hat{X}^r, \hat{U}^r\right). \qquad (59)$$

To prove the C-tightness of $\{\hat{Z}^r\}$, we first verify that Assumption 4.1 of [14, Sect. 4] is satisfied.

For any $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$ and $r > 0$, let $\gamma^{r,\mathcal{K}}(y, x) \triangleq \tilde{\gamma}^{\mathcal{K}}$ for all $x, y \in \mathbb{R}^N$, $\alpha^r \triangleq 0 \in \mathbb{D}^N$, $\beta^r = \{\beta^{r,\mathcal{K}} : \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ where $\beta^{r,\mathcal{K}} \triangleq 0 \in \mathbb{D}$, $\delta^r = 1/r$, and $\hat{Y}^r = \{\hat{Y}^{r,\mathcal{K}} : \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ where $\hat{Y}^{r,\mathcal{K}} = \|\gamma^{\mathcal{K}}\| \hat{U}^{r,\mathcal{K}}$. With these definitions, the conditions (i)–(vi) of Assumption 4.1 of [14] are satisfied with $\{(\hat{W}^r, \hat{X}^r, \hat{Y}^r)\}$ in place of $\{(W^n, X^n, Y^n)\}$. Here

$$\hat{Y}^{r,\mathcal{K}}(t) = \int_0^t 1_{\{\mathrm{dist}(\hat{W}^r(s), \partial G^{\mathcal{K}} \cap \partial G) \leq \delta^r\}} \, d\hat{Y}^{r,\mathcal{K}}(s) \qquad (60)$$

because $\hat{U}^{r,\mathcal{K}}$ can increase only when $\hat{W}^r$ is on $\partial G^{\mathcal{K}} \cap \partial G$ (see (35)), and $\{\hat{X}^r\}$ is C-tight by Lemma 7.1. It then follows from Theorem 4.2 of [14, Sect. 4], that $\{(\hat{W}^r, \hat{X}^r, \hat{Y}^r)\}$, and hence $\{\hat{Z}^r\}$, is C-tight and the theorem is proved. □

7.4 SRBM with extensive data

In this subsection, we show that any weak limit point of the sequence of processes $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$ defines an SRBM with extensive data. Before presenting the

theorem and its proof, we need to define an SRBM with extensive data. The following definition is adapted from the definition in [14, Sect. 2]. Recall the definition of G from (43), θ and Γ from (13) and (26), and $\{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ from (44). Let ν be a probability measure on $(\overline{G}, \mathcal{B}(\overline{G}))$, where $\mathcal{B}(\overline{G})$ denotes the σ-algebra of Borel subsets of the closure, $\overline{G}$, of G.

Definition 7.1 (SRBM with Extensive Data) An SRBM with the extensive data $(\overline{G}, \theta, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}, \nu)$ is an $\{\mathcal{F}_t\}$-adapted, N-dimensional process W defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that

(i) P-a.s., for all $t \geq 0$,

$$W(t) = X(t) + \sum_{\emptyset \neq \mathcal{K} \subseteq \mathcal{N}} \gamma^{\mathcal{K}} U^{\mathcal{K}}(t), \tag{61}$$

(ii) P-a.s., W has continuous paths and $W(t) \in \overline{G}$ for all $t \geq 0$,
(iii) under P, X is an N-dimensional $\{\mathcal{F}_t\}$-Brownian motion with drift vector θ, covariance matrix Γ, and initial distribution ν,
(iv) for each $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $U^{\mathcal{K}}$ is an $\{\mathcal{F}_t\}$-adapted, one-dimensional process such that P-a.s.,
  (a) $U^{\mathcal{K}}(0) = 0$,
  (b) $U^{\mathcal{K}}$ is continuous and non-decreasing,
  (c) for all $t \geq 0$,

$$U^{\mathcal{K}}(t) = \int_0^t 1_{\{W(s) \in \partial G^{\mathcal{K}} \cap \partial G\}} \, dU^{\mathcal{K}}(s). \tag{62}$$

When $\nu = \delta_x$, for $x \in \overline{G}$, we say that W is an SRBM associated with the data $(\overline{G}, \theta, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\})$ that starts from x.

Remark We have introduced the terminology "extensive" data in this work to differentiate between the above SRBM which has reflection on the lower-dimensional faces and the simpler SRBM introduced in Definition 6.1.

With this definition in hand, we can now state and prove the main result of this subsection.

Theorem 7.5 *Any weak limit point* $(W(\cdot), X(\cdot), U(\cdot))$ *of the sequence of processes* $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$ *defines an SRBM,* W, *with the extensive data* $(\overline{G}, \theta, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\})$ *that starts from the origin.*

We need the following lemma for our proof of Theorem 7.5. So as to not disrupt the flow of this section, we defer the proof of this lemma to Appendix A.

Lemma 7.6 *Suppose that* $Z = (W, X, U)$ *is a weak limit point of the sequence* $\{(\hat{W}^r, \hat{X}^r, \hat{U}^r)\}$. *Let* $\mathcal{F}_t = \sigma\{Z(s) : 0 \leq s \leq t\}, t \geq 0$. *Then* $\{X(t) - X(0) - \theta t, \mathcal{F}_t, t \geq 0\}$ *is a martingale.*

Proof See Appendix A. □

Proof of Theorem 7.5 The result follows from Theorem 4.3 of [14] provided Assumption 4.1 and Assumptions (vi)′ and (vii) of Theorem 4.3 in [14] hold for $\{(\hat{W}^r, \hat{X}^r, \hat{Y}^r)\}$ where $\hat{Y}^r = \{\hat{Y}^{r,\mathcal{K}} : \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ and $\hat{Y}^{r,\mathcal{K}} = \|\gamma^{\mathcal{K}}\|\hat{U}^{r,\mathcal{K}}$. Our proof of Theorem 7.2 shows that Assumption 4.1 of [14] holds. Assumption (vi)′ of Theorem 4.3 in [14] follows immediately from Lemma 7.1. Assumption (vii) of Theorem 4.3 in [14] follows from Lemma 7.6 and the simple relationship between $\hat{U}^{r,\mathcal{K}}$ and $\hat{Y}^{r,\mathcal{K}}$. □

7.5 Pushing on the lower-dimensional faces

In this subsection, we show a result, which when combined with Theorem 7.5 implies that for any weak limit point, $(W(\cdot), X(\cdot)), U(\cdot))$, of the sequence of processes $\{(\hat{W}^r(\cdot), \hat{X}^r(\cdot), \hat{U}^r(\cdot))\}$, the amount of pushing done by U at any of the faces of ∂G of dimension $N-2$ or less is negligible. Formally, we prove the following. For this, recall that for any $\mathcal{K} \subseteq \mathcal{N}$, $F_{\mathcal{K}}$ is defined in (1).

Theorem 7.7 *Let* $(W(\cdot), X(\cdot), U(\cdot))$ *define an SRBM,* $W(\cdot)$*, with extensive data* $(\overline{G}, \theta, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\})$ *that starts from the origin. Then for each* $\mathcal{K} \subseteq \mathcal{N}$*,* $|\mathcal{K}| \geq 2$*, for each* $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}$*,*

$$\int_0^\infty 1_{F_{\mathcal{K}}}\big(W(s)\big)\, dU^{\mathcal{L}}(s) = 0 \quad \textit{almost surely.} \tag{63}$$

Consequently, almost surely,

$$W(t) = X(t) + \sum_{i \in \mathcal{N}} \gamma^{\{i\}} U^{\{i\}}(t), \quad t \geq 0. \tag{64}$$

Our proof of Theorem 7.7 is a generalization of the proof of the main theorem in Reiman and Williams [20]. However, there are some differences, since in [20] there were only N directions of reflection—one for each $(N-1)$-dimensional boundary face, whereas here there are $2^N - 1$, one for each boundary face. We prove the theorem in three steps. We assume that $N \geq 2$, otherwise the result is vacuous and hence trivially true. We first prove that for the case of zero drift ($\theta = 0$) the amount of pushing done when W is at the origin is negligible (see Lemma 7.8). We then use a backwards induction argument on $|\mathcal{K}|$ to show that for the case of zero drift the amount of pushing done on $F_{\mathcal{K}}$ is negligible provided $|\mathcal{K}| \geq 2$ (see Lemma 7.9). Finally, using a Girsanov transformation, the result is extended to all constant drifts θ (see Lemma 7.10). We then complete the proof.

Lemma 7.8 *Suppose* (W, X, U) *is as in the hypothesis of Theorem 7.7 and* $\theta = 0$*. Then for* $N \geq 2$ *and* $\mathcal{K} = \mathcal{N}$*, (63) holds for all* $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$*.*

Proof From the semimartingale representation (61) of W and Itô's formula, for any function f that is twice continuously differentiable in some domain containing $\overline{G}$,

we have almost surely for all $t \geq 0$:

$$f\big(W(t)\big) - f\big(W(0)\big) = \int_0^t \big\langle \nabla f\big(W(s)\big), dX(s)\big\rangle + \sum_{\emptyset \neq \mathcal{L} \subseteq \mathcal{N}} \int_0^t \big\langle \gamma^{\mathcal{L}}, \nabla f\big(W(s)\big)\big\rangle dU^{\mathcal{L}}(s) + \int_0^t Lf\big(W(s)\big)\, ds, \tag{65}$$

where

$$Lf = \frac{1}{2} \sum_{i=1}^{N} \Gamma_{ii} \frac{\partial^2 f}{\partial x_i^2}. \tag{66}$$

We shall substitute functions into (65) that allow us to estimate the left-hand side of (63). Each such function will be L-harmonic in some domain containing $\overline{G}$ and for each $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$, its directional derivative in the direction of $\gamma^{\mathcal{L}}$ will be bounded below on $\overline{G}$ and be very large and positive near the origin. These functions are chosen such that they are uniformly bounded on compact subsets of $\overline{G}$.

Define

$$\tilde{\beta} \triangleq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}_+^N. \tag{67}$$

Then from (58) with $x = 0$, we have for all $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$,

$$\big\langle \gamma^{\mathcal{L}}, \tilde{\beta} \big\rangle = \sqrt{N} \|\gamma^{\mathcal{L}}\| \big\langle \tilde{\gamma}^{\mathcal{L}}, n^{\mathcal{N}} \big\rangle > 0. \tag{68}$$

Therefore, there exists a vector $\beta \in \mathbb{R}_+^N$ having all components strictly positive such that for all $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$,

$$\big\langle \gamma^{\mathcal{L}}, \beta \big\rangle \triangleq \delta^{\mathcal{L}} \in [1, \infty). \tag{69}$$

Define

$$\alpha \triangleq \Gamma \beta. \tag{70}$$

Note that $\alpha_i > 0$ for all $i \in \mathcal{N}$. For each $x \in \overline{G} = \mathbb{R}_+^N$ and $s \in (0, 1)$, define a squared distance function:

$$\begin{aligned} d^2(x, s) &\triangleq (x + s\alpha)' \Gamma^{-1} (x + s\alpha) \\ &= x' \Gamma^{-1} x + 2s\alpha' \Gamma^{-1} x + s^2 \alpha' \Gamma^{-1} \alpha \\ &= x' \Gamma^{-1} x + 2s\beta' x + s^2 \alpha' \Gamma^{-1} \alpha \\ &\geq s^2 \hat{\alpha}, \end{aligned} \tag{71}$$

where

$$\hat{\alpha} \triangleq \alpha' \Gamma^{-1} \alpha = \beta' \Gamma \beta > 0. \tag{72}$$

We have used the facts that Γ (and hence Γ^{-1}) is symmetric and strictly positive definite, and $\beta, x \in \mathbb{R}^N_+$. Then for each fixed $\varepsilon \in (0,1)$,

$$\phi_\varepsilon(x) \triangleq \begin{cases} \frac{1}{2-N} \int_\varepsilon^1 s^{N-2} (d^2(x,s))^{\frac{2-N}{2}} ds, & N \geq 3, \\ \frac{1}{2} \int_\varepsilon^1 \ln(d^2(x,s))\, ds, & N = 2, \end{cases} \tag{73}$$

is twice continuously differentiable in some domain containing $\overline{G}$, and on each compact subset of $\overline{G}$, it is bounded, uniformly in ε. Moreover, since the integrand in (73), for a fixed s, is L-harmonic as a function of $x \in \mathbb{R}^N \setminus \{-s\alpha\}$, it is readily verified that for each $\varepsilon \in (0,1)$,

$$L\phi_\varepsilon = 0 \tag{74}$$

in some domain containing $\overline{G}$.

For the verification of the directional derivative properties of ϕ_ε, for each $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$, let

$$u^{\mathcal{L}} \triangleq \Gamma^{-1} \gamma^{\mathcal{L}}. \tag{75}$$

Then by (69) and (70),

$$\langle u^{\mathcal{L}}, \alpha \rangle = \langle \Gamma^{-1} \gamma^{\mathcal{L}}, \alpha \rangle = \big(\gamma^{\mathcal{L}}\big)' \Gamma^{-1} \alpha = \big(\gamma^{\mathcal{L}}\big)' \beta = \delta^{\mathcal{L}} \geq 1. \tag{76}$$

Combining (76) with

$$\nabla \phi_\varepsilon(x) = \int_\varepsilon^1 s^{N-2} \Gamma^{-1} (x + s\alpha) \big(d^2(x,s)\big)^{-N/2} ds, \tag{77}$$

we get

$$\big\langle \gamma^{\mathcal{L}}, \nabla \phi_\varepsilon(x) \big\rangle = \int_\varepsilon^1 s^{N-2} \big(\langle u^{\mathcal{L}}, x \rangle + s\delta^{\mathcal{L}}\big) \big(d^2(x,s)\big)^{-N/2} ds. \tag{78}$$

Let

$$\xi^{\mathcal{L}} \triangleq \frac{\delta^{\mathcal{L}}}{\|u^{\mathcal{L}}\|}. \tag{79}$$

Then for $\varepsilon \in (0,1)$ and $x \in \overline{G}$ satisfying $\|x\| < \varepsilon \xi^{\mathcal{L}}$, we have $|\langle u^{\mathcal{L}}, x \rangle| < \varepsilon \delta^{\mathcal{L}}$ and for $s > \varepsilon$,

$$\begin{aligned} d^2(x,s) &\leq \big\| \Gamma^{-1} \big\| \, \|x + s\alpha\|^2 \\ &\leq \big\| \Gamma^{-1} \big\| \big(\|x\| + \|s\alpha\|\big)^2 \\ &\leq \big\| \Gamma^{-1} \big\| \big(\xi^{\mathcal{L}} + \|\alpha\|\big)^2 s^2, \end{aligned} \tag{80}$$

where $\|\Gamma^{-1}\|$ denotes the norm of Γ^{-1} as an operator from $\mathbb{R}^N$ to $\mathbb{R}^N$ with the Euclidean norm. Setting

$$\zeta^{\mathcal{L}} \triangleq \delta^{\mathcal{L}}\big(\|\Gamma^{-1}\|\big(\xi^{\mathcal{L}} + \|\alpha\|\big)^2\big)^{-N/2} \tag{81}$$

and substituting the above in (78) yields:

$$\begin{aligned}\big\langle \gamma^{\mathcal{L}}, \nabla\phi_\varepsilon(x)\big\rangle &\geq \zeta^{\mathcal{L}} \int_\varepsilon^1 s^{N-2}(s-\varepsilon)s^{-N}\,ds \\ &\geq -\zeta^{\mathcal{L}}[\ln\varepsilon + 1]\end{aligned} \tag{82}$$

for all $x \in \overline{G}$ satisfying $\|x\| < \varepsilon\xi^{\mathcal{L}}$. Note that for small ε, the term in the last line above is large and positive.

Now for any $x \in \overline{G}$, $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$,

$$\big\langle \gamma^{\mathcal{L}}, \nabla\phi_\varepsilon(x)\big\rangle = -\delta^{\mathcal{L}} \int_\varepsilon^1 s^{N-2}\big(\rho^{\mathcal{L}}(x) - s\big)\big(d^2(x,s)\big)^{-N/2}\,ds, \tag{83}$$

where

$$\rho^{\mathcal{L}}(x) \triangleq -\frac{\langle u^{\mathcal{L}}, x\rangle}{\delta^{\mathcal{L}}}. \tag{84}$$

If $\rho^{\mathcal{L}}(x) \leq \varepsilon$, then the right-hand side of (83) is non-negative. Thus, to obtain a lower bound for $\langle \gamma^{\mathcal{L}}, \nabla\phi_\varepsilon(x)\rangle$ on $\overline{G}$, it suffices to consider $x \in \overline{G}$ such that $\rho^{\mathcal{L}}(x) > \varepsilon$. For such x,

$$\begin{aligned}&\int_\varepsilon^1 s^{N-2}\big(\rho^{\mathcal{L}}(x) - s\big)\big(d^2(x,s)\big)^{-N/2}\,ds \\ &\quad\leq \int_\varepsilon^{\rho^{\mathcal{L}}(x)} s^{N-2}\big(\rho^{\mathcal{L}}(x) - s\big)\big(d^2(x,s)\big)^{-N/2}\,ds \\ &\quad\leq \big(\rho^{\mathcal{L}}(x) - \varepsilon\big) \max_{s\in[\varepsilon,\rho^{\mathcal{L}}(x)]} \frac{\rho^{\mathcal{L}}(x) - s}{d^2(x,s)} \max_{s\in[\varepsilon,\rho^{\mathcal{L}}(x)]} \frac{s^{N-2}}{(d^2(x,s))^{(N-2)/2}}.\end{aligned} \tag{85}$$

Since $d^2(x,s)$ is quadratic in s with positive coefficients, the first maximum above is achieved at $s = \varepsilon$, and by (71), the second maximum is crudely dominated by $\hat{\alpha}^{(2-N)/2}$. Thus, the last term of (85) is bounded from above by

$$\frac{(\rho^{\mathcal{L}}(x) - \varepsilon)^2}{d^2(x,\varepsilon)}\hat{\alpha}^{(2-N)/2}. \tag{86}$$

Since Γ^{-1} is strictly positive definite, there is an $\eta > 0$ such that $x'\Gamma^{-1}x \geq \eta\,\|x\|^2$ and so (see (71)),

$$\begin{aligned}d^2(x,\varepsilon) &\geq \eta\,\|x\|^2 + \varepsilon^2\hat{\alpha} \\ &\geq (\eta \wedge \hat{\alpha})\big(\|x\|^2 + \varepsilon^2\big).\end{aligned} \tag{87}$$

On the other hand, by the definition of $\rho^{\mathcal{L}}(x)$,

$$\begin{aligned}\bigl(\rho^{\mathcal{L}}(x)-\varepsilon\bigr)^2 &\le 2\bigl(\bigl(\rho^{\mathcal{L}}(x)\bigr)^2+\varepsilon^2\bigr)\\ &\le 2\bigl(\bigl\|u^{\mathcal{L}}\bigr\|^2\|x\|^2\bigl(\delta^{\mathcal{L}}\bigr)^{-2}+\varepsilon^2\bigr)\\ &\le 2\max\bigl(\bigl\|u^{\mathcal{L}}\bigr\|^2\bigl(\delta^{\mathcal{L}}\bigr)^{-2},1\bigr)\bigl(\|x\|^2+\varepsilon^2\bigr).\end{aligned}\tag{88}$$

It follows from (87) and (88) that (85) is bounded from above by a constant not depending on x or ε. Hence, there is a $\tilde{\zeta}^{\mathcal{L}}\ge 0$ such that for all $x\in\overline{G}$ and $\varepsilon\in(0,1)$,

$$\bigl\langle\gamma^{\mathcal{L}},\nabla\phi_\varepsilon(x)\bigr\rangle\ge-\tilde{\zeta}^{\mathcal{L}}.\tag{89}$$

We are now ready to prove that when $\mathcal{K}=\mathcal{N}$, (63) holds almost surely for each $\emptyset\ne\mathcal{L}\subseteq\mathcal{N}$. For each positive integer m, define

$$T_m\triangleq\inf\bigl\{t\ge0:\bigl\|W(t)\bigr\|\ge m\text{ or }U^{\mathcal{L}}(t)\ge m\text{ for some }\emptyset\ne\mathcal{L}\subseteq\mathcal{N}\bigr\}\wedge m.\tag{90}$$

Replacing f by ϕ_ε and t by T_m in (65), we see from (74) that almost surely:

$$\begin{aligned}\phi_\varepsilon\bigl(W(T_m)\bigr)-\phi_\varepsilon\bigl(W(0)\bigr)&=\int_0^{T_m}\bigl\langle\nabla\phi_\varepsilon\bigl(W(s)\bigr),dX(s)\bigr\rangle\\ &\quad+\sum_{\emptyset\ne\mathcal{L}\subseteq\mathcal{N}}\int_0^{T_m}\bigl\langle\gamma^{\mathcal{L}},\nabla\phi_\varepsilon\bigl(W(s)\bigr)\bigr\rangle dU^{\mathcal{L}}(s).\end{aligned}\tag{91}$$

Since ϕ_ε and its first derivatives are bounded on each compact subset of $\overline{G}$, by the definition of the stopping time T_m and since $\theta=0$, the stochastic integral with respect to dX in (91) has zero expectation. Thus, taking expectations in (91) yields:

$$\begin{aligned}&\mathbf{E}\bigl[\phi_\varepsilon\bigl(W(T_m)\bigr)-\phi_\varepsilon\bigl(W(0)\bigr)\bigr]\\ &=\sum_{\emptyset\ne\mathcal{L}\subseteq\mathcal{N}}\mathbf{E}\biggl[\int_0^{T_m}\bigl\langle\gamma^{\mathcal{L}},\nabla\phi_\varepsilon\bigl(W(s)\bigr)\bigr\rangle dU^{\mathcal{L}}(s)\biggr]\\ &\ge-(\ln\varepsilon+1)\sum_{\emptyset\ne\mathcal{L}\subseteq\mathcal{N}}\zeta^{\mathcal{L}}\mathbf{E}\biggl[\int_0^{T_m}1_{\{\|W(s)\|<\varepsilon\xi^{\mathcal{L}}\}}\,dU^{\mathcal{L}}(s)\biggr]\\ &\quad-\sum_{\emptyset\ne\mathcal{L}\subseteq\mathcal{N}}\tilde{\zeta}^{\mathcal{L}}\mathbf{E}\bigl[U^{\mathcal{L}}(T_m)\bigr],\end{aligned}\tag{92}$$

where the lower bounds (82) and (89) have been used to obtain the last inequality. Now, the left-hand side of (92) is bounded as $\varepsilon\downarrow0$, since for $\varepsilon\in(0,1)$, ϕ_ε is uniformly bounded on compact subsets of $\overline{G}$. Also, the last sum in (92) is positive and independent of ε. Thus, dividing (92) by $-(\ln\varepsilon+1)$ and letting $\varepsilon\downarrow0$ yields:

$$\lim_{\varepsilon\downarrow0}\sum_{\emptyset\ne\mathcal{L}\subseteq\mathcal{N}}\zeta^{\mathcal{L}}\mathbf{E}\biggl[\int_0^{T_m}1_{\{\|W(s)\|<\varepsilon\xi^{\mathcal{L}}\}}\,dU^{\mathcal{L}}(s)\biggr]\le0.\tag{93}$$

Since each term in the above sum is non-negative and $\zeta^{\mathcal{L}} > 0$, it follows by Fatou's lemma that for each $\emptyset \neq \mathcal{L} \subseteq \mathcal{N}$,

$$\int_0^{T_m} 1_{F_{\mathcal{N}}}\big(W(s)\big)\, dU^{\mathcal{L}}(s) = 0 \quad \text{almost surely.} \tag{94}$$

Letting $m \to \infty$ yields the desired result. □

Lemma 7.9 *Suppose* (W, X, U) *is as in the hypothesis of Theorem* 7.7 *and* $\theta = 0$. *Then* (63) *holds for all* $\emptyset \neq \mathcal{L} \subseteq \mathcal{K} \subseteq \mathcal{N}$ *where* $|\mathcal{K}| \geq 2$.

Proof Our proof is by backwards induction on $|\mathcal{K}|$. Without loss of generality, we assume $N \geq 2$ (otherwise there is no $\mathcal{K} \subseteq \mathcal{N}$ with $|\mathcal{K}| \geq 2$). By Lemma 7.8, the result holds for $|\mathcal{K}| = N$ in which case the only possible $\mathcal{K}$ is $\mathcal{K} = \mathcal{N}$. Fix $2 \leq k < N$ and suppose that (63) holds for all $\mathcal{K} \subseteq \mathcal{N}$ and $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}$, such that $k < |\mathcal{K}| \leq N$. Fix some $\mathcal{K} \subseteq \mathcal{N}$ such that $|\mathcal{K}| = k$. We need to show that for all $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}$,

$$\int_0^{\infty} 1_{F_{\mathcal{K}}}\big(W(s)\big)\, dU^{\mathcal{L}}(s) = 0 \quad \text{almost surely.} \tag{95}$$

To this end, fix $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}$ and recall that for any non-empty set $\mathcal{L} \subseteq \mathcal{N}$ and any $w \in \mathbb{R}^N$, $w_{\mathcal{L}}$ denotes the vector whose components are those of w with indices in $\mathcal{L}$. Then

$$\begin{aligned}\int_0^{\infty} 1_{F_{\mathcal{K}}}\big(W(s)\big)\, dU^{\mathcal{L}}(s) &= \int_0^{\infty} 1_{\{\mathcal{K}(W(s))=\mathcal{K}\}}\, dU^{\mathcal{L}}(s) \\ &\quad + \int_0^{\infty} 1_{\{W(s)\in \cup_{\mathcal{K}\subsetneq\mathcal{M}} F_{\mathcal{M}}\}}\, dU^{\mathcal{L}}(s) \\ &\overset{\text{a.s.}}{=} \int_0^{\infty} 1_{\{W_{\mathcal{K}}(s)=0,\, W_{\mathcal{K}^c}(s)>0\}}\, dU^{\mathcal{L}}(s), \end{aligned} \tag{96}$$

where by the induction assumption the second integral on the right-hand side of the first equation is almost surely zero. Thus, by monotone convergence, it suffices to prove that for each $\eta \in \mathbb{R}_+^{N-k}$, satisfying $\eta > 0$, we have

$$\int_0^{\infty} 1_{\{W_{\mathcal{K}}(s)=0,\, W_{\mathcal{K}^c}(s)>\eta\}}\, dU^{\mathcal{L}}(s) = 0 \quad \text{almost surely.} \tag{97}$$

For this, fix an $\eta \in \mathbb{R}_+^{N-k}$ with $\eta > 0$, and define a sequence of stopping times $\{T_m\}_{m=1}^{\infty}$ as follows. (Here, for notational convenience, we regard the entries in η as being indexed by $i \in \mathcal{K}^c$.)

$$\begin{aligned} T_0 &\triangleq 0, \\ T_1 &\triangleq \inf\big\{s \geq 0 : W_i(s) < \eta_i/2 \text{ for some } i \in \mathcal{K}^c\big\}, \\ T_2 &\triangleq \inf\big\{s \geq T_1 : W_{\mathcal{K}^c}(s) > \eta\big\}, \end{aligned} \tag{98}$$

and for $m \geq 1$,

$$\begin{aligned} T_{2m+1} &\triangleq \inf\bigl\{t \geq T_{2m} : W_i(s) < \eta_i/2 \text{ for some } i \in \mathcal{K}^c\bigr\}, \\ T_{2m+2} &\triangleq \inf\{t \geq T_{2m+1} : W_{\mathcal{K}^c} > \eta\}. \end{aligned} \tag{99}$$

By the continuity of the paths of W, $T_m \to \infty$ as $m \to \infty$, and we have almost surely:

$$\int_0^\infty 1_{\{W_{\mathcal{K}}(s)=0,\, W_{\mathcal{K}^c}(s)>\eta\}}\, dU^{\mathcal{L}}(s) \leq \sum_{m=0}^{\infty} \int_{T_{2m}}^{T_{2m+1}} 1_{\{W_{\mathcal{K}}(s)=0\}}\, dU^{\mathcal{L}}(s). \tag{100}$$

Consider $m \geq 0$. Then on $\{T_{2m} < \infty\}$, for $\emptyset \neq \mathcal{M} \subseteq \mathcal{N}$, $\mathcal{M} \nsubseteq \mathcal{K}$, $U^{\mathcal{M}}$ can increase only when $W_{\mathcal{M}} = 0$ and so, almost surely, for all such $\mathcal{M}$,

$$U^{\mathcal{M}}(t + T_{2m}) - U^{\mathcal{M}}(T_{2m}) = 0 \quad \text{for all } t \in [0, T_{2m+1} - T_{2m}]. \tag{101}$$

Thus, on $\{T_{2m} < \infty\}$, we have almost surely for all $t \in [0, T_{2m+1} - T_{2m}]$

$$\begin{aligned} &W_{\mathcal{K}}(t + T_{2m}) - W_{\mathcal{K}}(T_{2m}) \\ &\quad = X_{\mathcal{K}}(t + T_{2m}) - X_{\mathcal{K}}(T_{2m}) + \sum_{\emptyset \neq \mathcal{M} \subseteq \mathcal{K}} \gamma_{\mathcal{K}}^{\mathcal{M}} \bigl(U^{\mathcal{M}}(t + T_{2m}) - U^{\mathcal{M}}(T_{2m})\bigr). \end{aligned} \tag{102}$$

Then Itô's formula, (65), holds on $\{T_{2m} < \infty\}$ for $f \in C^2(\mathbb{R}_+^k)$ with $(X, \{U^{\mathcal{L}} : \emptyset \neq \mathcal{L} \subseteq \mathcal{N}\}, W)$ and $\{\gamma^{\mathcal{L}} : \emptyset \neq \mathcal{L} \subseteq \mathcal{N}\}$ replaced by

$$(X_{\mathcal{K}}, \{U^{\mathcal{M}} : \emptyset \neq \mathcal{M} \subseteq \mathcal{K}\}, W_{\mathcal{K}})((\cdot + T_{2m}) \wedge T_{2m+1}) \quad \text{and} \quad \{\gamma_{\mathcal{K}}^{\mathcal{M}} : \emptyset \neq \mathcal{M} \subseteq \mathcal{K}\},$$

and with

$$Lf = \frac{1}{2} \sum_{i \in \mathcal{K}} \Gamma_{ii} \frac{\partial^2 f}{\partial x_i^2}. \tag{103}$$

The same proof as in Lemma 7.8, but with the dimension reduced from N to $k = |\mathcal{K}|$, shows that

$$\sum_{\emptyset \neq \mathcal{M} \subseteq \mathcal{K}} 1_{\{T_{2m} < \infty\}} \int_{T_{2m}}^{T_{2m+1}} 1_{\{W_{\mathcal{K}}(s)=0\}}\, dU^{\mathcal{M}}(s) = 0 \quad \text{almost surely,} \tag{104}$$

and hence for all $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}$,

$$\int_{T_{2m}}^{T_{2m+1}} 1_{\{W_{\mathcal{K}}(s)=0\}}\, dU^{\mathcal{M}}(s) = 0 \quad \text{almost surely on } \{T_{2m} < \infty\}. \tag{105}$$

For this, one uses the martingale property of the Brownian motion X and the fact that there is a $\beta^k \in \mathbb{R}_+^k$ and $\delta^{\mathcal{M},k} \in [1, \infty)$ such that $\langle \gamma_{\mathcal{K}}^{\mathcal{M}}, \beta^k \rangle = \delta^{\mathcal{M},k}$ for any $\emptyset \neq \mathcal{M} \subseteq \mathcal{K}$ (this follows from the fact that (58) holds with $\mathcal{K}(x) = \mathcal{K}$ where $n_i^{\mathcal{K}(x)} = 0$ if $i \notin \mathcal{K}(x)$ and $n_i^{\mathcal{K}(x)} = 1/\sqrt{|\mathcal{K}(x)|}$ if $i \in \mathcal{K}(x)$). Substituting (105) in (100) then yields the desired result. □

Lemma 7.10 *Suppose* (W, X, U) *is as in the hypothesis of Theorem* 7.7 *and* $\theta \in \mathbb{R}^N$. *Then* (63) *holds for all* $\emptyset \neq \mathcal{L} \subseteq \mathcal{K} \subseteq \mathcal{N}$ *where* $|\mathcal{K}| \geq 2$.

Proof Let $\mathcal{K} \subseteq \mathcal{N}$ satisfy $|\mathcal{K}| \geq 2$, $\emptyset \neq \mathcal{L} \subseteq \mathcal{K}$ and $\theta \in \mathbb{R}^N$. Without loss of generality (by considering a canonical representation on path space for example), we may assume that $(\Omega, \mathcal{F})$ is a standard measurable space and for $t \geq 0$, $\mathcal{F}_t \triangleq \sigma\{(W(s), X(s), U(s)) : 0 \leq s \leq t\}$. Let the associated probability measure be P^θ. Then X is a (θ, Γ)-Brownian motion on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P^\theta)$. By the Girsanov transformation (see Ikeda and Watanabe [12, p. 176]), there is a probability measure P^0 on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ such that under P^0, X is a $(0, \Gamma)$-Brownian motion starting from 0 and for each positive integer m, P^θ and P^0 are mutually absolutely continuous on $\mathcal{F}_m$. It follows that W with the probability measure P^0 on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ is an SRBM with extensive data $(\mathbb{R}^N_+, 0, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\})$ that starts from the origin. Then by Lemma 7.9, for each $\emptyset \neq \mathcal{L} \subseteq \mathcal{K} \subseteq \mathcal{N}$, (63) holds almost surely under P^0. But since P^θ and P^0 are mutually absolutely continuous on $\mathcal{F}_m$, it follows that (63) holds almost surely under P^θ with m in place of the upper limit ∞ there. Letting $m \to \infty$ yields the desired result. □

Proof of Theorem 7.7 Combining Lemmas 7.8, 7.9, and 7.10, we have proved the first part of the theorem.

To prove the second part of the theorem, we use (63). From the definition of an SRBM with extensive data (Definition 7.1) and the remark following it, W has the form:

$$W(t) = X(t) + \sum_{\emptyset \neq \mathcal{K} \subseteq \mathcal{N}} \gamma^{\mathcal{K}} U^{\mathcal{K}}(t), \quad t \geq 0, \tag{106}$$

where for $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$,

$$U^{\mathcal{K}}(t) = \int_0^t 1_{F_{\mathcal{K}}}\big(W(s)\big)\, dU^{\mathcal{K}}(s), \quad t \geq 0. \tag{107}$$

From (63), for $\mathcal{K} \subseteq \mathcal{N}$ with $|\mathcal{K}| \geq 2$,

$$U^{\mathcal{K}}(t) = \int_0^t 1_{F_{\mathcal{K}}}\big(W(s)\big)\, dU^{\mathcal{K}}(s) = 0 \quad \text{almost surely.} \tag{108}$$

Thus the only non-trivial terms in the sum in (106) are those indexed by $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$ where $|\mathcal{K}| = 1$. Equation (64) immediately follows. □

7.6 Proof of Theorem 6.1

Proof By Theorems 7.2 and 7.5, it suffices to prove that whenever (W, X, U) defines an SRBM with extensive data $(\overline{G}, \theta, \Gamma, \{\gamma^{\mathcal{K}}, \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\})$ that starts from the origin, then W is an SRBM with data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ that starts from the origin and the law of the latter is unique.

By Theorem 7.7, $W(\cdot)$ has the representation given by (64). For $i \in \mathcal{N}$, define

$$Y^i \triangleq c_i^{\emptyset} U^{\{i\}}. \tag{109}$$

Note that from (62), a.s.,

$$Y^i(t) = \int_0^t 1_{F_i}\big(W(s)\big)\, dY^i(s) \quad \text{for all } t \geq 0. \tag{110}$$

Therefore Y satisfies condition (iv) of Definition 6.1. From (64), (109), and the representation for $[\gamma^{\{1\}}, \gamma^{\{2\}}, \ldots, \gamma^{\{N\}}]$ given by (46), we have that for $t \geq 0$,

$$W(t) = X(t) + RY(t), \tag{111}$$

where by Lemma 5.3, R satisfies the HR condition, and W and X satisfy the other conditions of Definition 6.1 with $\nu = \delta_0$. Therefore, (W, X, Y) defines an SRBM with data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ that starts from the origin. Since R satisfies the HR condition, by Harrison and Reiman [10], the law of W is unique. It follows that

$$\hat{W}^r \Rightarrow W \quad \text{as } r \to \infty, \tag{112}$$

where W is an SRBM with data $(\mathbb{R}^N_+, \theta, \Gamma, R)$ that starts from the origin. □

Appendix A: Proof of Lemma 7.6

To prove Lemma 7.6, we use Proposition 4.4 of [14]. Specifically, we prove the following lemma, a restatement of condition (II) of Proposition 4.4 in [14], from which Lemma 7.6 follows. Our proof of Lemma A.1 is similar to the proof of Lemma 8.4 in Williams [25].

Lemma A.1 *For each $r > 0$, $\hat{X}^r = \check{X}^r + \varepsilon^r$, where ε^r is a process that converges to 0 in probability as $r \to \infty$, and*

(i) $\{\check{X}^r(t) - \check{X}(0) : r > 0\}$ *is uniformly integrable for each* $t \geq 0$,
(ii) *there is a sequence of constants* $\{\theta^r\}$ *in* $\mathbb{R}^N$ *such that* $\lim_{r\to\infty} \theta^r = \theta$,
(iii) *for each* r, $\{\check{X}^r(t) - \check{X}^r(0) - \theta^r t, t \geq 0\}$ *is a martingale with respect to the filtration generated by* $(\hat{W}^r, \check{X}^r, \hat{U}^r)$.

We need to develop some preliminaries before proving Lemma A.1.
For $r > 0$, $i \in \mathcal{N}$, and $n \in \mathbb{N}$, define

$$A^r_i(n) \triangleq \sum_{k=1}^{n} u^r_i(k), \tag{113}$$

where an empty sum is defined to be zero. Then for $r > 0$, the exogenous arrival process is defined for $i \in \mathcal{N}$, and $t \geq 0$, by

$$E^r_i(t) \triangleq \max\big\{n \geq 0 : A^r_i(n) \leq t\big\}. \tag{114}$$

Recall the definition of $V(\cdot)$ from (5).

For each $p \in \mathbb{N}^N$, let

$$\mathcal{G}_p^r \triangleq \sigma\left\{A^r\left(\cdot \wedge (p + e_{\mathcal{N}})\right), V^r(\cdot \wedge p)\right\}, \tag{115}$$

where

$$A^r\left(\cdot \wedge (p + e_{\mathcal{N}})\right) \triangleq \left(A_i^r\left(\cdot \wedge (p_i + 1)\right) : i \in \mathcal{N}\right) \tag{116}$$

and

$$V^r(\cdot \wedge p) \triangleq \left(V_i^r(\cdot \wedge p_i) : i \in \mathcal{N}\right). \tag{117}$$

Then $\{\mathcal{G}_p^r : p \in \mathbb{N}^N\}$ is multi-parameter filtration (see [8, Sect. 2.8]).

Definition A.1 A multi-parameter stopping time relative to $\{\mathcal{G}_p^r : p \in \mathbb{N}^N\}$ is a random variable τ taking values in $\mathbb{N}^N$ such that

$$\{\tau = p\} \in \mathcal{G}_p^r \tag{118}$$

for all $p \in \mathbb{N}^N$.

Lemma A.2 *For each $t \geq 0$,*

$$\tau^r(t) \triangleq E^r(t) \tag{119}$$

is a stopping time relative to $\{\mathcal{G}_p^r : p \in \mathbb{N}^N\}$.

Remark The reader will note that in defining $\mathcal{G}_p^r$, $e_{\mathcal{N}}$ is added to the argument of $A^r(\cdot)$. This has to be done because we need to know the first $p_i + 1$ interarrival times for the i-th user before we can determine whether $E_i^r(t) = p_i$ or not.

Proof For $i \in \mathcal{N}$ and $p \in \mathbb{N}^N$,

$$\left\{E_i^r(t) = p_i\right\} = \left\{A_i^r(p_i) \leq t < A_i^r(p_i + 1)\right\} \in \mathcal{G}_p^r. \tag{120}$$

Therefore, $\tau^r(t) = E^r(t)$ is a stopping time relative to $\{\mathcal{G}_p^r : p \in \mathbb{N}^N\}$. □

We next show that the diffusion-scaled workload process is adapted to the multi-parameter filtration stopped at the stopping time $\tau^r(r^2 t)$. The proof of the following lemma is based on the proof of Lemma 8.3 in [25] that proves the stopping time property of certain renewal processes for the system of interest. The following lemma, on the other hand, proves the adaptedness of the workload, a result that in [25], unlike here, follows from the structure of the system.

Lemma A.3 *The process $\hat{W}^r(\cdot)$ is adapted to the filtration $\{\mathcal{G}^r_{\tau^r(r^2 t)}, t \geq 0\}$, where $\tau^r(r^2 t) = E^r(r^2 t)$.*

Remark As a consequence of the adaptedness of $\hat{W}^r$, the processes $\{\hat{U}^{r,\mathcal{K}}(\cdot), \emptyset \neq \mathcal{K} \subseteq \mathcal{N}\}$ are adapted to the filtration $\{\mathcal{G}^r_{\tau^r(r^2 t)}, t \geq 0\}$ as well.

Proof From the definition of $\hat{W}^r(\cdot)$, it suffices to show that W^r is adapted to $\{\mathcal{G}^r_{\tau^r(t)}, t \geq 0\}$. Our proof is for fixed r and so the superscript r will be suppressed in the following proof.

Since $W(0) = 0$ (and for all $\emptyset \neq \mathcal{K} \subseteq \mathcal{N}$, $U^{\mathcal{K}}(0) = 0$), it follows that $W(0)$ and $U(0)$ are $\mathcal{G}_0$-measurable. Furthermore, the process $\{(A(p + e_{\mathcal{N}}), V(p)) : p \in \mathbb{N}^N\}$ is adapted to the multi-parameter filtration $\{\mathcal{G}_p : p \in \mathbb{N}^N\}$. Then by [8, Proposition 2.8.5] and the stopping time property of $\tau(t)$ (Lemma A.2), we have that for each $t \geq 0$:

$$\big(A\big(E(t) + e_{\mathcal{N}}\big), V\big(E(t)\big)\big) \in \mathcal{G}_{\tau(t)}. \tag{121}$$

Therefore, from (6), we only need to show that $T(t)$ (as defined by (32)) is adapted to the filtration $\{\mathcal{G}_{\tau(t)}, t \geq 0\}$.

Next, we define a strictly increasing sequence of real-valued random times $\{\eta_l\}_{l=0}^{\infty}$ for the (discrete event) queueing system such that $\eta_0 \triangleq 0$ and for $l = 1, 2, \ldots$, η_l is the l-th time that there is a new arrival for some queue or that a queue newly becomes empty. We have $\eta_l < \infty$ for each l, and $\eta_l \to \infty$ as $l \to \infty$. (This follows by the assumption concerning the exclusion of exceptional null sets made at the end of Sect. 3.2.)

For each $t \geq 0$, $p \in \mathbb{N}^N$,

$$\big\{E(t) = p\big\} = \bigcup_{j=0}^{\infty} \bigcap_{l=j}^{\infty} \big\{E(t \wedge \eta_l) = p\big\}. \tag{122}$$

For each $l \geq 0$, $p \in \mathbb{N}^N$, define

$$B^l_p \triangleq \big\{E(t \wedge \eta_l) = p\big\}. \tag{123}$$

Fix $t \geq 0$. It will be shown by induction that for each $l \geq 0$, the following two properties hold for all $p \in \mathbb{N}^N$:

(i) $B^l_p \in \mathcal{G}_p$,
(ii) for

$$\mathcal{I}^l \triangleq \big(t \wedge \eta_l, E(\cdot \wedge t \wedge \eta_l), T(\cdot \wedge t \wedge \eta_l), W(\cdot \wedge t \wedge \eta_l)\big), \tag{124}$$

we have $1_{B^l_p} \mathcal{I}^l \in \mathcal{G}_p$.

We now proceed with the induction proof. For $l = 0$, one has $\eta_0 = 0$ and $E(0) = 0$. Moreover, for all $p \in \mathbb{N}^N$, $W(0) = 0 \in \mathcal{G}_p$ and $T(0) = 0 \in \mathcal{G}_p$. Then, (i) and (ii) are easily verified to hold for $l = 0$.

For the induction step, assume that for some $l \geq 0$, (i) and (ii) hold for all $p \in \mathbb{N}^N$. Now,

$$B^{l+1}_p = \bigcup_m \big(B^{l+1}_p \cap B^l_m\big), \tag{125}$$

where the union is over all $m \in \mathbb{N}^N$ such that $m \leq p$. By the induction assumption, for fixed $p \in \mathbb{N}^N$ and any $m \in \mathbb{N}^N$ such that $m \leq p$, we have

$$B^l_m \in \mathcal{G}_m, \qquad 1_{B^l_m} \mathcal{I}^l \in \mathcal{G}_m. \tag{126}$$

Hence, from (124), $B^l_m \cap \{\eta_l \geq t\} \in \mathcal{G}_m$ and $B^l_m \cap \{\eta_l < t\} \in \mathcal{G}_m$.

Now, on $B_m^l \cap \{\eta_l \geq t\}$, $\eta_{l+1} \wedge t = \eta_l \wedge t$, $E(t \wedge \eta_{l+1}) = E(t \wedge \eta_l) = m$, and $\mathcal{I}^{l+1} = \mathcal{I}^l$. Thus, if $m = p$ we have

$$B_p^{l+1} \cap B_m^l \cap \{\eta_l \geq t\} = B_m^l \cap \{\eta_l \geq t\} \in \mathcal{G}_m, \tag{127}$$

or if $m \neq p$, then the left member of (127) is the empty set which is still in $\mathcal{G}_m$. Thus, combining the above with the induction assumption (126), we obtain

$$1_{B_p^{l+1} \cap B_m^l \cap \{\eta_l \geq t\}} \mathcal{I}^{l+1} = 1_{\{m=p\}} 1_{B_m^l \cap \{\eta_l \geq t\}} \mathcal{I}^l \in \mathcal{G}_m. \tag{128}$$

On the other hand, on $B_m^l \cap \{\eta_l < t\}$, $E(\eta_l) = E(t \wedge \eta_l) = m$ and the first time after η_l that a new external arrival occurs is $\eta = \min_{i \in \mathcal{N}} A_i(m_i + 1)$. Furthermore, on the set $B_m^l \cap \{\eta_l < t\}$, we have

$$\mathcal{I}^l = \big(\eta_l, E(\cdot \wedge \eta_l), T(\cdot \wedge \eta_l), W(\cdot \wedge \eta_l)\big). \tag{129}$$

Recall that the rate of service given to each of the users over the period $[\eta_l, \eta_{l+1})$ is given by $\sigma^l \triangleq \Lambda(W(\eta_l))$ where, from (7), $\Lambda(\cdot)$ is a measurable function on $\mathbb{R}_+^N$. It follows that if we define

$$\zeta \triangleq \eta_l + \inf\big\{s \geq 0 : W_i(\eta_l) - \sigma_i^l s = 0 \text{ for some } i \text{ such that } \sigma_i^l > 0, i \in \mathcal{N}\big\}, \tag{130}$$

then on $B_m^l \cap \{\eta_l < t\}$, $\eta_{l+1} = \eta \wedge \zeta$ where η_{l+1} is a measurable function of $(A(\cdot \wedge (m + e_{\mathcal{N}})), \eta_l, W(\eta_l))$, and hence by the induction assumption (126), (129), and the definition of $\mathcal{G}_m$, we have

$$1_{B_m^l \cap \{\eta_l < t\}} \eta_{l+1} \in \mathcal{G}_m. \tag{131}$$

Moreover, on $B_m^l \cap \{\eta_l < t\}$, we can express $E(\eta_{l+1})$, $T(\eta_{l+1})$, and $W(\eta_{1+1})$ as measurable functions of η_l, η_{l+1}, $E(\eta_l)$, $A(m + e_{\mathcal{N}})$, $T(\eta_l)$, $V(E(\eta_{l+1}))$, and $W(\eta_l)$ as follows. For $i \in \mathcal{N}$,

$$\begin{aligned} E_i(\eta_{l+1}) &= E_i(\eta_l) + 1_{\{A_i(m_i+1)=\eta_{l+1}\}}, \\ T_i(\eta_{l+1}) &= T_i(\eta_l) + \sigma_i^l(\eta_{l+1} - \eta_l), \\ W_i(\eta_{l+1}) &= V_i\big(E_i(\eta_{l+1})\big) - T_i(\eta_{l+1}). \end{aligned} \tag{132}$$

Since on $[\eta_l, \eta_{l+1})$, E is constant and T is linearly increasing at a fixed rate, given by σ^l, on combining the above with the induction assumption (126), (129), and (131), we have that

$$1_{B_m^l \cap \{\eta_l < t\}}\big(\eta_{l+1}, E(\cdot \wedge \eta_{l+1}), T(\cdot \wedge \eta_{l+1})\big) \in \mathcal{G}_m. \tag{133}$$

In particular,

$$1_{B_m^l \cap \{\eta_l < t\}}\big(E(t \wedge \eta_{l+1})\big) \in \mathcal{G}_m \tag{134}$$

and hence

$$B_p^{l+1} \cap B_m^l \cap \{\eta_l < t\} \in \mathcal{G}_m. \tag{135}$$

On combining this with (127), we see that $B_p^{l+1} \cap B_m^l \in \mathcal{G}_m \subset \mathcal{G}_p$ and hence by (125),

$$B_p^{l+1} \in \mathcal{G}_p. \tag{136}$$

Thus, (i) holds with $l+1$ in place of l. Similarly,

$$B_p^{l+1} \cap \{\eta_{l+1} \le t\} = \bigcup_m \left(B_p^{l+1} \cap B_m^l \cap \{\eta_l < t\} \cap \{\eta_{l+1} \le t\}\right) \in \mathcal{G}_p, \tag{137}$$

where the union is over all $m \in \mathbb{N}^N$ such that $m \le p$.

It remains to verify (ii) with $l+1$ in place of l. On $B_p^{l+1} \cap \{\eta_{l+1} \le t\}$, we have $V(E(\eta_{l+1})) = V(p) \in \mathcal{G}_p$. Combining this with (132), the fact that W is linearly decreasing on $[\eta_l, \eta_{l+1})$, and with (126), (133), (135), and (137), we have that

$$1_{B_p^{l+1} \cap B_m^l \cap \{\eta_l < t\} \cap \{\eta_{l+1} \le t\}} \mathcal{I}^{l+1} \in \mathcal{G}_p. \tag{138}$$

Combining the above with (125), (128) and the fact that

$$1_{B_p^{l+1} \cap B_m^l \cap \{\eta_l < t < \eta_{l+1}\}} \mathcal{I}^{l+1} = 1_{B_p^{l+1} \cap B_m^l \cap \{\eta_l < t < \eta_{l+1}\}} \mathcal{I}^l \in \mathcal{G}_p,$$

we conclude that

$$1_{B_p^{l+1}} \mathcal{I}^{l+1} \in \mathcal{G}_p. \qquad \square \tag{139}$$

Proof of Lemma A.1 An outline of our proof is as follows. The idea of the proof of part (iii) is that apart from small error terms associated with residual interarrival times, by suitably centering and scaling the primitive processes (A^r, V^r), we can re-express $\hat{X}^r$, as given by (34), in terms of a martingale evaluated at a stopping time. Indeed, we use the i.i.d. and independence assumptions on the primitive sequences $\{u_i^r(k), k = 1, 2, \dots, \}$, $\{v_i(k), k = 1, 2, \dots\}$, $i \in \mathcal{N}$, to establish the martingale property. In order to conclude that the stopped process is also a martingale, we establish L^2-bounds on the martingale and on the mean of the stopping time $\tau^r(t) = E^r(t)$. The martingale property in part (iii) of the lemma follows from this stopped martingale property and the fact that U^r and W^r are adapted to $\mathcal{G}^r_{\tau^r(t)}$. The asymptotic negligibility of error terms associated with the martingale property of the renewal process $E^r(t)$ is used to show that the residual process converges in probability to 0. Part (ii) of the lemma follows from the heavy traffic assumption (Assumption 4.1). Finally, the uniform integrability property in part (i) follows from L^2 bounds used in obtaining the stopped martingale property mentioned above. Now we provide the details of the proof.

For the moment, let r be fixed. Now,

$$\{\mathcal{G}_p^r\} \triangleq \{\mathcal{G}_p^r : p \in \mathbb{N}^N\} \tag{140}$$

defined by (115) is a (multi-parameter) filtration and for each $t \ge 0$, by Lemma A.2,

$$\tau^r(t) = E^r(t) \tag{141}$$

is a (multi-parameter) stopping time relative to this filtration. If $(\Omega^r, \mathcal{F}^r)$ is the measurable space on which all of the processes indexed by r are defined, then for each $t \geq 0$ we can define a σ-algebra associated with the multi-parameter stopping time $\tau^r(t)$ as follows:

$$\mathcal{G}^r_{\tau^r(t)} \triangleq \left\{B \in \mathcal{F}^r : B \cap \left\{\tau^r(t) \leq p\right\} \in \mathcal{G}^r_p \text{ for all } p \in \mathbb{N}^N\right\}. \tag{142}$$

Then $\{\mathcal{G}^r_{\tau^r(t)}, t \geq 0\}$ is a filtration in the usual single-parameter sense. From Lemma A.3, we have that the process W^r (and hence U^r) is adapted to this filtration.

We now introduce the fundamental multi-parameter martingales $\mathcal{M}^r$ and $\mathcal{O}^r$, and martingales associated with squares of their components. For each $p \in \mathbb{N}^N$ and $i \in \mathcal{N}$, let

$$\mathcal{M}^r_i(p_i) \triangleq \lambda^r_i A^r_i(p_i + 1) - (p_i + 1), \tag{143}$$

$$\mathcal{N}^r_i(p_i) \triangleq \left(\mathcal{M}^r_i(p_i)\right)^2 - (p_i + 1)\alpha^2_i, \tag{144}$$

$$\mathcal{O}^r_i(p_i) \triangleq V^r_i(p_i) - p_i m_i, \tag{145}$$

$$\mathcal{P}^r_i(p_i) \triangleq \left(\mathcal{O}^r_i(p_i)\right)^2 - p_i m^2_i \beta^2_i. \tag{146}$$

Let $\mathcal{M}^r(p) \triangleq (\mathcal{M}^r_i(p_i) : i \in \mathcal{N})$, $\mathcal{N}^r(p) \triangleq (\mathcal{N}^r_i(p_i) : i \in \mathcal{N})$, $\mathcal{O}^r(p) \triangleq (\mathcal{O}^r_i(p_i) : i \in \mathcal{N})$, $\mathcal{P}^r(p) \triangleq (\mathcal{P}^r_i(p_i) : i \in \mathcal{N})$. Because of the independence and i.i.d. assumptions of Sect. 3, we have that the $4N$-dimensional process:

$$\left\{\mathcal{Q}^r(p) \triangleq \left(\mathcal{M}^r(p), \mathcal{N}^r(p), \mathcal{O}^r(p), \mathcal{P}^r(p)\right) : p \in \mathbb{N}^N\right\}, \tag{147}$$

is a multi-parameter martingale relative to $\{\mathcal{G}^r_p\}$.

For each $p \in \mathbb{N}^N$, let

$$\mathcal{R}^r(p) \triangleq \left(\mathcal{M}^r(p), \mathcal{O}^r(p)\right). \tag{148}$$

We aim to show that $\{\mathcal{R}^r(\tau^r(t)), \mathcal{G}^r_{\tau^r(t)}, t \geq 0\}$ is a martingale. However, we cannot immediately deduce this from the martingale property of $\mathcal{Q}^r$, since $\tau^r(t)$ is a possibly unbounded stopping time. So we first truncate time, apply the multi-parameter stopping theorem, and then pass to the limit in the truncation using uniform integrability to deduce the desired result. The bounds obtained for the uniform integrability will also prove useful in verifying part (i) of the lemma. For $n \in \mathbb{N}$, let n^N denote the N-dimensional vector whose components all have value n. Then, we can verify (in a similar manner to that for $\mathcal{Q}^r$) that

$$\left\{\mathcal{Q}^{r,n}(p) \triangleq \mathcal{Q}^r\left(p \wedge n^N\right) : p \in \mathbb{N}^N\right\} \tag{149}$$

is a multi-parameter martingale relative to $\{\mathcal{G}^r_p\}$. Then by the multi-parameter optional stopping theorem (see [8, Theorem 2.8.7]) we have that

$$\left\{\mathcal{Q}^{r,n}\left(\tau^r(t)\right), \mathcal{G}^r_{\tau^r(t)}, t \geq 0\right\} \tag{150}$$

is a martingale for each $n \in \mathbb{N}$. Now, for $p \in \mathbb{N}^N$ and $n \in \mathbb{N}$, let

$$\mathcal{R}^{r,n}(p) \triangleq \big(\mathcal{M}^r\big(p \wedge n^N\big), \mathcal{O}^r\big(p \wedge n^N\big)\big). \tag{151}$$

For each $n \in \mathbb{N}$, it follows from the martingale property of $\{\mathcal{Q}^{r,n}(\tau^r(t)), \mathcal{G}^r_{\tau^r(t)}, t \geq 0\}$ that

$$\big\{\mathcal{R}^{r,n}\big(\tau^r(t)\big), \mathcal{G}^r_{\tau^r(t)}, t \geq 0\big\} \tag{152}$$

is a martingale. We aim to prove that the same is true with $\mathcal{R}^r$ in place of $\mathcal{R}^{r,n}$. For $t \geq 0$ fixed, $\mathcal{R}^{r,n}(\tau^r(t)) \to \mathcal{R}^r(\tau^r(t))$ pointwise as $n \to \infty$, and so it suffices to show that $\{\mathcal{R}^{r,n}(\tau^r(t))\}_{n=1}^{\infty}$ is L^2-bounded for each $t \geq 0$, since this implies that it is uniformly integrable. By the martingale properties of the $\mathcal{N}^r$ and $\mathcal{P}^r$ elements of $\mathcal{Q}^{r,n}(\tau^r(\cdot))$ we have for all $i \in \mathcal{N}$, $n \geq 1$:

$$\mathbf{E}\big[\big(\mathcal{M}^r_i\big(E^r_i(t) \wedge n\big)\big)^2 - \big(\big(E^r_i(t) + 1\big) \wedge n\big)\alpha_i^2\big] = 0, \tag{153}$$

$$\mathbf{E}\big[\big(\mathcal{O}^r_i\big(E^r_i(t) \wedge n\big)\big)^2 - \big(E^r_i(t) \wedge n\big)m_i^2\beta_i^2\big] = 0. \tag{154}$$

From Lorden's inequality for renewal processes (see Lindvall [16, pp. 77–78]; Carlsson and Nerman [5]), we obtain the following upper bound for $i \in \mathcal{N}$:

$$\mathbf{E}\big[E^r_i(t) + 1\big] \leq \lambda^r_i t + \alpha_i^2 + 2 \triangleq h^r_i(t), \tag{155}$$

where $h^r_i(t)$ is finite. It then follows from (153)–(155) that for all $n \geq 1$, $i \in \mathcal{N}$,

$$\mathbf{E}\big[\big(\mathcal{M}^r_i\big(E^r_i(t) \wedge n\big)\big)^2\big] \leq \alpha_i^2 h^r_i(t), \tag{156}$$

$$\mathbf{E}\big[\big(\mathcal{O}^r_i\big(E^r_i(t) \wedge n\big)\big)^2\big] \leq m_i^2\beta_i^2 h^r_i(t). \tag{157}$$

This establishes the desired L^2-boundedness and hence

$$\big\{\mathcal{R}^r\big(\tau^r(t)\big), \mathcal{G}^r_{\tau^r(t)}, t \geq 0\big\} \tag{158}$$

is a martingale for each r.

We now apply the above martingale properties to establish part (iii) of the lemma. For $i \in \mathcal{N}$, define

$$\check{X}^r_i(t) \triangleq r^{-1}\big(\mathcal{O}^r_i\big(E^r_i(r^2t)\big) - m_i\mathcal{M}^r_i\big(E^r_i(r^2t)\big) + (\lambda^r_i - \lambda_i)m_i r^2 t\big), \tag{159}$$

$$\varepsilon^r_i(t) \triangleq r^{-1}m_i\big(\lambda^r_i A^r_i\big(E^r_i\big(r^2t\big) + 1\big) - \big(\lambda^r_i r^2 t + 1\big)\big), \tag{160}$$

$$\theta^r_i \triangleq r\big(\lambda^r_i - \lambda_i\big)m_i. \tag{161}$$

Then from (14), (19), (20), and (34), for $i \in \mathcal{N}$, $t \geq 0$,

$$\hat{X}^r_i(t) = \check{X}^r_i(t) + \varepsilon^r_i(t). \tag{162}$$

Since

$$\mathcal{R}^r\big(\tau^r\big(r^2t\big)\big) = \big(\mathcal{M}^r\big(E^r\big(r^2t\big)\big), \mathcal{O}^r\big(E^r\big(r^2t\big)\big)\big), \tag{163}$$

it follows, from the martingale property of (158), that

$$\left\{\check{X}^r(t) - \check{X}^r(0) - \theta^r t, \mathcal{G}^r_{\tau^r(r^2 t)}, t \geq 0\right\} \tag{164}$$

is a martingale. Note that by Lemma A.3, $\hat{U}^r$ and $\hat{W}^r$ are adapted to the filtration $\{\mathcal{G}^r_{\tau^r(r^2 t)}, t \geq 0\}$. Hence, $\{\check{X}^r(t) - \check{X}^r(0) - \theta^r t, t \geq 0\}$ is a martingale relative to the filtration generated by $(\hat{W}^r, \check{X}^r, \hat{U}^r)$.

We next show that ε^r converges in probability to the zero process as $r \to \infty$. By the definition of E^r_i from A^r_i for $i \in \mathcal{N}$, for each $T \geq 0$,

$$\left\|\varepsilon^r(\cdot)\right\|_T \leq 2 \max_{i \in \mathcal{N}} \left|m_i \lambda^r_i\right| \left\|r^{-1} u^r_i \left(E^r_i\left(r^2 \cdot\right) + 1\right)\right\|_T + \max_{i \in \mathcal{N}} |m_i| / r, \tag{165}$$

where, as a consequence of the functional central limit theorem (Proposition 5.1), the right-hand side above goes to zero in probability as $r \to \infty$ (see the proof of Lemma 6 in Iglehart and Whitt [11]).

Part (ii) of the lemma follows from the heavy traffic assumption (Assumption 4.1).

It remains to show part (i) of the lemma. For this it suffices to show that $\check{X}^r(t)$ as r varies is uniformly integrable for each fixed $t \geq 0$. Now by Fatou's lemma, (156)–(157) hold with the n's removed. Fix $t \geq 0$. By (155), we have

$$\sup_r \max_{i \in \mathcal{N}} \frac{h^r_i(r^2 t)}{r^2} < \infty. \tag{166}$$

Replacing t by $r^2 t$ in (156)–(157), and combining with the above, we see that

$$\left\{r^{-1}\left(\mathcal{M}^r\left(E^r\left(r^2 t\right)\right), \mathcal{O}^r\left(E^r\left(r^2 t\right)\right)\right)\right\} \tag{167}$$

as a collection indexed by r is L^2-bounded, and hence uniformly integrable. The uniform integrability of $\{\check{X}^r(t)\}$ follows. □

References

1. Bell, S.L., Williams, R.J.: Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: Asymptotic optimality of a threshold policy. Ann. Appl. Probab. **11**(3), 608–649 (2001)
2. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York (1979)
3. Bhardwaj, S., Williams, R.J., Acampora, A.S.: On the performance of a two-user MIMO downlink system in heavy traffic. IEEE Trans. Inf. Theory **53**(5), 1851–1859 (2007)
4. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, New York (1999)
5. Carlsson, H., Nerman, O.: An alternative proof of Lorden's renewal inequality. Adv. Appl. Probab. **18**(4), 1015–1016 (1986)
6. Choi, W., Andrews, J.G.: The capacity gain from intercell scheduling in multi-antenna systems. IEEE Trans. Wirel. Commun. **7**(2), 714–725 (2008)
7. Costa, M.H.M.: Writing on dirty paper. IEEE Trans. Inf. Theory **IT-29**(3), 439–441 (1983)
8. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1986)
9. Georgiadis, L., Neely, M.J., Tassiulas, L.: Resource allocation and cross-layer control in wireless networks. Found. Trends Netw. **1**(1), 1–144 (2006)

10. Harrison, J.M., Reiman, M.I.: Reflected Brownian motion on an orthant. Ann. Probab. **9**(2), 302–308 (1981)
11. Iglehart, D.L., Whitt, W.: The equivalence of functional central limit theorems for counting processes and associated partial sums. Ann. Math. Stat. **42**(4), 1372–1378 (1971)
12. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, 2nd edn. North-Holland Mathematical Library. Birkhäuser, Amsterdam (1989)
13. Jindal, N., Vishwanath, S., Goldsmith, A.: On the duality of Gaussian multiple-access and broadcast channels. IEEE Trans. Inf. Theory **50**(05), 768–783 (2004)
14. Kang, W., Williams, R.J.: An invariance principle for semimartingale reflecting Brownian motions in domains with piecewise smooth boundaries. Ann. Appl. Probab. **17**(2), 741–779 (2007)
15. Kelly, F.P., Laws, C.N.: Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling. Queueing Syst. Theory Appl. **13**(1–3), 47–86 (1993)
16. Lindvall, T.: Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1992)
17. Ng, B.L., Evans, J.S., Hanly, S.V., Aktas, D.: Transmit beamforming with cooperating base stations. In: Proc. of IEEE International Symposium on Information Theory, Adelaide, Australia, September 4–9, 2005, pp. 1431–1435
18. Peterson, W.P.: A heavy traffic limit theorem for networks of queues with multiple customer types. Math. Oper. Res. **16**(1), 90–118 (1991)
19. Prokhorov, Y.V.: Convergence of random processes and limit theorems in probability theory. Theory Probab. Appl. **1**(2), 157–214 (1956)
20. Reiman, M.I., Williams, R.J.: A boundary property of semimartingale reflecting Brownian motions. Probab. Theory Relat. Fields **77**(1), 87–97 (1988)
21. Shakkotai, S., Srikant, R., Stolyar, A.L.: Pathwise optimality of the exponential scheduling rule for wireless channels. Adv. Appl. Probab. **36**(4), 1021–1045 (2004)
22. Shamai (Shitz), S., Zaidel, B.M.: Enhancing the cellular downlink capacity via co-processing at the transmitting end. In: Proc. of Spring IEEE Vehicular Technology Conf., Rhodes, Greece, May 6–9, 2001, vol. 3, pp. 1745–1749
23. Stolyar, A.L.: Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. Ann. Appl. Probab. **14**(1), 1–53 (2004)
24. Weingarten, H., Steinberg, Y., Shamai (Shitz), S.: The capacity region of the Gaussian multiple-input multiple-output broadcast channel. IEEE Trans. Inf. Theory **52**(09), 3936–3964 (2006)
25. Williams, R.J.: Diffusion approximations for open multiclass queueing networks: Sufficient conditions involving state space collapse. Queueing Syst. Theory Appl. **30**(1–2), 27–88 (1998)