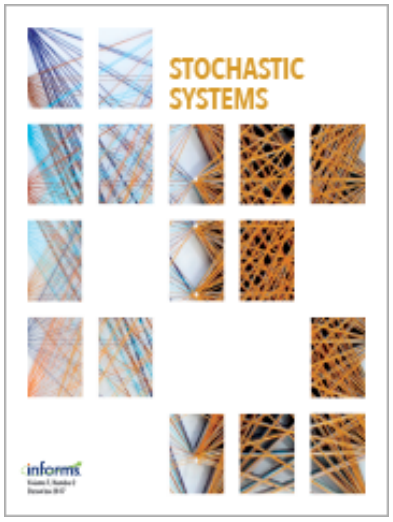
This article was downloaded by: [216.73.217.172] On: 15 November 2026, At: 06:34
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Diffusion Approximation for an Input-queued Switch Operating under a Maximum Weight Matching Policy

W. N. Kang, R. J. Williams

To cite this article:
W. N. Kang, R. J. Williams (2012) Diffusion Approximation for an Input-queued Switch Operating under a Maximum Weight Matching Policy. Stochastic Systems 2(2):277-321. https://doi.org/10.1287/12-SSY061

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*Stochastic Systems*. Copyright 2012 The Author(s). https://doi.org/10.1287/12-SSY061, used under a Creative Commons Attribution License: http://creativecommons.org/licenses/by/4.0/."

Copyright © 2012, The author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Stochastic Systems
2012, Vol. 2, No. 2, 277–321
DOI: 10.1214/12-SSY061

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

DIFFUSION APPROXIMATION FOR AN INPUT-QUEUED SWITCH OPERATING UNDER A MAXIMUM WEIGHT MATCHING POLICY

By W. N. Kang and R. J. Williams*

University of Maryland, Baltimore County and
University of California, San Diego

For $N \geq 2$, we consider an $N \times N$ input-queued switch operating under a maximum weight matching policy. We establish a diffusion approximation for a $(2N-1)$-dimensional workload process associated with this switch when all input ports and output ports are heavily loaded. The diffusion process is a semimartingale reflecting Brownian motion living in a polyhedral cone with N^2 boundary faces, each of which has an associated constant direction of reflection. Our proof builds on our own prior work [13] on an invariance principle for semimartingale reflecting Brownian motions in piecewise smooth domains and on a multiplicative state space collapse result for switched networks established by Shah and Wischik in [19].

1. Introduction. Input-queued crossbar switches are widely used in Internet routers. The main control feature in such switches is the choice of a scheduling policy for the transfer of packets from input ports to output ports in each time slot. A policy that has received considerable attention is the so-called maximum weight matching (MWM) policy. At each time step, this policy chooses a matching (or bijection) from the input ports to the output ports that maximizes the sum of the lengths of the virtual output queues served by the matching. Under various assumptions on the arrival processes, rate stability has been established for the MWM policy, provided the load placed on each of the input and output ports is less than its maximum capacity [7, 11, 21].

Recently, Shah and Wischik [19] have studied the asymptotic behavior of fluid models for heavily loaded switched networks operating under maximum-weight-like policies and used this behavior to prove a multiplicative state space collapse result for such networks. In this paper, assuming

Received January 2012.

*Research of R. J. Williams was supported in part by NSF grants DMS-0906535 and DMS-1206772.

AMS 2000 subject classifications: Primary 60J60, 60J70, 60K30, 90B36.

Keywords and phrases: Input-queued switch, maximum weight matching policy, diffusion approximation, heavy traffic, semimartingale reflecting Brownian motion (SRBM).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

that multiplicative state space collapse holds, we establish a diffusion approximation for an $N \times N$ (where $N \geq 2$) input-queued switch operating under the MWM policy when all of the input ports and output ports are heavily loaded. (This contrasts with the situation considered in Stolyar [20] for a generalized switch model, where in effect only one input port or one output port is heavily loaded.) The diffusion is a semimartingale reflecting Brownian motion living in a polyhedral cone in $\mathbb{R}_+^{2N-1}$ with N^2 boundary faces, each of which has an associated constant direction of reflection. Our proof builds on our own prior work [13] on an invariance principle for semimartingale reflecting Brownian motions in piecewise smooth domains. When combined with the multiplicative state space collapse result of Shah and Wischik [19], specialized to single-hop networks operating under the MWM policy, this yields a new diffusion approximation result for a heavily loaded $N \times N$ input-queued switch operating under the MWM policy.

Our interest in studying this problem stems not only from the application, but also from the fact that an input-queued switch can be viewed as an example of a stochastic processing network with head-of-the-line (HL) service [10] in which each activity can simultaneously process packets from more than one queue. This is similar to an assembly or joining operation familiar in manufacturing systems. Another type of stochastic processing network that involves simultaneous actions arises from bandwidth sharing models considered recently by several authors, see e.g., [12, 17, 23]. However, in these models, the activities simultaneously use multiple resources or servers to process jobs; in other words, the simultaneous aspect is at the opposite end of the activity than in the case of an input-queued switch. Furthermore, in both HL multiclass queueing networks and in bandwidth sharing models, the workload for a single server is typically one-dimensional, whereas for a single $N \times N$ input-queued switch, which has only one server, the workload dimension is $2N-1$ when all input and output ports are heavily loaded. The analysis of stochastic processing networks is still in its early stages of development and the study of this concrete switch example, with its simultaneous processing action, provides valuable information that can guide development of a general theory.

There are several possible directions for further research related to diffusion approximations for switches. Here we have focused on the situation where all input and output ports are heavily loaded. At the other end of the spectrum, Stolyar [20] has established a diffusion approximation for the situation when only one input or one output port is heavily loaded. It would be interesting to develop results where only some input and output ports are heavily loaded. Such an approximation does not follow from our results

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

as we require a non-degenerate covariance matrix for the uniqueness of our diffusion. Indeed, in light of the complexities of the geometry of the state space and directions of reflection for our diffusion approximation, it is likely that some care is required to interpolate between the single heavily loaded port case of Stolyar and the all ports heavily loaded case we treat here. In addition to this interesting direction, one might also consider investigating other scheduling policies. In this paper, we have restricted attention to the maximum weight matching (MWM) policy. However, Shah and Wischik [19] have established multiplicative state space collapse for some generalizations of this policy. In particular, one class of such policies are obtained when the virtual output queuelength is replaced by its α^{th} power in the criterion for determining an optimal matching, yielding the MW-α policies for $\alpha \in (0, \infty)$. (The MWM policy considered in this paper corresponds to $\alpha = 1$.) The multiplicative state space collapse result established in [19] under such a policy leads to a natural conjecture for a diffusion approximation to the workload process; this diffusion lives in a cone with piecewise smooth curved boundaries when $\alpha \neq 1$. While we believe that the invariance principle in [13] can be adapted to justify this approximation for $\alpha \in (1, \infty)$ for small N, the complexities of the geometry make this a daunting task as N increases. For $\alpha \in (0, 1)$, in addition to the geometric complexity, there is not yet even an existence and uniqueness theory (nor an associated invariance principle) for the proposed diffusion process. Shah and Wischik [19] also considered multi-hop switched networks, and it would be interesting to see if a diffusion approximation could be established for that more general network setting. The interested reader may also wish to consult the recent work of Shah et al. [15, 16, 18] who, for several policies, have investigated bounds on the total system workload for input-queued switches as a function of both N and the distance from the heavy-traffic boundary.

This paper is organized as follows. In Section 2, we introduce our model for an input-queued switch operating under a maximum weight matching policy. We also introduce the key performance processes there, namely, an N^2-dimensional queue-length process with one component for each input-output pair and a $(2N - 1)$-dimensional workload process which tracks the work (measured in packets) waiting in each of the first $(N - 1)$ input ports, waiting for the first $(N - 1)$-output ports and the total work waiting in the system. In Section 3, we introduce the setup for our heavy traffic diffusion approximation result. In particular, we consider an $N \times N$ input-queued switch with a sequence of arrival processes, where the load on all input and output ports approaches full capacity as one moves along the sequence. The arrival processes are assumed to satisfy functional central limit theorems.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

In Section 4, we introduce a critical fluid model associated with the switch model. We recall the characterization of its invariant states and a sufficient condition for multiplicative state space collapse given in [19]. The definition of our diffusion process (which is a semimartingale reflecting Brownian motion living in a polyhedral cone) is given in Section 5, where the main result of this paper is also stated. The proof of this result is given in Section 6. When coupled with the multiplicative state space collapse result of Shah and Wischik [19], our main result yields a diffusion approximation for the workload process (and hence for the queue-length process) for a heavily loaded input-queued switch. In Section 7, with a view to future research, we discuss the main difficulty in proving diffusion approximations for an input-queued switch operating under the MW-α policies for $\alpha \neq 1$.

1.1. *Notation and terminology.* The following notation will be used throughout the paper. The symbol $\mathbb{N}$ will denote the set of strictly positive integers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ will denote the set of non-negative integers. The symbol $\mathbb{R}$ will denote the set of real numbers and $\mathbb{R}_+ = [0, \infty)$ is the set of non-negative real numbers. For $d \in \mathbb{N}$, $\mathbb{R}^d$ denotes d-dimensional Euclidean space, $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : \ x_i \geq 0, i = 1, \ldots, d\}$ and $\mathbb{N}^d = \{x \in \mathbb{R}^d : x_i \in \mathbb{N}, i = 1, \ldots, d\}$. For $a, b \in \mathbb{R}$, $a \vee b$ denotes the maximum of a and b and $a \wedge b$ denotes the minimum of a and b. The indicator function of the set B is denoted by 1_B (that is, $1_B(x) = 1$ if $x \in B$ and $1_B(x) = 0$ otherwise). All vectors and matrices in this paper are assumed to have real-valued entries. We denote the inner product on $\mathbb{R}^d$ by $\langle \cdot, \cdot \rangle$, that is, $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$, for $x, y \in \mathbb{R}^d$. The usual Euclidean norm on $\mathbb{R}^d$ will be denoted by $|\cdot|$ so that $|x| = (\sum_{i=1}^d x_i^2)^{1/2}$ for $x \in \mathbb{R}^d$. For $x, y \in \mathbb{R}^d$, $x \geq y$ means $x_i \geq y_i$ for all $i = 1, \ldots, d$. We will let A' denote the transpose of A, where A is a matrix or a vector. Let $|A|$ denote the norm of an $m \times n$ matrix A (for $m, n \in \mathbb{N}$), so that $|A| = (\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2)^{1/2}$. Let $\mathcal{N}(A)$ denote the null space of such a matrix A, that is, $\mathcal{N}(A) = \{x \in \mathbb{R}^n : \ Ax = 0\}$. For a function f from $[0, \infty)$ into $\mathbb{R}^d$ and any $T \geq 0$, let $||f||_T = \sup_{0 \leq t \leq T} |f(t)|$.

We denote by $\mathbb{D}([0, \infty), \mathbb{R}^d)$ (for some $d \in \mathbb{N}$) the space of right continuous functions with finite left limits (r.c.l.l. functions) from $\mathbb{R}_+$ into $\mathbb{R}^d$ and we endow this space with the usual Skorokhod J_1-topology (see Section 5 of Chapter 3 of Ethier and Kurtz [9]) under which it is a Polish space. We denote by $\mathbb{C}([0, \infty), \mathbb{R}^d)$ the space of continuous functions from $[0, \infty)$ into $\mathbb{R}^d$. The Borel σ-algebra on $\mathbb{D}([0, \infty), \mathbb{R}^d)$ or $\mathbb{C}([0, \infty), \mathbb{R}^d)$ will be denoted by $\mathcal{M}^d$. All continuous time stochastic processes used in this paper will be assumed to have r.c.l.l. paths in some Euclidean space. Consider $W, W_1, W_2, \ldots$, each of which is a d-dimensional process (possibly defined on different probabil-

ity spaces). The sequence $\{W_n\}_{n=1}^{\infty}$ is said to be C-tight if the probability measures induced by the sequence $\{W_n\}_{n=1}^{\infty}$ on $(\mathbb{D}([0,\infty),\mathbb{R}^d),\mathcal{M}^d)$ form a tight sequence and if each limit point, obtained as a limit in distribution along a subsequence, almost surely has sample paths in $\mathbb{C}([0,\infty),\mathbb{R}^d)$. The notation "$W_n \Rightarrow W$" will mean that "$W_n$ converges in distribution to W as $n \to \infty$".

Throughout this paper, we shall use Propositions to state previously known or standard results. Lemmas will state supporting results proved here and the Theorem will indicate our main result.

2. Input-queued switch model.

2.1. *Model description.* Fix an integer $N \geq 2$. The structure for an $N \times N$ input-queued switch is illustrated in Figure 1.

An $N \times N$ input-queued switch has N input ports and N outport ports. Time is slotted so that packets of fixed size arrive at the switch at the beginning of a time slot. For concreteness, time slot n denotes the time interval $[n-1,n)$, $n \in \mathbb{N}$, and we assume zero packets arrive at time zero. Packets arriving at input port i and destined for output port j are stored in a first-in-first-out (FIFO) buffer called a "Virtual Output Queue" (VOQ), denoted here by VOQ_{ij}. The use of VOQs avoids the so-called head-of-line blocking phenomenon (cf. [1] and [11]). Thus, there are N^2 separate VOQs, one for each input–output pair. The packets arriving at an input port are switched from that port to the appropriate output port by a crossbar fabric. At the beginning of each time slot (just after packets have arrived for the time slot), a matching policy specifies which input ports are to be connected to which output ports during the time slot. At the end of that time slot, input port i transfers one packet to output port j if they are matched to one another and VOQ_{ij} is non-empty. In each time slot, each input port can

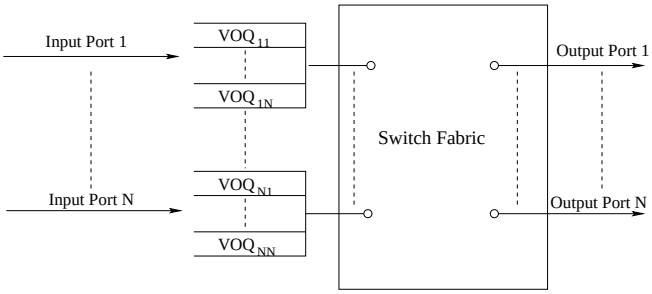


FIG 1. *An $N \times N$ input-queued switch*

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

transmit at most one packet and each output port can receive at most one packet. Due to the constraints just mentioned, this scheduling amounts to choosing a bipartite matching (or bijection) between the sets of input ports and output ports at the beginning of each time slot. Such a matching may connect input port i to output port j even if VOQ_{ij} is empty, in which case no packet is transmitted and the system is said to "transmit a blank".

2.2. *Stochastic primitives.* For each $i, j \in \mathcal{I} \doteq \{1, 2, \ldots, N\}$ and integer $n \in \mathbb{N}$, let $I_{ij}(n)$ denote the number of packets that have arrived to input port i that are destined for output port j in the time interval $(0, n]$, and set $I_{ij}(0) = 0$. Note that these packets are stored in VOQ_{ij}. We assume that, for $i, j \in \mathcal{I}$, $I_{ij}(\cdot)$ is defined from a sequence of i.i.d., nonnegative, integer valued random variables $\{\vartheta_{ij}(k) : k \in \mathbb{N}\}$ with mean $\lambda_{ij} \in (0, \infty)$ and variance $b_{ij} \in (0, \infty)$, where for each $k \in \mathbb{N}$, $\vartheta_{ij}(k)$ denotes the number of packets that arrive to input port i and that are destined for output port j at time k. Then $I_{ij}(n)$ has the representation:

$$I_{ij}(n) = \sum_{k=1}^{n} \vartheta_{ij}(k), \qquad n \in \mathbb{N}_0, \tag{1}$$

where an empty sum is defined to be zero. We assume that the sequences $\{\vartheta_{ij}(k) : k \in \mathbb{N}\}$ for $i, j \in \mathcal{I}$ are mutually independent, and so the processes $I_{ij}(\cdot)$ for $i, j \in \mathcal{I}$ are mutually independent. For each $n \in \mathbb{N}_0$, let $I(n)$ denote the N^2-dimensional vector determined by $\{I_{ij}(n) : i, j \in \mathcal{I}\}$ such that $I_{ij}(n)$ is the $((i-1)N + j)^{th}$ entry of $I(n)$. This notation allows us to refer to the N^2-dimensional vector $I(n)$ rather than an $N \times N$ matrix.

2.3. *Scheduling policy: Maximum weight matching.* The $N \times N$ input-queued switch is assumed to be operated under a maximum weight matching policy, denoted by MWM policy. At the beginning of each time slot (just after the arrival of new packets), a matching is chosen by the MWM policy to connect input ports and output ports during that time slot. The matching is chosen as a function of the number of packets in each of the VOQs at the beginning of the time slot. For $i, j \in \mathcal{I}$ and $n \in \mathbb{N}_0$, let $Q_{ij}(n)$ denote the number of packets in VOQ_{ij} at time n (immediately after the arrival of any packets at time n), and let $Q(n)$ denote the N^2-dimensional vector whose $((i-1)N + j)^{th}$ entry is $Q_{ij}(n)$. The process $Q_{ij}(\cdot)$ satisfies the equation:

$$Q_{ij}(n) = Q_{ij}(0) + I_{ij}(n) - D_{ij}(n), \qquad n \in \mathbb{N}_0, \tag{2}$$

where $D_{ij}(n)$ denotes the cumulative number of packets that have departed from VOQ_{ij} by time n. For concreteness, we imagine the packets departing

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

at the end of each time slot, just before the beginning of the next time slot, e.g., $D_{ij}(1)$ is the number of pockets that departed during the time slot $[0,1)$. We next describe the MWM policy more precisely, which, in turn, specifies the form of $D_{ij}(\cdot)$.

We can represent a matching π by an $N \times N$ matrix of zeros and ones $\pi = [\pi_{ij}]$, where $\pi_{ij} = 1$ if input port i is connected to output port j, otherwise $\pi_{ij} = 0$. The matrix π has exactly one 1 in each row and column. It can be viewed as representing a permutation of N elements and there are a total of $N!$ distinct matchings. We denote the set of distinct matchings by Π. At time n for $n \in \mathbb{N}_0$, the MWM policy chooses a matching π in the following way: immediately after the arrivals of packets at time n, for each matching π, the weight of π over the time interval $[n, n+1)$, denoted by $\omega_\pi(Q(n))$, is computed via

$$\omega_\pi(Q(n)) = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij} Q_{ij}(n). \tag{3}$$

Then, the MWM policy chooses the matching

$$\pi^*(Q(n)) \equiv \operatorname{argmax}_{\pi \in \Pi} \{\omega_\pi(Q(n))\}, \tag{4}$$

for use in the time interval $[n, n+1)$. (If there is more than one matching with maximal weight, a deterministic ordering of matchings is assumed and the earliest matching in the ordered list is used.) At the end of the interval $[n, n+1)$, packets are transferred according to the matching $\pi^*(Q(n))$.

For each $i, j \in \mathcal{I}$ and $n \in \mathbb{N}_0$, we can now give a description of $D_{ij}(n)$. Let $D_{ij}(0) = 0$. For each matching π, let $T_\pi(0) = 0$ and for $n \in \mathbb{N}$, let $T_\pi(n)$ denote the total number of time slots in $[0, n)$ that the matching π has been used by the MWM policy. Then,

$$T_\pi(\cdot) \text{ is non-decreasing and } \sum_{\pi \in \Pi} T_\pi(n) = n, \tag{5}$$

$$T_\pi(n) - T_\pi(n-1) = 0 \text{ if } \exists\ \sigma \in \Pi \text{ such that } \omega_\sigma(Q(n-1)) > \omega_\pi(Q(n-1)), \tag{6}$$

$$D_{ij}(n) = \sum_{\pi \in \Pi} \sum_{l=1}^{n} \pi_{ij} 1_{\{Q_{ij}(l-1)>0\}} (T_\pi(l) - T_\pi(l-1)). \tag{7}$$

For each $n \in \mathbb{N}_0$, let $D(n)$ denote the N^2-dimensional vector determined by $\{D_{ij}(n) : i, j \in \mathcal{I}\}$ such that $D_{ij}(n)$ is the $((i-1)N + j)^{th}$ entry of $D(n)$. Combining (2) and (7) we see that

$$Q_{ij}(n) = Q_{ij}(0) + I_{ij}(n) - \sum_{\pi \in \Pi} \pi_{ij} T_\pi(n) + U_{ij}(n), \tag{8}$$

where

$$U_{ij}(n) = \sum_{\pi \in \Pi} \sum_{l=1}^{n} \pi_{ij} 1_{\{Q_{ij}(l-1)=0\}} (T_\pi(l) - T_\pi(l-1)) \tag{9}$$

is the cumulative number of "blanks" transmitted from input port i to output port j up to time n.

2.4. *Workload process.* In this subsection we define a $(2N-1)$-dimensional workload process W from the queue-length process Q.

For $n \in \mathbb{N}_0$ and $i, j \in \mathcal{I} \setminus \{N\}$, define

$$\begin{aligned} W_i(n) &= \sum_{l \in \mathcal{I}} Q_{il}(n), \\ W_{N-1+j}(n) &= \sum_{k \in \mathcal{I}} Q_{kj}(n), \\ W_{2N-1}(n) &= \sum_{k,l \in \mathcal{I}} Q_{kl}(n). \end{aligned}$$

For $i, j \in \mathcal{I} \setminus \{N\}$ and $n \in \mathbb{N}_0$, the quantity $W_i(n)$ is the amount of work (measured in packets) that is waiting at input port i at time n, $W_{N-1+j}(n)$ is the amount of work destined for output port j that is waiting to be transmitted at time n, and $W_{2N-1}(n)$ is the total amount of work in the system at time n. We have used this non-symmetric form of workload process, rather than the symmetric $2N$-dimensional process that has one workload component for each input port and each output port, because the former has no redundant components whereas the latter has one redundant component. Note in particular that the workload at input port N is given by $W_{2N-1}(n) - \sum_{i=1}^{N-1} W_i(n)$. Let A be the $(2N-1) \times N^2$ matrix such that $(\zeta_i', \zeta_j', 1)'$ is the $((i-1)N + j)^{th}$ column of A for $i, j \in \mathcal{I}$, where $\zeta_i \in \mathbb{R}^{N-1}$ is the i^{th} unit coordinate vector in $\mathbb{R}^{N-1}$ for $1 \le i \le N-1$ and ζ_N is the zero vector in $\mathbb{R}^{N-1}$. For example, when $N = 2$, the matrix A is

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{10}$$

Then the workload process W can also be written in the following compact form:

$$W(n) = AQ(n) \text{ for } n \in \mathbb{N}_0. \tag{11}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Let A^{ij} denote the $((i-1)N+j)^{th}$ column of A for $i,j \in \mathcal{I}$ and A^{ij}_k denote the k^{th} element of this column for $k = 1,\dots,2N-1$. Then

$$A^{ij}_k = \begin{cases} 1_{\{k=i\}} & \text{for } 1 \le k \le N-1, \\ 1_{\{k-N+1=j\}} & \text{for } N \le k \le 2N-2, \\ 1 & \text{for } k = 2N-1, \end{cases} \tag{12}$$

and for any $\pi \in \Pi$,

$$\sum_{i,j\in\mathcal{I}} A^{ij}_k \pi_{ij} = \begin{cases} \sum_{j\in\mathcal{I}} \pi_{kj} = 1 & \text{for } 1 \le k \le N-1, \\ \sum_{i\in\mathcal{I}} \pi_{i(k-N+1)} = 1 & \text{for } N \le k \le 2N-2, \\ \sum_{i,j\in\mathcal{I}} \pi_{ij} = N & \text{for } k = 2N-1. \end{cases} \tag{13}$$

Thus, combining (5), (8), (11) and (13), we have that for each $n \in \mathbb{N}_0$,

$$\begin{aligned} W(n) &= W(0) + AI(n) - \sum_{\pi\in\Pi} \sum_{i,j\in\mathcal{I}} A^{ij}\pi_{ij} T_\pi(n) + AU(n) \\ &= W(0) + AI(n) - vn + AU(n), \end{aligned} \tag{14}$$

where $v \in \mathbb{R}^{2N-1}_+$ is such that $v_k = 1$ for $1 \le k \le 2N-2$ and $v_{2N-1} = N$.

The following lemma shows that the matrix A has full row rank.

LEMMA 2.1. *The rows of the matrix A are linearly independent, hence A has row rank $2N-1$.*

PROOF. Let $\{A_k : k = 1,\dots,2N-1\}$ denote the set of rows of the matrix A and let $\{c_k : k = 1,\dots,2N-1\}$ be a set of constants such that $\sum_{k=1}^{2N-1} c_k A_k = 0$. For $1 \le i,j \le N-1$, the $((i-1)N+j)^{th}$ component of this equation yields $c_i + c_{N-1+j} + c_{2N-1} = 0$ and the $((i-1)N+N)^{th}$ component yields $c_i + c_{2N-1} = 0$. It follows that $c_{N-1+j} = 0$ for each $1 \le j \le N-1$. In particular, $c_N = 0$. Moreover, we can also observe that the $((N-1)N+1)^{th}$ entry of $\sum_{k=1}^{2N-1} c_k A_k = 0$ yields $c_N + c_{2N-1} = 0$. Then we get that $c_{2N-1} = 0$, and hence $c_i = 0$ for each $1 \le i \le N-1$. That proves the lemma. □

Next we introduce two additional matrices. Let B be the $2N \times N^2$ matrix such that $(e_i', e_j')'$ is the $((i-1)N+j)^{th}$ column of B for $i,j \in \mathcal{I}$, where for each $i \in \mathcal{I}$, $e_i \in \mathbb{R}^N$ is the i^{th} unit coordinate vector in $\mathbb{R}^N$. For example, when $N = 2$, the matrix B has the form

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \tag{15}$$

It is not difficult to see that the matrix B has row rank $2N-1$. In fact, BQ would define a symmetric version of workload. Let C be the $2N \times (2N-1)$ matrix such that

$$C_{ij} = \begin{cases} 1, & 1 \le i = j \le N-1 \text{ or } N+1 \le i = j+1 \le 2N-1, \\ 1, & i = N,\ j = 2N-1 \text{ or } i = 2N,\ j = 2N-1, \\ -1, & i = N,\ 1 \le j \le N-1 \text{ or } i = 2N,\ N \le j \le 2N-2, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

It is readily verified that

$$CA = B. \tag{17}$$

We end this subsection by introducing the following rank property of the matrix AB'.

LEMMA 2.2. *The $(2N-1) \times 2N$ matrix AB' has full row rank.*

PROOF. It suffices to show the null space relation:

$$\mathcal{N}(AB') = \mathcal{N}(B'), \tag{18}$$

since if this holds, from the fact that B' has column rank $2N-1$, we have $\mathcal{N}(AB')$ is one-dimensional, and then since AB' is a $(2N-1) \times 2N$ matrix, AB' has full row rank.

To establish (18), it is enough to show that $\mathcal{N}(AB') \subseteq \mathcal{N}(B')$ since the reverse inclusion is clear. Suppose $x \in \mathcal{N}(AB')$, that is, $AB'x = 0$. Then, by (17), $BB'x = CAB'x = 0$. Hence $|B'x|^2 = x'BB'x = 0$. This implies that $B'x = 0$, thus, $x \in \mathcal{N}(B')$ and so (18) holds. □

2.5. *State descriptor.* The dynamics of the switch are described by the following collection of processes:

$$\chi(n) = (I(n), \{T_\pi(n) :\ \pi \in \Pi\}, D(n), U(n), Q(n), W(n)), \qquad n \in \mathbb{N}_0.$$

We extend $\chi(\cdot)$, and its constituent processes, to be defined on $\mathbb{R}_+$ such that for each $t \in \mathbb{R}_+$,

$$\chi(t) = \chi(\lfloor t \rfloor), \tag{19}$$

where $\lfloor t \rfloor$ denotes the integer part of t.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

3. Sequence of systems and scaling. To establish a diffusion approximation for our switch model, we now consider a sequence of switch models indexed by r where r tends to infinity through a sequence of positive real values. (To ease notation, we suppress the sequence indexing on r.) The basic switch structure with associated matrices A, B and C does not vary with r. Each member of the sequence is a stochastic system as described in the previous section. We append a superscript of r to any process, sequence of random variables or parameter associated with the r^{th} system that can vary with r. Thus, we have processes $I^r, \{T^r_\pi : \pi \in \Pi\}, U^r, D^r, Q^r, W^r$, sequences of random variables $\{\vartheta^r_{ij}(k) : k \in \mathbb{N}\}$ for $i, j \in \mathcal{I}$, and parameters λ^r_{ij} and b^r_{ij} for $i, j \in \mathcal{I}$. The r^{th} switch model has the associated state descriptor

$$\chi^r(\cdot) = (I^r(\cdot), \{T^r_\pi(\cdot) : \pi \in \Pi\}, D^r(\cdot), U^r(\cdot), Q^r(\cdot), W^r(\cdot)).$$

3.1. *Heavy traffic assumption.* To obtain a heavy traffic diffusion approximation for our sequence of switch models, we impose the following heavy traffic condition. Let λ^r denote the N^2-dimensional vector determined by $\{\lambda^r_{ij} : i, j \in \mathcal{I}\}$ such that λ^r_{ij} is the $((i-1)N + j)^{th}$ entry of λ^r.

Assumption 3.1 (Heavy Traffic). *There exist $\lambda \in \mathbb{R}^{N^2}_+$ with $\lambda > 0$ and $\theta \in \mathbb{R}^{2N-1}$ such that as $r \to \infty$,*

$$\lambda^r \to \lambda, \tag{20}$$

$$r(A\lambda^r - v) \to \theta, \tag{21}$$

where $v \in \mathbb{R}^{2N-1}$ is the vector in (14).

We note that (20)–(21) imply that $A\lambda = v$. This condition can be interpreted as meaning that in the heavy traffic limit, all of the input ports and output ports are heavily loaded, that is, $\sum_{j=1}^N \lambda_{ij} = 1$ for $i \in \mathcal{I}$ and $\sum_{i=1}^N \lambda_{ij} = 1$ for $j \in \mathcal{I}$. Thus, λ is a doubly stochastic matrix and by the Birkhoff-von Neumann theorem (see [2], Theorem 5.6 for a proof), it can be expressed as a convex combination of the permutation matrices $\pi \in \Pi$.

3.2. *Diffusion scaling.* Let us consider the following diffusion scaling. For $r > 0$ and $t \in \mathbb{R}_+$, we define

$$\widehat{I}^r(t) = \frac{1}{r}(I^r(r^2 t) - r^2 \lambda^r t), \tag{22}$$

$$\widehat{U}^r(t) = \frac{1}{r} U^r(r^2 t), \tag{23}$$

$$\widehat{Q}^r(t) = \frac{1}{r}Q^r(r^2t), \tag{24}$$

$$\widehat{W}^r(t) = \frac{1}{r}W^r(r^2t) = A\widehat{Q}^r(t). \tag{25}$$

For each $t \in \mathbb{R}_+$, it follows from (14) and (22)–(25) that

$$\begin{aligned}\widehat{W}^r(t) &= A\widehat{Q}^r(t)\\ &= \widehat{W}^r(0) + A\widehat{I}^r(t) + r(A\lambda^r - v)t + v\left(rt - \frac{\lfloor r^2t\rfloor}{r}\right) + A\widehat{U}^r(t).\end{aligned} \tag{26}$$

For each $i, j \in \mathcal{I}$, recall that A^{ij} denotes the $((i-1)N + j)^{th}$ column of the matrix A, and let

$$\gamma^{ij} = \frac{A^{ij}}{|A^{ij}|} \quad \text{and} \quad \widehat{Y}^r_{ij}(\cdot) = |A^{ij}|\widehat{U}^r_{ij}(\cdot). \tag{27}$$

Note that since T^r_π, Q^r only jump at positive integer times, it follows from (9) that $\widehat{Y}^r_{ij}$ can jump at time s only if $\widehat{Q}^r_{ij}(s-) = 0$ and so for each $i, j \in \mathcal{I}$,

$$\widehat{Y}^r_{ij}(t) = \int_{(0,t]} 1_{\{\widehat{Q}^r_{ij}(s-)=0\}}\, d\widehat{Y}^r_{ij}(s), \qquad t \geq 0. \tag{28}$$

Now by (26), $\widehat{W}^r(\cdot)$ and $\{\widehat{Y}^r_{ij}(\cdot) : i, j \in \mathcal{I}\}$ satisfy the following for all $t \geq 0$,

$$\widehat{W}^r(t) = \widehat{W}^r(0) + \widehat{X}^r(t) + \sum_{i,j\in\mathcal{I}} \widehat{Y}^r_{ij}(t)\gamma^{ij}, \tag{29}$$

where

$$\widehat{X}^r(t) = A\widehat{I}^r(t) + r(A\lambda^r - v)t + v\left(rt - \frac{\lfloor r^2t\rfloor}{r}\right). \tag{30}$$

3.3. *Functional central limit theorem for stochastic primitives.* We now introduce assumptions which will imply a functional central limit theorem for the diffusion scaled external packet arrival processes $\{\widehat{I}^r : r > 0\}$.

ASSUMPTION 3.2. *For each r and $i, j \in \mathcal{I}$, $\{\vartheta^r_{ij}(k) : k \in \mathbb{N}\}$ is independent of the initial queue-length $Q^r(0)$, $\{\vartheta^r_{ij}(k) : k \in \mathbb{N}\}$ are i.i.d. with finite mean λ^r_{ij} and finite variance b^r_{ij} satisfying (20)–(21) and*

$$b^r_{ij} \to b_{ij} \text{ as } r \to \infty, \tag{31}$$

and the following condition (of Lindeberg type) holds:

$$\sup_r E\left[(\vartheta^r_{ij}(1))^2;\ \vartheta^r_{ij}(1) > K\right] \to 0 \text{ as } K \to \infty. \tag{32}$$

For each r, the sequences $\{\vartheta^r_{ij}(k) : k \in \mathbb{N}\}$ for $i, j \in \mathcal{I}$ are mutually independent.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

PROPOSITION 3.1. *Let ν be a probability measure on $\mathbb{R}^{2N-1}$. Suppose that Assumptions 3.1 and 3.2 hold and $\widehat{W}^r(0)$ converges in distribution as $r \to \infty$ to a $(2N-1)$-dimensional random variable $\widetilde{W}_0$ with distribution ν. Then,*

$$(\widehat{W}^r(0), \widehat{I}^r) \text{ converges in distribution to } (\widetilde{W}_0, \widetilde{I}) \text{ as } r \to \infty, \tag{33}$$

where $\widetilde{W}_0$ is independent of $\widetilde{I}$ and $\widetilde{I}$ is an N^2-dimensional Brownian motion starting from the origin that has zero drift and $N^2 \times N^2$ diagonal covariance matrix Ξ, whose $((i-1)N+j)^{th}$ diagonal entry is b_{ij} for $i, j \in \mathcal{I}$.

This proposition follows directly from the standard functional central limit theorems for triangular arrays (cf. Theorem 18.2 of [3]). We assume henceforth that Assumptions 3.1 and 3.2 hold.

4. Fluid model, invariant states and multiplicative state space collapse. In this section, we consider a critical fluid model for the input-queued switch. We specialize to our situation some items developed for switched networks with maximum-weight-like policies in [19], namely, the form of the model, the characterization of invariant states and sufficient conditions for multiplicative state space collapse. Throughout this section, we assume the critical loading condition $A\lambda = v$ for $\lambda \in \mathbb{R}_+^{N^2}$ satisfying $\lambda > 0$.

4.1. *Fluid model solutions.* Fluid model solutions can be thought of as being obtained as formal limits of $\{\chi^r(\cdot)\}$ under law of large numbers scaling. The following terminology is used below. A function $f = (f_1, \ldots, f_d) : [0,\infty) \to \mathbb{R}_+^d$ is absolutely continuous if each of its components $f_i : [0,\infty) \to \mathbb{R}_+$, $i = 1, \ldots, d$, is absolutely continuous. A regular time for an absolutely continuous function $f : [0,\infty) \to \mathbb{R}_+^d$ is a value of $t \in (0,\infty)$ at which each component of f is differentiable. (Since f is absolutely continuous, almost every time $t \in (0,\infty)$ is a regular time for f and f can be recovered via integration of its a.e. defined derivative.) A uniformly Lipschitz continuous function is absolutely continuous.

DEFINITION 4.1. A fluid model solution is a continuous function $\overline{Q} : [0,\infty) \to \mathbb{R}_+^{N^2}$ such that there exist two families of nondecreasing, nonnegative, continuous functions $\{\overline{U}_{ij} : i, j \in \mathcal{I}\}$ and $\{\overline{T}_\pi : \pi \in \Pi\}$ defined from $[0,\infty)$ into $\mathbb{R}_+$ satisfying for all $i, j \in \mathcal{I}$ and $t \in \mathbb{R}_+$,

$$\overline{Q}_{ij}(t) = \overline{Q}_{ij}(0) + \lambda_{ij} t - \sum_{\pi \in \Pi} \pi_{ij} \overline{T}_\pi(t) + \overline{U}_{ij}(t), \tag{34}$$

$$\sum_{\pi\in\Pi} \overline{T}_\pi(t) = t, \tag{35}$$

$$\overline{U}_{ij}(0) = 0, \quad \int_0^\infty 1_{\{\overline{Q}_{ij}(s)>0\}} d\overline{U}_{ij}(s) = 0, \tag{36}$$

and for all $\pi \in \Pi$,

$$\int_0^\infty 1_{\{\omega_\pi(\overline{Q}(s))<\max_{\sigma\in\Pi}\omega_\sigma(\overline{Q}(s))\}} d\overline{T}_\pi(s) = 0. \tag{37}$$

It is clear that $\{\overline{T}_\pi : \pi \in \Pi\}$ are Lipschitz continuous with Lipschitz constant one and it follows from the oscillation inequality (cf. Theorem 5.1 of Williams [22]) for solutions of the one-dimensional Skorokhod problem that $\{\overline{Q}_{ij},\ \overline{U}_{ij} : i, j \in \mathcal{I}\}$ are uniformly Lipschitz continuous as well. In particular, almost every $t > 0$ is a regular time for $\{\overline{T}_\pi : \pi \in \Pi\}, \overline{Q}, \overline{U}$, and at such a time, for all $i, j \in \mathcal{I}$,

$$\frac{d}{dt}\overline{Q}_{ij}(t) = \begin{cases} \lambda_{ij} - \sum_{\pi\in\Pi} \pi_{ij} \frac{d}{dt}\overline{T}_\pi(t) & \text{if } \overline{Q}_{ij}(t) > 0, \\ 0 & \text{if } \overline{Q}_{ij}(t) = 0, \end{cases} \tag{38}$$

and all $\pi \in \Pi$,

$$\frac{d}{dt}\overline{T}_\pi(t) = 0 \text{ if } \exists\ \sigma \in \Pi \text{ such that } \omega_\sigma(\overline{Q}(t)) > \omega_\pi(\overline{Q}(t)). \tag{39}$$

REMARK 4.1. Our fluid model description is equivalent to that in Definition 4.1 of [19].

4.2. *Invariant states.*

DEFINITION 4.2. A state $q \in \mathbb{R}_+^{N^2}$ is an invariant state for the fluid model if there exists a fluid model solution $\overline{Q}(\cdot)$ such that $\overline{Q}(t) = q$ for all $t \in \mathbb{R}_+$.

For the situation of critical loading ($A\lambda = v$) treated here, the following is a version of an optimization problem considered by Shah and Wischik [19]; see Definition 5.3 therein with $f(x) = x$. (In fact, Shah and Wischik [19] treated a more general situation, where some input and output ports can be underloaded.) Using this optimization problem, results in [19] yield a characterization of the invariant states of the fluid model and the property that fluid model solutions converge towards the manifold of invariant states as time goes to infinity. Indeed, as shown in [19], any fluid model solution $\overline{Q}$ satisfies the constraints of the optimization problem for all time with $w = A\overline{Q}(0)$ and the objective function in the problem can be used to create a Lyapunov function to prove the convergence towards the invariant manifold.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

DEFINITION 4.3. Given $w \in \mathbb{R}_+^{2N-1}$, consider the optimization problem:

$$\begin{aligned} &\text{minimize} && |x|^2 \\ &\text{subject to} && Bx \geq Cw, \\ &\text{over} && x \in \mathbb{R}_+^{N^2}. \end{aligned} \tag{40}$$

Since the function $x \mapsto |x|^2$ is strictly convex, and tends to ∞ as $|x| \to \infty$, and the feasible set of (40) is non-empty, closed, convex, and bounded below, then (40) has a unique solution. Let $\Delta(w)$ denote the unique optimal solution. Since the constraints are linear, we can represent $\Delta(w)$ in terms of Lagrange multipliers (see [5], Proposition 3.4.1):

$$\Delta(w) = B'p, \qquad p \in \mathbb{R}_+^{2N}, \qquad \langle p, (B\Delta(w) - Cw)\rangle = 0. \tag{41}$$

Note that p here need not be unique. The function Δ is called the lifting map, which has the following properties.

PROPOSITION 4.1. *The function $\Delta : \mathbb{R}_+^{2N-1} \to \mathbb{R}_+^{N^2}$ is continuous. Furthermore, for each $w \in \mathbb{R}_+^{2N-1}$ and $c > 0$,*

$$\Delta(cw) = c\Delta(w). \tag{42}$$

PROOF. The first property follows from Corollary II.3.1 and Corollary I.3.4 of Dantzig et al. [6]. The second property follows by the same argument as in the proof of the second property in Proposition 4.1 of [12]. □

The next proposition provides a representation of the invariant states.

PROPOSITION 4.2. *A state $q \in \mathbb{R}_+^{N^2}$ is an invariant state if and only if $q = \Delta(Aq)$.*

PROOF. It follows from Lemma 5.11 of [19] that $q \in \mathbb{R}_+^{N^2}$ is an invariant state if and only if $q = \breve{\Delta}\breve{W}(q)$, where the map $\breve{\Delta}\breve{W}$ is defined in Definition 5.3 of [19]. (Here we use $\breve{\Delta}\breve{W}$ instead of the ΔW in [19] to distinguish from the symbols used in this paper). The map $\breve{\Delta}\breve{W}$ is the composition of two maps $\breve{\Delta}$ and $\breve{W}$, where $\breve{\Delta}$ is defined by (35) of [19] and $\breve{W}$ is a workload map defined right before (35) of [19]. In the setting of input-queued switches, Theorem 8.3 of [19] shows that $\breve{W}(q) = Bq = CAq$ and with Lemma A.1 of [19] we see that $\breve{\Delta}\breve{W}(q) = \Delta(Aq)$, and the desired result follows. □

We now show that the lifting map Δ maps $\mathbb{R}_+^{2N-1}$ into the set of invariant states and we give a representation for it.

LEMMA 4.1. *For each $w \in \mathbb{R}_+^{2N-1}$, $\Delta(w)$ is an invariant state. Furthermore, for $w \in \mathcal{W} \doteq \{AB'p : \ p \in \mathbb{R}_+^{2N}\}$,*

$$\Delta(w) = B'(AB')^\dagger w, \tag{43}$$

where the $2N \times (2N-1)$ matrix $(AB')^\dagger$ is the Moore-Penrose pseudo-inverse of AB' which satisfies $AB'(AB')^\dagger = I_{2N-1}$ and I_{2N-1} is the $(2N-1)\times(2N-1)$ identity matrix.

PROOF. Fix $w \in \mathbb{R}_+^{2N-1}$. It follows from (41) that there is a $p \in \mathbb{R}_+^{2N}$ such that $\Delta(w) = B'p$. By Proposition 4.2, to show that $\Delta(w)$ is an invariant state, it suffices to show that

$$B'p = \Delta(AB'p). \tag{44}$$

As a consequence of (41) again, there is a $\tilde{p} \in \mathbb{R}_+^{2N}$ such that

$$\Delta(AB'p) = B'\tilde{p} \qquad \text{and} \qquad \big\langle \tilde{p}, (B\Delta(AB'p) - CAB'p)\big\rangle = 0.$$

Then, by using (17) and the above display, we obtain

$$\big\langle B'\tilde{p}, (B'\tilde{p} - B'p)\big\rangle = 0. \tag{45}$$

On the other hand, since $B'\tilde{p} = \Delta(AB'p)$ is the unique solution to the optimization problem in (40) with $AB'p$ in place of w, then $B'\tilde{p}$ satisfies the constraint in (40). Thus,

$$BB'\tilde{p} \geq CAB'p.$$

This and (17) imply that $B(B'\tilde{p} - B'p) \geq 0$. Since $p \geq 0$, we then have that $\langle p, B(B'\tilde{p} - B'p)\rangle \geq 0$, and so, $\langle B'p, (B'\tilde{p} - B'p)\rangle \geq 0$. It follows from this and (45) that

$$0 \leq |B'\tilde{p} - B'p|^2 = \big\langle B'\tilde{p} - B'p, B'\tilde{p} - B'p\big\rangle \leq 0.$$

As a result, $B'p = B'\tilde{p} = \Delta(AB'p)$.

To establish the second part of the lemma, let $w \in \mathcal{W}$. There exists a $p \in \mathbb{R}_+^{2N}$ such that $w = AB'p$. It follows from the argument above that

$$\Delta(w) = \Delta(AB'p) = B'p. \tag{46}$$

On the other hand, since AB' has full row rank, its Moore-Penrose pseudoinverse $(AB')^\dagger$ satisfies $AB'(AB')^\dagger = I_{2N-1}$. Then we have

$$AB'(AB')^\dagger w = w = AB'p.$$

Thus, by (18), $p - (AB')^\dagger w \in \mathcal{N}(AB') = \mathcal{N}(B')$ and then $B'p = B'(AB')^\dagger w$. This and (46) imply that (43) holds. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

4.3. *Multiplicative state space collapse.*

DEFINITION 4.4 (MSSC). We say that multiplicative state space collapse holds if for each $T \in \mathbb{R}_+$,

$$\frac{\left|\left|\widehat{Q}^r(\cdot) - \Delta(\widehat{W}^r(\cdot))\right|\right|_T}{\left|\left|\widehat{Q}^r(\cdot)\right|\right|_T \vee 1} \to 0 \text{ in probability as } r \to \infty. \tag{47}$$

REMARK 4.2. If the above holds without the factor in the denominator of (47), then *state space collapse* is said to hold, that is, for each $T \in \mathbb{R}_+$,

$$\left|\left|\widehat{Q}^r(\cdot) - \Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_T \to 0 \text{ in probability as } r \to \infty. \tag{48}$$

Shah and Wischik [19] adapted a method of Bramson [4] and combined it with asymptotic behavior of fluid model solutions to give sufficient conditions for multiplicative state space collapse to hold for switched networks operating under maximum-weight-like policies. The following proposition is a consequence of their results specialized to single-hop networks and the maximum weight matching policy considered here.

PROPOSITION 4.3. *Suppose that Assumptions 3.1 and 3.2 hold and for each $r > 0$, the initial queue sizes $\{Q^r_{ij}(0) : i, j \in \mathcal{I}\}$ are non-random and satisfy $\lim_{r\to\infty} \widehat{Q}^r(0) = q_0$ for some invariant state q_0. In addition, suppose that there exists a sequence $\{\delta_n : n \in \mathbb{N}\} \subset \mathbb{R}_+$ such that $\delta_n \to 0$ as $n \to \infty$ and*

$$n(\log n)^2 \sup_r P\left(\sup_{k\le n} |I^r_{ij}(k) - \lambda^r_{ij}k| \ge n\delta_n\right) \to 0 \text{ as } n \to \infty. \tag{49}$$

Then multiplicative state space collapse holds.

PROOF. The input queued switch model in this paper is a special case of the single-hop switched network considered in [19], and the maximum weight matching policy considered here is the max-weight policy (or MW-f policy with $f(x) = x$) considered in [19]. For each $r > 0$, the arrival process $I^r(\cdot)$ has stationary increments with mean arrival rate vector λ^r. It follows from (20)–(21) that $\lambda^r \to \lambda$ as $r \to \infty$ and $A\lambda = v$, which means all of the input ports and output ports are heavily loaded in the limit. By virtue of the fact that λ can be expressed as a convex combination of permutation matrices (see the discussion at the end of Section 3.1), λ is admissible in the sense of

Definition 5.1 of [19]. Although r is restricted to $\mathbb{N}$ in [19], the proofs apply whenever $r > 0$ tends to infinity through an increasing sequence of values. With this observation, the proposition follows directly from Theorem 7.1 of [19]. □

REMARK 4.3. As noted following Assumption 2.5 in [19], the last condition of Proposition 4.3 is satisfied in our setting if for each $i, j \in \mathcal{I}$, the sequence $\{\vartheta_{ij}^r(k) : k \in \mathbb{N}\}$ satisfies a uniform fourth moment bound of the following type:

$$\sup_r E\left[|\vartheta_{ij}^r(1)|^4\right] < \infty \quad \text{for all } i, j \in \mathcal{I}.$$

5. Diffusion approximation. Let

$$\mathcal{W} = \left\{AB'p : p \in \mathbb{R}_+^{2N}\right\}. \tag{50}$$

It will be shown in Lemma 6.1 that the polyhedron $\mathcal{W}$ has N^2 boundary faces given by

$$\mathcal{B}_{ij} = \left\{AB'p : p \in \mathbb{R}_+^{2N} \text{ satisfying } p_i = p_{N+j} = 0\right\} \quad \text{for } i, j \in \mathcal{I}. \tag{51}$$

Recall the diagonal matrix Ξ from Proposition 3.1. Let $\Gamma = A\Xi A'$, which is a $(2N-1)\times(2N-1)$ covariance matrix. We assume that Γ is non-degenerate, i.e., it is strictly positive definite. (This holds if $b_{ij} > 0$ for all $i, j \in \mathcal{I}$, for example.) Also let μ be a probability measure on $\mathcal{W}$, where $\mathcal{W}$ is endowed with the Borel σ-algebra of $\mathbb{R}^{2N-1}$. Recall γ^{ij} defined in (27) for each $i, j \in \mathcal{I}$ and θ in (21).

DEFINITION 5.1. A Semimartingale Reflecting Brownian Motion that lives in the polyhedral cone $\mathcal{W}$, has direction of reflection γ^{ij} on the boundary face $\mathcal{B}_{ij}$ for each $i, j \in \mathcal{I}$, has drift θ and covariance matrix Γ, and has initial distribution μ on $\mathcal{W}$, is an $\{\mathcal{F}_t\}$-adapted, $(2N-1)$-dimensional process $\widetilde{W}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that

(i) P-a.s., $\widetilde{W}(t) = \widetilde{W}(0) + \widetilde{X}(t) + \sum_{i,j\in\mathcal{I}} \widetilde{Y}_{ij}(t)\gamma^{ij}$ for all $t \in \mathbb{R}_+$,

(ii) P-a.s., $\widetilde{W}$ has continuous paths, $\widetilde{W}(t) \in \mathcal{W}$ for all $t \in \mathbb{R}_+$, and $\widetilde{W}(0)$ has distribution μ,

(iii) under P,

- **(a)** $\widetilde{X}$ is a $(2N-1)$-dimensional Brownian motion starting from the origin with drift θ and covariance matrix Γ,
- **(b)** $\{\widetilde{X}(t) - \theta t,\ \mathcal{F}_t,\ t \in \mathbb{R}_+\}$ is a martingale,

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

(iv) for each $i, j \in \mathcal{I}$, $\widetilde{Y}_{ij}$ is an $\{\mathcal{F}_t\}$-adapted, one-dimensional process such that P-a.s.,

(a) $\widetilde{Y}_{ij}(0) = 0$,

(b) $\widetilde{Y}_{ij}$ is continuous and non-decreasing,

(c) $\widetilde{Y}_{ij}(t) = \int_0^t 1_{\{\widetilde{W}(s) \in \mathcal{B}_{ij}\}} d\widetilde{Y}_{ij}(s),\ t \in \mathbb{R}_+$.

REMARK 5.1. We call a process that satisfies the above properties an SRBM associated with the data $(\mathcal{W}, \{\gamma^{ij} : i, j \in \mathcal{I}\}, \theta, \Gamma, \mu)$. Geometric conditions for existence and uniqueness in law of an SRBM are given in Dai and Williams [8]. It will be shown in the course of proving our main result, Theorem 5.1, that these are satisfied by $(\mathcal{W}, \{\gamma^{ij} : i, j \in \mathcal{I}\})$.

We now state the main result of this paper.

THEOREM 5.1. *Suppose that Assumptions 3.1 and 3.2 and multiplicative state space collapse all hold. Let $\mathcal{W}$, $\{\gamma^{ij} : i, j \in \mathcal{I}\}$, θ, Γ be as described at the beginning of this section. Suppose that $\widehat{W}^r(0)$ converges in distribution to a random variable with distribution μ on $\mathcal{W}$. Then $(\widehat{W}^r, \widehat{Q}^r)$ converges in distribution as $r \to \infty$ to a continuous process $(\widetilde{W}, \widetilde{Q})$, where $\widetilde{W}$ is an SRBM associated with the data $(\mathcal{W}, \{\gamma^{ij} : i, j \in \mathcal{I}\}, \theta, \Gamma, \mu)$ and $\widetilde{Q} = \Delta(\widetilde{W})$.*

The following corollary follows directly from Theorem 5.1 and Proposition 4.3.

COROLLARY 5.1. *Suppose that all of the assumptions in Proposition 4.3 hold and Γ is non-degenerate. Then the conclusion in Theorem 5.1 holds, where μ is the point mass at Aq_0.*

As an illustration of Corollary 5.1, consider a 2×2 input-queued switch operated under the MWM policy. The workload space $\mathcal{W}$ in this case is a polyhedral cone lying strictly inside $\mathbb{R}^3_+$ that has the following representation:

$$\mathcal{W} = \left\{ \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2p_1 + p_3 + p_4 \\ p_1 + p_2 + 2p_3 \\ 2p_1 + 2p_2 + 2p_3 + 2p_4 \end{bmatrix} : \ p \in \mathbb{R}^4_+ \right\}. \tag{52}$$

The set of directions of reflection $\{\gamma^{ij} : \ 1 \le i, j \le 2\}$ associated with the four boundary faces is given by

$$\gamma^{11} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \gamma^{12} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \gamma^{21} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \gamma^{22} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

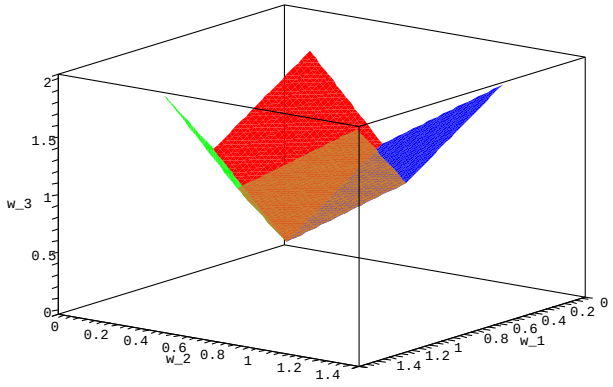


FIG 2. *A portion of the polyhedral workload cone $\mathcal{W}$ is shown for a 2×2 input-queued switch.*

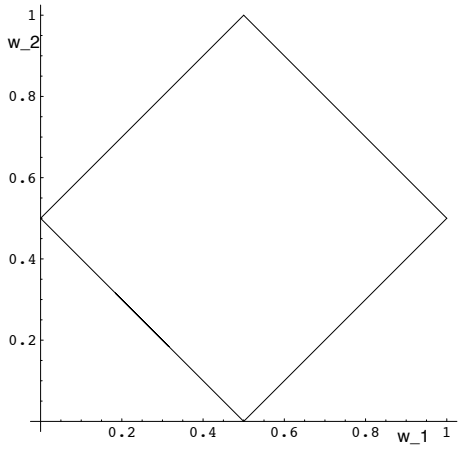


FIG 3. *Cross-section of the workload cone $\mathcal{W}$ depicted in Figure 2 taken at $w_3 = 1$.*

The workload space $\mathcal{W}$ in this case is depicted in Figures 2–3. The conclusion of Corollary 5.1 applies under Assumptions 3.1 and 3.2, provided Γ is non-degenerate and the initial queue sizes $\{Q_{ij}^r(0) : \ i, j \in \mathcal{I}\}$ are non-random and satisfy $\lim_{r\to\infty} \widehat{Q}^r(0) = q_0$ for some invariant state q_0. Note that the latter holds if the VOQs in each of the switches indexed by r start empty and μ is the point mass at the origin in $\mathbb{R}^3_+$. Then the limiting SRBM $\widetilde{W}$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

lives in $\mathcal{W}$. It behaves like Brownian motion with a constant drift θ and covariance matrix Γ inside $\mathcal{W}$ and it is confined to $\mathcal{W}$ by instantaneous reflection (or pushing) at the boundary where the direction of reflection on $\mathcal{B}_{ij}$ is given by γ^{ij}, $i,j \in \mathcal{I}$. This pushing on the boundary corresponds to an underutilization of processing capacity (or transmittal of blanks) in the original system, where pushing on boundary face $\mathcal{B}_{ij}$ corresponds to the situation where VOQ_{ij} is empty and so there are no packets to transmit from there and processing effort allocated there is wasted. In the diffusion approximation, this results in an instantaneous increase in $\widetilde{W}_i$ (for $i \neq N$), $\widetilde{W}_{N-1+j}$ (for $j \neq N$) and $\widetilde{W}_{2N-1}$ on $\mathcal{B}_{ij}$. We thus see the constraints imposed by the switch architecture due to simultaneous processing from more than one queue are encoded in the geometry of the approximating diffusion.

6. Proof of diffusion approximation. In this section we prove the diffusion approximation result, Theorem 5.1. Throughout this section, we assume that the hypotheses of Theorem 5.1 hold. For the proof of this theorem, we apply an invariance principle for SRBMs that we developed in [13]. To obtain the convergence of $\widehat{W}^r$, we will apply Theorem 5.4 of [13]. To justify use of this theorem, we need to verify Assumptions (A1)–(A5), 4.1, and 5.1 of [13]. We first verify, in Section 6.1, that the workload space $\mathcal{W}$ and the directions of reflection $\{\gamma^{ij} : i,j \in \mathcal{I}\}$ satisfy the geometric properties in Assumptions (A1)–(A5) and 5.1 of [13]. The verification of Assumption 4.1 of [13] relies on an oscillation inequality that we prove in this paper in Section 6.2. In Section 6.3, using the oscillation inequality, we prove that state space collapse follows from multiplicative state space collapse. Finally, in Section 6.4 we verify Assumption 4.1 of [13], which ensures that $\widehat{W}^r$ satisfies a perturbed version of the SRBM definition, and we prove the convergence of $\widehat{W}^r$ by applying Theorem 5.4 of [13]. The convergence jointly of $\widehat{Q}^r$ with $\widehat{W}^r$ then follows from the state space collapse already established in Section 6.3.

6.1. *Verification of geometric conditions.* Recall that the workload space $\mathcal{W}$ is given by (50), where AB' is a $(2N-1)\times(2N)$ matrix that has the following entries: for $1 \le k \le 2N-1$ and $1 \le l \le 2N$,

$$(AB')_{kl} = \begin{cases} N, & 1 \le k = l \le N-1, \\ N, & N \le k \le 2N-2,\ l = k+1, \\ N, & k = 2N-1,\ 1 \le l \le 2N, \\ 1, & 1 \le k \le N-1,\ N+1 \le l \le 2N, \\ 1, & N \le k \le 2N-2,\ 1 \le l \le N, \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

For example, when $N = 2$, the matrix AB' is

$$AB' = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 2 & 2 & 2 & 2 \end{bmatrix}. \tag{54}$$

Recall from Lemma 2.2 that AB' has full row rank. Let $(AB')^l$ denote the l^{th} column of AB' for $1 \le l \le 2N$. Then from (53), we have

$$\sum_{l=1}^{N} (AB')^l = \sum_{l=1}^{N} (AB')^{N+l}. \tag{55}$$

Define a set of unit vectors $\{n^{ij} : i, j \in \mathcal{I}\}$ as follows:

$$n^{ij} = \begin{cases} \frac{N\eta_i + N\eta_{N-1+j} - \eta_{2N-1}}{(2N^2+1)^{1/2}}, & 1 \le i, j \le N-1, \\ \frac{N\eta_i - \sum_{k=N}^{2N-2} N\eta_k + (N-1)\eta_{2N-1}}{(N^3+(N-1)^2)^{1/2}}, & 1 \le i \le N-1, j = N, \\ \frac{N\eta_{N-1+j} - \sum_{k=1}^{N-1} N\eta_k + (N-1)\eta_{2N-1}}{(N^3+(N-1)^2)^{1/2}}, & 1 \le j \le N-1, i = N, \\ \frac{(2N-1)\eta_{2N-1} - \sum_{k=1}^{2N-2} N\eta_k}{((2N-2)N^2+(2N-1)^2)^{1/2}}, & i = N, j = N, \end{cases} \tag{56}$$

where η_k denotes the unit vector that is parallel to the positive k^{th} coordinate direction in $\mathbb{R}_+^{2N-1}$. We also define a set of strictly positive constants $\{K_{ij} : i, j \in \mathcal{I}\}$ as follows:

$$K_{ij} = \begin{cases} \frac{N^2}{(2N^2+1)^{1/2}}, & 1 \le i, j \le N-1, \\ \frac{N^2}{(N^3+(N-1)^2)^{1/2}}, & 1 \le i \le N-1, j = N, \\ \frac{N^2}{(N^3+(N-1)^2)^{1/2}}, & 1 \le j \le N-1, i = N, \\ \frac{N^2}{((2N-2)N^2+(2N-1)^2)^{1/2}}, & i = N, j = N. \end{cases} \tag{57}$$

Note that for each $i, j \in \mathcal{I}$ and $1 \le l \le 2N$,

$$\left\langle n^{ij}, (AB')^l \right\rangle = \begin{cases} K_{ij}, & \text{if } l = i \text{ or } l = N + j, \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

Define

$$\mathcal{G} = \bigcap_{i,j \in \mathcal{I}} \mathcal{G}_{ij}, \tag{59}$$

where for $i, j \in \mathcal{I}$,

$$\mathcal{G}_{ij} = \left\{ x \in \mathbb{R}^{2N-1} : \left\langle n^{ij}, x \right\rangle \ge 0 \right\}.$$

The following lemma shows that the polyhedral cone $\mathcal{W} = \mathcal{G}$ and it has N^2 boundary faces given by $\{\mathcal{B}_{ij} : i, j \in \mathcal{I}\}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

LEMMA 6.1. *$\mathcal{W} = \mathcal{G}$ and $\mathcal{B}_{ij} = \partial\mathcal{G}_{ij} \cap \partial\mathcal{G}$ for $i, j \in \mathcal{I}$.*

PROOF. We first show that $\mathcal{W} \subseteq \mathcal{G}$. Let $x \in \mathcal{W}$. Then $x = AB'p$ for some $p \in \mathbb{R}^{2N}_+$. It follows from (58) that for each $i, j \in \mathcal{I}$,

$$\langle n^{ij}, x\rangle = \langle n^{ij}, AB'p\rangle = K_{ij}(p_i + p_{N+j}) \geq 0. \tag{60}$$

Hence $x \in \mathcal{G}$ and then $\mathcal{W} \subseteq \mathcal{G}$.

We next show that $\mathcal{G} \subseteq \mathcal{W}$. Let $x \in \mathcal{G} \subseteq \mathbb{R}^{2N-1}$. There is a $p \in \mathbb{R}^{2N}$ such that $x = AB'p$, since AB' has rank $2N-1$. Now we show that there is a $p^* \in \mathbb{R}^{2N}_+$ such that $x = AB'p^*$. We use p to construct p^*. If $p_1 \geq 0$, then we let $p^1 = p$. If $p_1 < 0$, we let p^1 be such that $p^1_i = p_i - p_1$ and $p^1_{N+i} = p_{N+i} + p_1$ for $i \in \mathcal{I}$. Then we have $p^1_1 = 0$ and by (55),

$$AB'p^1 = AB'p - \sum_{i\in\mathcal{I}}(AB')^i p_1 + \sum_{i\in\mathcal{I}}(AB')^{N+i}p_1 = AB'p = x.$$

If $p^1_2 \geq 0$, we let $p^2 = p^1$. If $p^1_2 < 0$, we let p^2 be such that $p^2_i = p^1_i - p^1_2$ and $p^2_{N+i} = p^1_{N+i} + p^1_2$ for $i \in \mathcal{I}$. Then by (55) again, we have that $x = AB'p^2$ and $p^2_1 \geq 0,\ p^2_2 \geq 0$. Continuing in this manner, we can construct $p^N \in \mathbb{R}^{2N}$ such that $x = AB'p^N$ and $p^N_i \geq 0$ for all $i \in \mathcal{I}$.

Now if $p^N_{N+1} \geq 0$, we let $p^{N+1} = p^N$. If $p^N_{N+1} < 0$, let p^{N+1} be such that $p^{N+1}_i = p^N_i + p^N_{N+1}$ and $p^{N+1}_{N+i} = p^N_{N+i} - p^N_{N+1}$ for $i \in \mathcal{I}$. By (60) and the fact that $x \in \mathcal{G}$, we know that for $1 \leq i \leq N$,

$$p^{N+1}_i = p^N_i + p^N_{N+1} = K^{-1}_{i1}\langle n^{i1}, AB'p^N\rangle = K^{-1}_{i1}\langle n^{i1}, x\rangle \geq 0$$

and $p^{N+1}_{N+1} = 0$. Using (55) again, we have that $x = AB'p^{N+1}$ and $p^{N+1}_i \geq 0$ for all $1 \leq i \leq N+1$. Continuing in this manner, we can construct $p^{2N} \in \mathbb{R}^{2N}$ such that $x = AB'p^{2N}$ and $p^{2N}_i \geq 0$ for all $1 \leq i \leq 2N$. Letting $p^* = p^{2N}$, we conclude that $\mathcal{G} \subseteq \mathcal{W}$.

To prove the second claim in the lemma, fix $i, j \in \mathcal{I}$. For each $x \in \mathcal{B}_{ij}$, we know that $x = AB'p$ for some $p \in \mathbb{R}^{2N}_+$ with $p_i = p_{N+j} = 0$. Then $\langle n^{ij}, x\rangle = 0$ by (60). Hence $x \in \partial\mathcal{G}_{ij}$ and since $x \in \mathcal{W} = \mathcal{G}$, it follows that $x \in \partial\mathcal{G}_{ij} \cap \partial\mathcal{G}$. On the other hand, for each $x \in \partial\mathcal{G}_{ij} \cap \partial\mathcal{G}$, since $\mathcal{W} = \mathcal{G}$, we have that $x = AB'p$ for some $p \in \mathbb{R}^{2N}_+$ and $\langle n^{ij}, x\rangle = 0$. Then by (60), we have that $p_i = p_{N+j} = 0$, hence $x \in \mathcal{B}_{ij}$. □

The next lemma establishes certain geometric properties of the set of directions of reflection $\{\gamma^{ij} :\ i, j \in \mathcal{I}\}$. For each $x \in \partial\mathcal{W}$, let $\mathcal{I}(x) = \{ij :\ x \in \mathcal{B}_{ij},\ i, j \in \mathcal{I}\}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

LEMMA 6.2. *The vectors* $\{\gamma^{ij} : i, j \in \mathcal{I}\}$ *satisfy the following condition: for each* $x \in \partial\mathcal{W}$, *there are positive constants* $\{b_{ij}(x) : ij \in \mathcal{I}(x)\}$ *and* $\{c_{ij}(x) : ij \in \mathcal{I}(x)\}$ *such that* $\sum_{ij\in\mathcal{I}(x)} b_{ij}(x) = 1$ *and* $\sum_{ij\in\mathcal{I}(x)} c_{ij}(x) = 1$,

$$\min_{ij\in\mathcal{I}(x)} \left\langle \sum_{kl\in\mathcal{I}(x)} b_{kl}(x) n^{kl}, \gamma^{ij} \right\rangle > 0, \tag{61}$$

$$\min_{ij\in\mathcal{I}(x)} \left\langle n^{ij}, \sum_{kl\in\mathcal{I}(x)} c_{kl}(x) \gamma^{kl} \right\rangle > 0. \tag{62}$$

PROOF. Recall that $N \geq 2$. We obtain by computation that for $i, j \in \mathcal{I}$ and $k, l \in \mathcal{I}$ such that $k \neq i$ and $l \neq j$,

$$\langle n^{ij}, \gamma^{ij} \rangle = \begin{cases} \frac{2N-1}{\sqrt{3}(2N^2+1)^{1/2}}, & 1 \leq i, j \leq N-1, \\ \frac{2N-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1 \leq i \leq N-1, j = N, \\ \frac{2N-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1 \leq j \leq N-1, i = N, \\ \frac{2N-1}{((2N-2)N^2+(2N-1)^2)^{1/2}}, & i = N, j = N, \end{cases} \tag{63}$$

$$\left\langle n^{ij}, \gamma^{il} \right\rangle = \begin{cases} \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}}, & 1 \leq i, j, l \leq N-1, \\ \frac{N-1}{\sqrt{2}(2N^2+1)^{1/2}}, & 1 \leq i, j \leq N-1, l = N, \\ \frac{N-1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}}, & 1 \leq i, l \leq N-1, j = N, \\ \frac{N-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1 \leq j, l \leq N-1, i = N, \\ \frac{N-1}{(N^3+(N-1)^2)^{1/2}}, & 1 \leq j \leq N-1, i = N, l = N, \\ \frac{N-1}{\sqrt{2}((2N-2)N^2+(2N-1)^2)^{1/2}}, & i = N, j = N, 1 \leq l \leq N-1, \end{cases} \tag{64}$$

$$\left\langle n^{ij}, \gamma^{kj} \right\rangle = \begin{cases} \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}}, & 1 \leq i, j, k \leq N-1, \\ \frac{N-1}{\sqrt{2}(2N^2+1)^{1/2}}, & 1 \leq i, j \leq N-1, k = N, \\ \frac{N-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1 \leq i, k \leq N-1, j = N, \\ \frac{N-1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}}, & 1 \leq j, k \leq N-1, i = N, \\ \frac{N-1}{(N^3+(N-1)^2)^{1/2}}, & 1 \leq i \leq N-1, j = N, k = N, \\ \frac{N-1}{\sqrt{2}((2N-2)N^2+(2N-1)^2)^{1/2}}, & i = N, j = N, 1 \leq k \leq N-1, \end{cases} \tag{65}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

(66)
$$\left\langle n^{ij},\gamma^{kl}\right\rangle = \begin{cases} \frac{-1}{\sqrt{3}(2N^2+1)^{1/2}}, & 1\le i,j,k,l\le N-1,\\ \frac{-1}{\sqrt{2}(2N^2+1)^{1/2}}, & 1\le i,j,k\le N-1, l=N,\\ \frac{-1}{\sqrt{2}(2N^2+1)^{1/2}}, & 1\le i,j,l\le N-1, k=N,\\ \frac{-1}{(2N^2+1)^{1/2}}, & 1\le i,j\le N-1, k=N, l=N,\\ \frac{-1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}}, & 1\le i,k,l\le N-1, j=N,\\ \frac{-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1\le i,l\le N-1, j=N, k=N,\\ \frac{-1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}}, & 1\le j,k,l\le N-1, i=N,\\ \frac{-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}}, & 1\le j,k\le N-1, i=N, l=N,\\ \frac{-1}{\sqrt{3}((2N-2)N^2+(2N-1)^2)^{1/2}}, & i=N, j=N, 1\le k,l\le N-1. \end{cases}$$

Fix $x\in\partial\mathcal{W}$. Notice that the set $\mathcal{I}(x)$ has the property that

$$ij,\ kl\in\mathcal{I}(x)\Longrightarrow il,\ kj\in\mathcal{I}(x). \tag{67}$$

In fact, if $ij,\ kl\in\mathcal{I}(x)$, then $x\in\mathcal{B}_{ij}\cap\mathcal{B}_{kl}$. Since $x\in\mathcal{W}=\mathcal{G}$, there is a $p\in\mathbb{R}^{2N}_+$ such that $x=AB'p$. Then by (60), we have that $p_i=p_{N+j}=p_k=p_{N+l}=0$. Hence by (60) again, we know that $\langle n^{il},x\rangle=\langle n^{kj},x\rangle=0$, then $il,\ kj\in\mathcal{I}(x)$.

For (62), it suffices to show that

$$\sum_{kl\in\mathcal{I}(x)}\left\langle n^{ij},\gamma^{kl}\right\rangle>0 \text{ for } ij\in\mathcal{I}(x), \tag{68}$$

because we can then set $c_{kl}(x)=\frac{1}{|\mathcal{I}(x)|}$ for $kl\in\mathcal{I}(x)$ and (62) holds. To show (68), fix $ij\in\mathcal{I}(x)$. Now,

$$\sum_{kl\in\mathcal{I}(x)}\gamma^{kl}=\sum_{k\in\mathcal{I}\setminus\{i\}}\sum_{l\in\mathcal{I}\setminus\{j\}}\left\{\gamma^{kl}1_{\{kl\in\mathcal{I}(x)\}} +\frac{1}{N-1}\left(\gamma^{il}1_{\{il\in\mathcal{I}(x)\}}+\gamma^{kj}1_{\{kj\in\mathcal{I}(x)\}}\right)\right\}+\gamma^{ij}. \tag{69}$$

Note from (63) that $\langle n^{ij},\gamma^{ij}\rangle>0$. Fix $k\neq i$, $l\neq j$, and consider the term between the braces $\{\}$ in (69). If $kl\notin\mathcal{I}(x)$, then the inner product of n^{ij} with the brace term is non-negative by (64)–(65). On the other hand, if $kl\in\mathcal{I}(x)$, then we know from (67) that $il,kj\in\mathcal{I}(x)$. We will show that, in this case, the inner product of n^{ij} with the brace term is strictly positive. There are a number of cases to consider.

Suppose first that $1\le i,j\le N-1$. If $1\le k,l\le N-1$ also, then

$$\left\langle n^{ij},\gamma^{kl}\right\rangle=\frac{-1}{\sqrt{3}(2N^2+1)^{1/2}},$$

$$\left\langle n^{ij}, \gamma^{il} \right\rangle = \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}} \quad \text{and} \quad \left\langle n^{ij}, \gamma^{kj} \right\rangle = \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}}.$$

Hence it follows that

$$\left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) = \frac{1}{\sqrt{3}(2N^2+1)^{1/2}} > 0.$$

If $k = N$ and $1 \le l \le N-1$ or $l = N$ and $1 \le k \le N-1$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \quad & \frac{-1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{1}{\sqrt{3}(2N^2+1)^{1/2}} \\ > \quad & 0. \end{aligned}$$

If $k = N$ and $l = N$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \quad & \frac{-1}{(2N^2+1)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} \\ > \quad & 0. \end{aligned}$$

Suppose next that $1 \le i \le N-1$ and $j = N$. If $1 \le k, l \le N-1$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \quad & \frac{-1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} + \frac{1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} \\ & \qquad + \frac{1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}} \\ > \quad & 0. \end{aligned}$$

If $k = N$ and $1 \le l \le N-1$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \quad & \frac{-1}{\sqrt{2}(N^3+(N-1)^2)^{1/2}} + \frac{1}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} \\ & \qquad + \frac{1}{(N^3+(N-1)^2)^{1/2}} \\ > \quad & 0. \end{aligned}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Suppose next that $1 \le j \le N-1$ and $i = N$. If $1 \le k, l \le N-1$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \; & \frac{-1}{\sqrt{3}(N^3 + (N-1)^2)^{1/2}} + \frac{1}{\sqrt{2}(N^3 + (N-1)^2)^{1/2}} \\ & + \frac{1}{\sqrt{3}(N^3 + (N-1)^2)^{1/2}} \\ > \; & 0. \end{aligned}$$

If $l = N$ and $1 \le k \le N-1$, then

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \; & \frac{-1}{\sqrt{2}(N^3 + (N-1)^2)^{1/2}} + \frac{1}{(N^3 + (N-1)^2)^{1/2}} \\ & + \frac{1}{\sqrt{3}(N^3 + (N-1)^2)^{1/2}} \\ > \; & 0. \end{aligned}$$

Lastly, suppose that $i = j = N$. Then $1 \le k, l \le N-1$ and

$$\begin{aligned} & \left\langle n^{ij}, \gamma^{kl} \right\rangle + \frac{1}{N-1}\left(\left\langle n^{ij}, \gamma^{il} \right\rangle + \left\langle n^{ij}, \gamma^{kj} \right\rangle\right) \\ = \; & \frac{-1}{\sqrt{3}((2N-2)N^2 + (2N-1)^2)^{1/2}} + \frac{1}{\sqrt{2}((2N-2)N^2 + (2N-1)^2)^{1/2}} \\ & + \frac{1}{\sqrt{2}((2N-2)N^2 + (2N-1)^2)^{1/2}} \\ > \; & 0. \end{aligned}$$

Combining all of the above, we see that $\sum_{kl \in \mathcal{I}(x)} \left\langle n^{ij}, \gamma^{kl} \right\rangle > 0$. Thus, since $\mathcal{I}(x)$ is finite, (62) follows.

We next establish (61). The argument is similar to that for (62), but it differs in some details as there is not exact symmetry between $\{n^{ij}\}$ and $\{\gamma^{kl}\}$. It suffices to show that

$$\sum_{kl \in \mathcal{I}(x)} \left\langle a_{kl} n^{kl}, \gamma^{ij} \right\rangle > 0 \text{ for each } ij \in \mathcal{I}(x), \tag{70}$$

where

$$a_{kl} = \begin{cases} 1 & \text{if } k \ne N,\ l \ne N, \\ N & \text{if } k = N,\ l \ne N, \\ N & \text{if } k \ne N,\ l = N, \\ 2N & \text{if } k = N,\ l = N, \end{cases}$$

because we can then set $b_{kl}(x) = \frac{a_{kl}}{H(x)}$ for $kl \in \mathcal{I}(x)$, where $H(x) = \sum_{kl\in\mathcal{I}(x)} a_{kl}$ and (61) holds. To show (70), fix $ij \in \mathcal{I}(x)$. Now,

$$\sum_{kl\in\mathcal{I}(x)} a_{kl}n^{kl} = \sum_{k\in\mathcal{I}\setminus\{i\}} \sum_{l\in\mathcal{I}\setminus\{j\}} \Big\{ a_{kl}n^{kl}1_{\{kl\in\mathcal{I}(x)\}} + \frac{1}{N-1}\left(a_{il}n^{il}1_{\{il\in\mathcal{I}(x)\}} + a_{kj}n^{kj}1_{\{kj\in\mathcal{I}(x)\}}\right)\Big\} + a_{ij}n^{ij}. \tag{71}$$

Note from (63) that $\langle a_{ij}n^{ij}, \gamma^{ij}\rangle > 0$. Fix $k \neq i$, $l \neq j$, and consider the term between the braces $\{\}$ in (71). If $kl \notin \mathcal{I}(x)$, then the inner product of γ^{ij} with the brace term is non-negative by (64)–(65). On the other hand, if $kl \in \mathcal{I}(x)$, then we know from (67) that $il, kj \in \mathcal{I}(x)$. We will show that, in this case, the inner product of γ^{ij} with the brace term is strictly positive. There are a number of cases to consider.

Suppose first that $1 \le i, j \le N-1$. If $1 \le k, l \le N-1$ also, then

$$\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle = \frac{-1}{\sqrt{3}(2N^2+1)^{1/2}},$$

$$\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle = \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}} \quad \text{and} \quad \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle = \frac{N-1}{\sqrt{3}(2N^2+1)^{1/2}}.$$

Hence it follows that

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ = \;& \frac{1}{\sqrt{3}(2N^2+1)^{1/2}} > 0. \end{aligned}$$

If $k = N$ and $1 \le l \le N-1$, then

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ = \;& \frac{-N}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} + \frac{1}{\sqrt{3}(2N^2+1)^{1/2}} + \frac{N}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} \\ > \;& 0. \end{aligned}$$

If $l = N$ and $1 \le k \le N-1$, then

$$\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right)$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

$$\begin{aligned} &= \frac{-N}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} + \frac{N}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} \\ &\quad + \frac{1}{\sqrt{3}(2N^2+1)^{1/2}} \\ &> 0. \end{aligned}$$

If $k = N$ and $l = N$, then

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ &= \frac{-2N}{\sqrt{3}((2N-2)N^2+(2N-1)^2)^{1/2}} + \frac{2N}{\sqrt{3}(N^3+(N-1)^2)^{1/2}} \\ &> 0. \end{aligned}$$

Suppose next that $1 \le i \le N-1$ and $j = N$. If $1 \le k, l \le N-1$, then

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ &= \frac{-1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{N}{\sqrt{2}(N^3+(N-1)^2)^{1/2}} \\ &> 0. \end{aligned}$$

If $k = N$ and $1 \le l \le N-1$, then

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ &= \frac{-N}{\sqrt{2}(N^3+(N-1)^2)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} \\ &\quad + \frac{2N}{\sqrt{2}((2N-2)N^2+(2N-1)^2)^{1/2}} \\ &> 0. \end{aligned}$$

The above display holds since the sum of the first term and the third term on the right-hand side of the equals sign is non-negative because $N \ge 2$.

Suppose next that $1 \le j \le N-1$ and $i = N$. If $1 \le k, l \le N-1$, then

$$\begin{aligned} &\left\langle a_{kl}n^{kl}, \gamma^{ij}\right\rangle + \frac{1}{N-1}\left(\left\langle a_{il}n^{il}, \gamma^{ij}\right\rangle + \left\langle a_{kj}n^{kj}, \gamma^{ij}\right\rangle\right) \\ &= \frac{-1}{\sqrt{2}(2N^2+1)^{1/2}} + \frac{N}{\sqrt{2}(N^3+(N-1)^2)^{1/2}} + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} \\ &> 0. \end{aligned}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

If $l = N$ and $1 \le k \le N-1$, then

$$\begin{aligned} & \left\langle a_{kl} n^{kl}, \gamma^{ij} \right\rangle + \frac{1}{N-1} \left(\left\langle a_{il} n^{il}, \gamma^{ij} \right\rangle + \left\langle a_{kj} n^{kj}, \gamma^{ij} \right\rangle \right) \\ = & \frac{-N}{\sqrt{2}(N^3 + (N-1)^2)^{1/2}} + \frac{2N}{\sqrt{2}((2N-2)N^2 + (2N-1)^2)^{1/2}} \\ & + \frac{1}{\sqrt{2}(2N^2+1)^{1/2}} \\ > & 0. \end{aligned}$$

Lastly, suppose that $i = j = N$. Then $1 \le k, l \le N-1$ and

$$\begin{aligned} & \left\langle a_{kl} n^{kl}, \gamma^{ij} \right\rangle + \frac{1}{N-1} \left(\left\langle a_{il} n^{il}, \gamma^{ij} \right\rangle + \left\langle a_{kj} n^{kj}, \gamma^{ij} \right\rangle \right) \\ = & \frac{-1}{(2N^2+1)^{1/2}} + \frac{2N}{(N^3 + (N-1)^2)^{1/2}} \\ > & 0. \end{aligned}$$

Combining all of the above, we obtain that $\sum_{kl \in \mathcal{I}(x)} \langle a_{kl} n^{kl}, \gamma^{ij} \rangle > 0$ for $ij \in \mathcal{I}(x)$ and (61) follows. □

COROLLARY 6.1. *Assumptions (A1)–(A5) of [13] are all satisfied by $\mathcal{W}$ and $\{\gamma^{ij} : i, j \in \mathcal{I}\}$, where $\mathcal{W} = \mathcal{G}$ given by (59) is a minimal description for $\mathcal{W}$.*

PROOF. In Lemma 6.1, we showed that $\mathcal{W}$ is a convex polyhedron with representation given by the intersection of the half-spaces $\mathcal{G}_{ij}$, $i, j \in \mathcal{I}$. We now prove by contradiction that such a representation is minimal in the sense that no proper subcollection defines $\mathcal{W}$. Suppose that there exist $i, j \in \mathcal{I}$ such that $\mathcal{W} = \bigcap_{(k,l) \neq (i,j)} \mathcal{G}_{kl}$. Let $\tilde{p} \in \mathbb{R}^{2N}$ be a vector such that

$$\tilde{p}_i = \tilde{p}_{N+j} = -1 \text{ and } \tilde{p}_k = \tilde{p}_{N+l} = 1 \text{ for } k \neq i \text{ and } l \neq j,$$

and let $\tilde{x} = AB'\tilde{p}$. By (60), we have that

$$\left\langle n^{kl}, \tilde{x} \right\rangle = K_{kl}(\tilde{p}_k + \tilde{p}_{N+l}) \ge 0 \text{ if } (k,l) \neq (i,j).$$

Then $\tilde{x} \in \mathcal{W}$. Hence there exists $p^* \in \mathbb{R}^{2N}_+$ such that $\tilde{x} = AB'p^*$. It follows that $AB'(p^* - \tilde{p}) = 0 \in \mathcal{W}$. By the second equality in (60), which holds for any $p \in \mathbb{R}^{2N}$, we obtain that

$$\left\langle n^{kl}, 0 \right\rangle = K_{kl}(p^*_k - \tilde{p}_k + p^*_{N+l} - \tilde{p}_{N+l}) = 0 \text{ for all } k, l \in \mathcal{I}.$$

In particular, $p_i^*+p_{N+j}^* = \tilde{p}_i+\tilde{p}_{N+j}$. Since $p_i^*+p_{N+j}^* \geq 0$ and $\tilde{p}_i+\tilde{p}_{N+j} = -2$, we have the contradiction, and this proves the minimality of the representation.

To see that $\mathcal{W}$ has non-empty interior, for $p^\sharp > 0$, let $x^\sharp = AB'p^\sharp$. Then it follows from (50) that $x^\sharp$ is in the interior of $\mathcal{W}$. As noted in Section 3 of [13], it follows that the conditions on the geometry of $\mathcal{W}$, Assumptions (A1)–(A3) of [13], are satisfied. Since the $\{\gamma^{ij} : i,j \in \mathcal{I}\}$ are constant vectors, they are trivially uniformly Lipschitz continuous, and since they are also of unit length, Assumption (A4) of [13] holds. Assumption (A5) of [13] is implied by Lemma 6.2. □

REMARK 6.1. As noted in [13], in the context of $\mathcal{W}$ being a convex polyhedron with minimal description and constant vector fields $\{\gamma^{ij} : i,j \in \mathcal{I}\}$ on the boundary faces, (A5) is equivalent to Assumption 5.1 of [13].

6.2. *Oscillation inequality.* The following oscillation inequality will be used in combination with the multiplicative state space collapse condition (Definition 4.4) to show that state space collapse (Theorem 6.2) holds. Also it is the key in verifying Assumption 4.1 of [13]. For the statement of the oscillation inequality, we need the following notation. For any $0 \leq s < t < \infty$ and any integer $k \in \mathbb{N}$, let $\mathbb{D}([s,t],\mathbb{R}^k)$ denote the set of functions $x : [s,t] \to \mathbb{R}^k$ that are right continuous on $[s,t)$ and have finite left limits on $(s,t]$, and for $x \in \mathbb{D}([s,t],\mathbb{R}^k)$, let

$$Osc(x,[s,t]) = \sup\{|x(u_2)-x(u_1)| : s \leq u_1 < u_2 \leq t\}. \tag{72}$$

LEMMA 6.3 (Oscillation Inequality). *There exists a constant $c_0 > 0$ such that for any $\delta > 0$ and any $0 \leq s < t < \infty$, w, $x \in \mathbb{D}([s,t],\mathbb{R}^{2N-1})$ and $y \in \mathbb{D}([s,t],\mathbb{R}^{N^2})$ satisfying*

(i) $w(u) = w(s) + x(u) - x(s) + \sum_{i,j\in\mathcal{I}}(y_{ij}(u) - y_{ij}(s))\gamma^{ij}$ *for all* $u \in [s,t]$,
(ii) $w(u) \in \mathcal{W}$ *for all* $u \in [s,t]$,
(iii) *for* $i,j \in \mathcal{I}$,
  (a) $y_{ij}(s) \geq 0$,
  (b) y_{ij} *is non-decreasing and* $y_{ij}(u) - y_{ij}(u-) \leq \delta$ *for all* $u \in (s,t]$,
  (c) $y_{ij}(u) = y_{ij}(s) + \int_{(s,u]} 1_{\{dist(w(v),\ \mathcal{B}_{ij})\leq\delta\}} dy_{ij}(v)$ *for all* $u \in [s,t]$,

the following hold:

$$Osc(w,[s,t]) \leq c_0(Osc(x,[s,t]) + \delta), \tag{73}$$

$$Osc(y,[s,t]) \leq c_0(Osc(x,[s,t]) + \delta). \tag{74}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

PROOF. A local version of this oscillation inequality is given in Theorem 4.1 of [13]. The main point of our argument below is to show that for the polyhedral cone $\mathcal{W}$ and constant directions of reflection $\{\gamma^{ij} : i,j \in \mathcal{I}\}$, the inequality actually holds globally. Since $\mathcal{W}$ and $\{\gamma^{ij} : i,j \in \mathcal{I}\}$ satisfy Assumptions (A1)–(A5) of [13], Theorem 4.1 of [13] holds with $\mathcal{W}$ and γ^{ij} in place of G and γ^i there. By the remark after Lemma A.3 and the construction of Π_m in the beginning of the proof of Theorem 4.1 of [13], we obtain that there is a constant $c_0 > 0$ such that $\Pi(u) \le c_0 u$ for each $u \ge 0$, where the function Π is constructed in the proof of Theorem 4.1 of [13]. By the fact that for each $i,j \in \mathcal{I}$, $\mathcal{B}_{ij}$ belongs to a hyperplane, we can choose $R(\cdot)$ in Assumption (A2) of [13] such that $R(\varepsilon) = \infty$ for each $\varepsilon \in (0,1)$. Since γ^{ij} is a constant vector for each $i,j \in \mathcal{I}$, then the Lipschitz constant L in Assumption (A4) of [13] can be arbitrarily small and hence, $\rho_0 = \frac{a}{4L}$ in [13] can be arbitrarily large. Then the quantity $\min\{\frac{\rho_0}{4}, \frac{R(a/4)}{4}\}$ in Theorem 4.1 of [13], which restricts the size of the neighborhood of a point in $\mathcal{W}$ in which the oscillation inequality holds, can be arbitrarily large. It follows that the oscillation inequality in Theorem 4.1 of [13] holds globally for paths w in $\mathcal{W}$ with $\Pi(u) \le c_0 u$ for all $u \ge 0$. □

6.3. *State space collapse.* We first state and prove the following two preliminary lemmas.

LEMMA 6.4. *$\{\widehat{W}^r(0) + \widehat{X}^r(\cdot) : r > 0\}$ converges in distribution to a $(2N-1)$-dimensional Brownian motion with initial distribution μ, drift θ and covariance matrix Γ.*

PROOF. The lemma follows directly from (30), the assumptions of Theorem 5.1, (21), Proposition 3.1, and the fact that $\left|rt - \frac{\lfloor r^2 t\rfloor}{r}\right| \le \frac{1}{r}$. □

LEMMA 6.5. *There exist two constants $c_1 > 0$, $c_2 > 1$ and $\overline{r} > 0$ such that for each $T > 0$,*

$$c_1 \left|\left|\widehat{W}^r(\cdot)\right|\right|_T \le \left|\left|\widehat{Q}^r(\cdot)\right|\right|_T \le c_2 \left|\left|\widehat{W}^r(\cdot)\right|\right|_T \quad \text{for all } r > \overline{r}. \tag{75}$$

PROOF. Since $\widehat{W}^r(\cdot) = A\widehat{Q}^r(\cdot)$, then for each $T > 0$, $||\widehat{W}^r(\cdot)||_T \le |A|\,||\widehat{Q}^r(\cdot)||_T$. By letting $c_1 = |A|^{-1}$, we see that the left hand inequality of (75) holds. On the other hand, for $i,j \in \mathcal{I}$ and $t \in \mathbb{R}_+$, $Q^r_{ij}(t)$ is the $((i-1)N + j)^{th}$ entry of $Q^r(t)$ and $Q^r_{ij}(t) \le W^r_{2N-1}(t)$. By letting $c_2 = N$, we obtain the right hand inequality of (75). □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

We are now ready to prove the main lemma used in proving the state space collapse. By (25), we obtain that for each $t \in \mathbb{R}_+$,

$$\widehat{W}^r(t) = \widetilde{W}^r(t) + \widehat{\xi}^r(t), \tag{76}$$

where

$$\widetilde{W}^r(t) = A\Delta(\widehat{W}^r(t)), \tag{77}$$

$$\widehat{\xi}^r(t) = A\left(\widehat{Q}^r(t) - \Delta\left(\widehat{W}^r(t)\right)\right). \tag{78}$$

By (41), we have that

$$\widetilde{W}^r(t) \in \mathcal{W} \text{ for all } t \in \mathbb{R}_+ \text{ and } r > 0. \tag{79}$$

Recall from (28)–(29) that for each $r > 0$ and $t \in \mathbb{R}_+$,

$$\widehat{W}^r(t) = \widehat{W}^r(0) + \widehat{X}^r(t) + \sum_{i,j\in\mathcal{I}} \widehat{Y}^r_{ij}(t)\gamma^{ij}, \tag{80}$$

where for each $\delta > 0$ fixed and each $r > 0,\ t \in \mathbb{R}_+,\ i, j \in \mathcal{I}$,

$$\widehat{Y}^r_{ij}(t) = \int_{(0,t]} 1_{\{\mathrm{dist}(\widetilde{W}^r(s),\mathcal{B}_{ij})\le\delta\}} d\widehat{Y}^r_{ij}(s) + \widehat{\zeta}^{r,\delta}_{ij}(t), \tag{81}$$

and

$$\widehat{\zeta}^{r,\delta}_{ij}(t) = \int_{(0,t]} 1_{\{\mathrm{dist}(\widetilde{W}^r(s),\mathcal{B}_{ij})>\delta\}} d\widehat{Y}^r_{ij}(s). \tag{82}$$

We then have the following estimate.

LEMMA 6.6. *For each $T > 0$, $\delta > 0$, there exists $r(T,\delta) > 0$ such that for all $r \ge r(T,\delta)$,*

$$P\left(\left|\left|\widehat{Q}^r(\cdot) - \Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_T \le \delta,\ \left|\left|\widehat{\xi}^r(\cdot)\right|\right|_T \le \delta,\ \left|\left|\widehat{\zeta}^{r,\delta}(\cdot)\right|\right|_T = 0\right) > 1 - \delta. \tag{83}$$

PROOF. Fix $T > 0$ and $\delta > 0$. By the convergence assumed for the initial random variables $\{\widehat{W}^r(0) :\ r > 0\}$, the convergence in distribution of $\{\widehat{W}^r(0) + \widehat{X}^r(\cdot) :\ r > 0\}$ established in Lemma 6.4, multiplicative state space collapse and the convergence in distribution of $\widehat{I}^r$ to the Brownian motion $\widetilde{I}$ in (33), we have that, for each $\varepsilon > 0$, there are constants $K_0 \ge 1$ (not depending on ε) and $r_0(\varepsilon) > 0$ such that for all $r \ge r_0(\varepsilon)$,

$$P\left(|\widehat{W}^r(0)| \le K_0,\ ||\widehat{X}^r(\cdot)||_T \le K_0\right) \ge 1 - \frac{\delta}{3}, \tag{84}$$

$$P\left(\sup_{0<t\le T}|\widehat{I}^r(t)-\widehat{I}^r(t-)|\le\frac{\delta}{8|A|}\right)\ge 1-\frac{\delta}{3}, \tag{85}$$

$$P\left(\left|\left|\widehat{Q}^r(\cdot)-\Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_T\le\varepsilon(||\widehat{Q}^r(\cdot)||_T\vee 1)\right)\ge 1-\frac{\delta}{3}. \tag{86}$$

The constants K_0 and $r_0(\varepsilon)$ depend on T and δ as well, but since these parameters are fixed throughout this proof, we do not explicitly indicate that dependence here. In the following, $\varepsilon>0$ will be fixed. A specific, suitably small value of ε will be chosen later (as a function of T and δ) to ensure that various inequalities hold. For $r>0$, let

$$O^{r,\varepsilon}=\left\{\begin{array}{c}\left|\widehat{W}^r(0)\right|\le K_0,\ \left|\left|\widehat{Q}^r(\cdot)-\Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_T\le\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot)\right|\right|_T\vee 1\right)\\ \left|\left|\widehat{X}^r(\cdot)\right|\right|_T\le K_0,\ \sup_{0<t\le T}|\widehat{I}^r(t)-\widehat{I}^r(t-)|\le\frac{\delta}{8|A|}\end{array}\right\}.$$

From (84)–(86) we have that for all $r\ge r_0(\varepsilon)$,

$$P(O^{r,\varepsilon})\ge 1-\delta. \tag{87}$$

Now, for $r\ge r_0(\varepsilon)$, on $O^{r,\varepsilon}$, by (78) we have

$$\left|\left|\widehat{\xi}^r(\cdot)\right|\right|_T\le|A|\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot)\right|\right|_T\vee 1\right). \tag{88}$$

We now focus on when $\widehat{Y}^r$ can increase. Fix $r\ge r_0(\varepsilon)$, $i,j\in\mathcal{I}$, and $\omega\in O^{r,\varepsilon}$. Recall the constraints on where $\widehat{Y}^r_{ij}$ can increase from (28). Fix a time instant $t^*\in(0,T]$ such that $\widehat{Q}^r_{ij}(t^*-,\omega)=0$. Since $\omega\in O^{r,\varepsilon}$, we have

$$\Delta\left(\widehat{W}^r(t^*-,\omega)\right)_{ij}\le\left|\left|\widehat{Q}^r(\cdot,\omega)-\Delta\left(\widehat{W}^r(\cdot,\omega)\right)\right|\right|_T\le\epsilon\left(\left|\left|\widehat{Q}^r(\cdot,\omega)\right|\right|_T\vee 1\right). \tag{89}$$

By (41), there is $p^r(t^*,\omega)\in\mathbb{R}^{2N}_+$ such that $\Delta\left(\widehat{W}^r(t^*-,\omega)\right)=B'p^r(t^*,\omega)$. Then it follows from the definition of B and (89) that

$$p^r_i(t^*,\omega)+p^r_{N+j}(t^*,\omega)\le\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot,\omega)\right|\right|_T\vee 1\right). \tag{90}$$

By the definition (77) of $\widetilde{W}^r$, we have

$$\widetilde{W}^r(t^*-,\omega)=AB'p^r(t^*,\omega). \tag{91}$$

It then follows from (60), (90) and (91) that

$$\mathrm{dist}(\widetilde{W}^r(t^*-,\omega),\mathcal{B}_{ij}) \;=\; \left\langle n^{ij},\widetilde{W}^r(t^*-,\omega)\right\rangle$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

$$\begin{aligned} &= \langle n^{ij}, AB' p^r(t^*,\omega)\rangle \\ &= K_{ij}(p_i^r(t^*,\omega) + p_{N+j}^r(t^*,\omega)) \\ &\le K_{ij}\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot,\omega)\right|\right|_T \vee 1\right). \end{aligned}$$

Notice that by (5) and (8), the jump of Q_{ij}^r at time t^* is bounded in size by the size of the jump of I_{ij}^r at time t^* plus one (bounding a possible jump of the departure process). Thus, for $r \ge 8|A|\sqrt{N}/\delta$,

$$\begin{aligned} |\widehat{Q}^r(t^*,\omega) - \widehat{Q}^r(t^*-,\omega)| &\le |\widehat{I}^r(t^*,\omega) - \widehat{I}^r(t^*-,\omega)| + \frac{\sqrt{N}}{r} \\ &\le \sup_{0<t\le T} |\widehat{I}^r(t,\omega) - \widehat{I}^r(t-,\omega)| + \frac{\sqrt{N}}{r} \le \frac{\delta}{4|A|}. \end{aligned}$$

Hence $|\widehat{W}^r(t^*,\omega) - \widehat{W}^r(t^*-,\omega)| \le \delta/4$. It then follows that

$$\begin{aligned} &\left|\widetilde{W}^r(t^*,\omega) - \widetilde{W}^r(t^*-,\omega)\right| \\ \le\ &\left|\widetilde{W}^r(t^*,\omega) - \widehat{W}^r(t^*,\omega)\right| + \left|\widehat{W}^r(t^*,\omega) - \widehat{W}^r(t^*-,\omega)\right| \\ &\quad + \left|\widehat{W}^r(t^*-,\omega) - \widetilde{W}^r(t^*-,\omega)\right| \\ \le\ &2\left|\left|\widehat{\xi}^r(\cdot)\right|\right|_T + \delta/4 \\ \le\ &2|A|\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot)\right|\right|_T \vee 1\right) + \delta/4. \end{aligned}$$

Let $K_1 = \sup\{K_{ij} :\ i,j \in \mathcal{I}\}$. Then we have

$$\operatorname{dist}(\widetilde{W}^r(t^*,\omega), \mathcal{B}_{ij}) \le (2|A| + K_1)\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot,\omega)\right|\right|_T \vee 1\right) + \delta/4. \tag{92}$$

Since $t^* \in (0,T]$ was an arbitrary point at which $\widehat{Q}_{ij}^r(t^*-) = 0$, it then follows from (28) that

$$\widehat{Y}_{ij}^r(t) = \int_{(0,t]} 1_{\{\operatorname{dist}(\widetilde{W}^r(s,\,\omega),\,\mathcal{B}_{ij}) \le (2|A|+K_1)\varepsilon\left(\left|\left|\widehat{Q}^r(\cdot,\,\omega)\right|\right|_T \vee 1\right) + \delta/4\}}\, d\widehat{Y}_{ij}^r(s), \tag{93}$$

for all $t \in (0,T]$. Furthermore, by (76) and (29), for all $t \in (0,T]$,

$$\begin{aligned} \widetilde{W}^r(t,\omega) &= \widehat{W}^r(t,\omega) - \widehat{\xi}^r(t,\omega) \\ &= \widehat{W}^r(0,\omega) + \widehat{X}^r(t,\omega) - \widehat{\xi}^r(t,\omega) + \sum_{i,j\in\mathcal{I}} \widehat{Y}_{ij}^r(t,\omega)\gamma^{ij}, \end{aligned}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

where $\widetilde{W}^r(t,\omega) \in \mathcal{W}$. Let $r_1(\varepsilon) = \max\{r_0(\varepsilon), 8|A|\sqrt{N}/\delta\}$ and consider $r \geq r_1(\varepsilon)$. By (9), U^r_{ij} can jump by at most one at any time and it follows from (27) that $\widehat{Y}^r_{ij}$ can jump by at most $|A_{ij}|/r \leq \delta/8$ at any time. It then follows from the oscillation inequality in Lemma 6.3 (with δ there replaced by $(2|A| + K_1)\varepsilon\big(\big|\big|\widehat{Q}^r(\cdot)\big|\big|_T \vee 1\big) + \delta/4$), and from (88), Lemma 6.5, and the fact that $c_2 > 1$, that on $O^{r,\varepsilon}$,

$$\begin{aligned}
& Osc(\widetilde{W}^r, [0,T]) \\
\leq\ & c_0 \left(Osc(\widehat{W}^r(0) + \widehat{X}^r - \widehat{\xi}^r, [0,T]) + (2|A| + K_1)\varepsilon \left(\left|\left| \widehat{Q}^r(\cdot) \right|\right|_T \vee 1 \right) + \delta/4 \right) \\
\leq\ & c_0 \left(2K_0 + (4|A| + K_1)\varepsilon \left(\left|\left| \widehat{Q}^r(\cdot) \right|\right|_T \vee 1 \right) + \delta/4 \right) \\
\leq\ & c_0 \left(2K_0 + (4|A| + K_1)\varepsilon c_2 \left(\left|\left| \widehat{W}^r(\cdot) \right|\right|_T \vee 1 \right) + \delta/4 \right).
\end{aligned}$$

Then it follows that on $O^{r,\varepsilon}$,

$$\begin{aligned}
& \left|\left| \widehat{W}^r(\cdot) \right|\right|_T \vee 1 \\
\leq\ & \left| \widehat{W}^r(0) \right| \vee 1 + Osc(\widehat{W}^r, [0,T]) \\
\leq\ & \left| \widehat{W}^r(0) \right| \vee 1 + Osc(\widetilde{W}^r, [0,T]) + Osc(\widehat{\xi}^r, [0,T]) \\
\leq\ & K_0 + (c_0 + 1) \left(2K_0 + (4|A| + K_1)\varepsilon c_2 \left(\left|\left| \widehat{W}^r(\cdot) \right|\right|_T \vee 1 \right) + \delta/4 \right).
\end{aligned} \tag{94}$$

Now, choose

$$\varepsilon = \min \left\{ \frac{1}{2(c_0+1)c_2(4|A| + K_1)}, \frac{\delta}{4(2|A| + K_1 + 1)c_2(2(2c_0+3)K_0 + (c_0+1)\delta/2)} \right\}.$$

Then we obtain from (94) that for $r \geq r_1(\varepsilon)$, on $O^{r,\varepsilon}$,

$$\left|\left| \widehat{W}^r(\cdot) \right|\right|_T \vee 1 \leq 2(2c_0 + 3)K_0 + (c_0 + 1)\delta/2, \tag{95}$$

and then, by Lemma 6.5,

$$\begin{aligned}
\varepsilon \left(\left|\left| \widehat{Q}^r(\cdot) \right|\right|_T \vee 1 \right) &\leq \varepsilon c_2 \left(\left|\left| \widehat{W}^r(\cdot) \right|\right|_T \vee 1 \right) \\
&\leq \varepsilon c_2 (2(2c_0 + 3)K_0 + (c_0 + 1)\delta/2) \\
&\leq \frac{\delta}{4(2|A| + K_1 + 1)}.
\end{aligned} \tag{96}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

It follows from this and (88) that for $r \geq r_1(\varepsilon)$, on $O^{r,\varepsilon}$,

$$\left|\left|\widehat{Q}^r(\cdot) - \Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_T \leq \delta/4 < \delta, \qquad \left|\left|\widehat{\xi}^r(\cdot)\right|\right|_T \leq \delta/8 < \delta.$$

Finally, on combining (96) with (93), we obtain that for $r \geq r_1(\varepsilon)$, on $O^{r,\varepsilon}$,

$$\widehat{Y}^r_{ij}(t) = \int_{(0,t]} 1_{\{\text{dist}(\widetilde{W}^r(s,\,\omega),\,\mathcal{B}_{ij}) \leq \delta/2\}}\, d\widehat{Y}^r_{ij}(s),$$

for all $t \in (0, T]$. Then by the definition of $\widehat{\zeta}^r_{ij}$, we have that $||\widehat{\zeta}^r_{ij}(\cdot)||_T = 0$ on $O^{r,\varepsilon}$ for all $r \geq r_1(\varepsilon)$. Let $r(T, \delta) = r_1(\varepsilon)$, then we have the desired result. □

The following is immediate from Lemma 6.6.

COROLLARY 6.2. *Under the assumptions of Theorem 5.1, state space collapse holds.*

REMARK 6.2. In Theorem 7.7 of [15], Shah et al. use an alternative method to give a proof that multiplicative state space collapse implies state space collapse under some conditions. Their method utilizes an a priori probabilistic bound on Q^r that they obtain using a Lyapunov drift technique. Indeed, their result applies to input-queued switches operating under MW-α policies for $\alpha \geq 1$. Although we focus on the case $\alpha = 1$ here, our methodology in fact extends to allow a proof for all $\alpha \in (0, \infty)$. Furthermore, Shah et al. require that the heavy traffic limit be reached through a sequence of strictly underloaded systems, whereas our result allows for the limit to be approached through critically loaded or even overloaded systems; in other words, the Shah et al. result assumes θ in our Assumption 3.1 has all components strictly negative, whereas we do not restrict the sign of the components of θ at all. In summary, we have included our proof here because it allows a more flexible heavy traffic assumption and our methodology readily extends to allow a proof of Corollary 6.2 for MW-α policies for all $\alpha \in (0, \infty)$.

6.4. *Proof of Theorem 5.1.* Recall that we are assuming that the hypotheses of Theorem 5.1 hold. It suffices to show that the conditions of Theorem 5.4 of [13] hold, from which it will follow that $\widehat{W}^r$ converges in distribution as $r \to \infty$ to an SRBM associated with the data $(\mathcal{W}, \{\gamma^{ij} : i, j \in \mathcal{I}\}, \theta, \Gamma, \mu)$. The joint convergence of $\widehat{Q}^r$ with $\widehat{W}^r$ and $\widetilde{Q} = \Delta(\widetilde{W})$ will then follow by the state space collapse of Corollary 6.2 and the continuity of Δ established in Proposition 4.1. The conditions of Theorem 5.4 of [13] fall into four groups. We treat each of these groups separately below.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Firstly, as verified in Corollary 6.1, the workload space $\mathcal{W}$ is a convex polyhedron having non-empty interior described as the intersection of a minimal set of half-spaces, and the directions of reflection $\{\gamma^{ij},\ i,j\in\mathcal{I}\}$ satisfy Assumption 5.1 of [13].

Secondly, we verify that Assumption 4.1 of [13] holds. Recall the existence of $r(T,\delta)$ from Lemma 6.6. Let $\{r_k:\ k\in\mathbb{N}\}$ be a strictly increasing sequence of positive constants such that for each $k\in\mathbb{N}$, $r_k\geq r(k,\frac{1}{k})$ and $r_k\to\infty$ as $k\to\infty$. Define δ^r such that $\delta^r=1$ when $r\leq r_1$ and $\delta^r=\frac{1}{k}$ when $r_k<r\leq r_{k+1}$ for $k\in\mathbb{N}$. Then $\delta^r\to 0$ as $r\to\infty$ and by Lemma 6.6, for each $k\in\mathbb{N}$ and $r_k<r\leq r_{k+1}$,

(97)
$$P\left(\left|\left|\widehat{Q}^r(\cdot)-\Delta\left(\widehat{W}^r(\cdot)\right)\right|\right|_k\leq\frac{1}{k},\ \left|\left|\widehat{\xi}^r(\cdot)\right|\right|_k\leq\frac{1}{k},\ \left|\left|\widehat{\zeta}^{r,\delta^r}(\cdot)\right|\right|_k=0\right)>1-\frac{1}{k}.$$

If for each $i,j\in\mathcal{I}$ and $t\in\mathbb{R}_+$, we define

$$\widehat{V}^r_{ij}(t)=\int_{(0,t]}1_{\{\mathrm{dist}(\widetilde{W}^r(s),\mathcal{B}_{ij})\leq\delta^r\}}d\widehat{Y}^r_{ij}(s), \tag{98}$$

then from this, (29), (76)–(82), we have that for each $r>0$ and $t\in\mathbb{R}_+$,

$$\widehat{W}^r(t)=\widehat{W}^r(0)+\widehat{X}^r(t)+\sum_{i,j\in\mathcal{I}}\widehat{Y}^r_{ij}(t)\gamma^{ij}, \tag{99}$$

$$\widehat{W}^r(t)=\widetilde{W}^r(t)+\widehat{\xi}^r(t)\quad\text{and}\quad\widetilde{W}^r(t)\in\mathcal{W}, \tag{100}$$

$$\widehat{Y}^r_{ij}(t)=\widehat{V}^r_{ij}(t)+\widehat{\zeta}^{r,\delta^r}_{ij}(t), \tag{101}$$

and

$$\widehat{V}^r_{ij}(t)=\int_{(0,t]}1_{\{\mathrm{dist}(\widetilde{W}^r(s),\mathcal{B}_{ij})\leq\delta^r\}}d\widehat{V}^r_{ij}(s). \tag{102}$$

LEMMA 6.7. *The process $\widehat{\xi}^r(\cdot)$ and the processes $\widehat{\zeta}^{r,\delta^r}_{ij}(\cdot)$, $i,j\in\mathcal{I}$, all converge in probability to zero processes as $r\to\infty$.*

PROOF. For $\widehat{\xi}^r(\cdot)$, for any $T>0$, it follows from (97) that for each $k>T$ and any $r>r_k$, we have

$$P\left\{||\widehat{\xi}^r(\cdot)||_T>\frac{1}{k}\right\}\leq P\left\{||\widehat{\xi}^r(\cdot)||_k>\frac{1}{k}\right\}\leq\frac{1}{k}. \tag{103}$$

It follows that $\widehat{\xi}^r(\cdot)$ converges in probability to the zero process as $r\to\infty$.

Fix $i, j \in \mathcal{I}$. By (97), we know that for each $k > T$ and any $r > r_k$,

$$P\left\{||\widehat{\zeta}_{ij}^{r,\delta^r}(\cdot)||_T > 0\right\} \leq \frac{1}{k}. \tag{104}$$

It follows that the non-decreasing process $\widehat{\zeta}_{ij}^{r,\delta^r}(\cdot)$ converges in probability to the zero process as $r \to \infty$. □

By Lemma 6.4, $\{\widehat{W}^r(0) + \widehat{X}^r(\cdot) : r > 0\}$ is C-tight. Combining the above, it follows that the conditions of Assumption 4.1 of [13] are satisfied with $\mathcal{W}$ in place of $\overline{G}$, r in place of n, δ^r in place of δ^n, ij in place of i, γ^{ij} in place of $\gamma^{i,n}$, $\widehat{W}^r$ in place of W^n, $\widetilde{W}^r$ in place of $\widetilde{W}^n$, $\widehat{W}^r(0) + \widehat{X}^r$ in place of X^n, $\widehat{Y}_{ij}^r$ in place of Y_i^n, $\widehat{V}_{ij}^r$ in place of $\tilde{Y}_i^n$, $\widehat{\xi}^r$ in place of α^n, and $\widehat{\zeta}^{r,\delta^r}$ in place of β^n.

Thirdly, we have the conclusion of Lemma 6.4.

Fourthly, and finally, we must verify condition (vii) of Theorem 4.3 of [13] (with θ in place of μ there). This condition requires that for any weak limit point $(\widetilde{W}, \widetilde{X}, \widetilde{Y})$ of $\{(\widehat{W}^r, \widehat{X}^r, \widehat{Y}^r)\}$, $\{\widetilde{X}(t) - \theta t : t \in \mathbb{R}_+\}$ is a martingale relative to the filtration generated by $(\widetilde{W}, \widetilde{X}, \widetilde{Y})$. (This condition is needed so that property (iii)(b) of Definition 5.1 will be satisfied). By Proposition 4.1 of [13], it suffices to prove the following lemma which implies that (vii) of Theorem 4.3 of [13] holds.

LEMMA 6.8. *The process $\widehat{X}^r$ as given by (30) has the decomposition:*

$$\widehat{X}^r(\cdot) = \check{X}^r(\cdot) + \check{\epsilon}^r(\cdot), \tag{105}$$

where $\check{X}^r$, $\check{\epsilon}^r$ are $(2N-1)$-dimensional processes satisfying the following conditions.

(i) *For each r, $\{\check{X}^r(t) - r(A\lambda^r - v)t : t \in \mathbb{R}_+\}$ is a martingale with respect to the filtration generated by $(\widehat{W}^r, \check{X}^r, \widehat{Y}^r)$.*

(ii) *For each $T \in \mathbb{R}_+$, $||\check{\epsilon}^r(\cdot)||_T \to 0$ in probability as $r \to \infty$.*

(iii) *For each $t \in \mathbb{R}_+$, the collection $\{\check{X}^r(t) : r > 0\}$ is uniformly integrable.*

PROOF. We first introduce some martingales that will be used in defining $\check{X}^r$. Let $(\Omega^r, \mathcal{F}^r)$ be the measurable space on which all of the processes indexed by r are defined. Let $\{\mathcal{H}_t^r : t \in \mathbb{R}_+\}$ be the filtration defined by

$$\mathcal{H}_t^r = \sigma\{I^r(s), Q^r(0) : 0 \leq s \leq t\}, \qquad t \in \mathbb{R}_+.$$

We first show that W^r and U^r are adapted to the filtration $\{\mathcal{H}_t^r : t \in \mathbb{R}_+\}$. From the definitions, it is easy to see that $Q^r(0), I^r(0), D^r(0) \in \mathcal{H}_0^r$ and $Q^r(0), I^r(1) \in \mathcal{H}_1^r$. Since the MWM scheduling policy is being used,

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

$D^r(1)$ depends only on $Q^r(0)$, thus $D^r(1) \in \mathcal{H}_1^r$. Hence by (2), we have that $Q^r(1) \in \mathcal{H}_1^r$. Using an induction argument, we can see that for each $n \in \mathbb{N}$, $Q^r(n) \in \mathcal{H}_n^r$ and $D^r(n) \in \mathcal{H}_n^r$. By (5), (7) and (9), $U_{ij}^r(n) + D_{ij}^r(n) = n$ for each $i, j \in \mathcal{I}$, and so $U^r(n) \in \mathcal{H}_n^r$ for each $n \in \mathbb{N}_0$. It then follows from (11) and (19) that $W^r(t), Q^r(t), U^r(t) \in \mathcal{H}_t^r$ for each $t \in \mathbb{R}_+$.

We now introduce the fundamental martingales $\mathcal{K}^r$ and $\mathcal{L}^r$. For each $t \in \mathbb{R}_+$ and $i, j \in \mathcal{I}$, let

$$\mathcal{K}_{ij}^r(t) = I_{ij}^r(t) - \lambda_{ij}^r \lfloor t \rfloor, \tag{106}$$

$$\mathcal{L}_{ij}^r(t) = (\mathcal{K}_{ij}^r(t))^2 - b_{ij}^r \lfloor t \rfloor. \tag{107}$$

Let $\mathcal{K}^r(t)$ and $\mathcal{L}^r(t)$ be the vectors whose $((N-1)i+j)^{th}$ entries are $\mathcal{K}_{ij}^r(t)$ and $\mathcal{L}_{ij}^r(t)$, respectively, for $i, j \in \mathcal{I}$. Because of the i.i.d. and integrability assumptions on the $\{\vartheta_{ij}^r(k) : k \in \mathbb{N}\}$, we have that the $2N^2$-dimensional process

$$\{(\mathcal{K}^r(n), \mathcal{L}^r(n)) : n \in \mathbb{N}_0\}$$

is a discrete time martingale relative to $\{\mathcal{H}_n^r : n \in \mathbb{N}_0\}$. Thus, it follows from (19) that

$$\{(\mathcal{K}^r(t), \mathcal{L}^r(t)), \mathcal{H}_t^r, t \in \mathbb{R}_+\} \text{ is a martingale.} \tag{108}$$

Now we are ready to prove the lemma. For each $r > 0$, $t \in \mathbb{R}_+$ and $i, j \in \mathcal{I}$, note that

$$\widehat{I}_{ij}^r(t) = r^{-1}\left(I_{ij}^r(r^2 t) - r^2 \lambda_{ij}^r t\right) = r^{-1}\mathcal{K}_{ij}^r(r^2 t) - \lambda_{ij}^r(rt - \lfloor r^2 t \rfloor / r).$$

Thus, we have using (30) that

$$\widehat{X}^r(\cdot) = \check{X}^r(\cdot) + \check{\epsilon}^r(\cdot),$$

where

$$\check{X}^r(t) = r^{-1} A \mathcal{K}^r(r^2 t) + r(A\lambda^r - v)t, \tag{109}$$

$$\check{\epsilon}^r(t) = (v - A\lambda^r)\left(rt - \frac{\lfloor r^2 t \rfloor}{r}\right). \tag{110}$$

Since for each r, $\{(\mathcal{K}^r(r^2 t), \mathcal{L}^r(r^2 t)), \mathcal{H}_{r^2 t}^r, t \in \mathbb{R}_+\}$ is a martingale and (using (27)), $\widehat{W}^r, \check{X}^r, \widehat{Y}^r$ are adapted to the filtration $\{\mathcal{H}_{r^2 t}^r : t \in \mathbb{R}_+\}$, then condition (i) of Lemma 6.8 holds. Condition (ii) follows from (20) and the fact that $rt - \frac{\lfloor r^2 t \rfloor}{r} \to 0$ uniformly in t as $r \to \infty$. For condition (iii), by taking expectations on both sides of (107), we have $r^{-2} E[(\mathcal{K}_{ij}^r(r^2 t))^2] \le b_{ij}^r t$. Together with the assumed convergence of b_{ij}^r for each $i, j \in \mathcal{I}$ and the convergence of $r(A\lambda^r - v)t$ as $r \to \infty$, we obtain the L^2-boundedness of $\{\check{X}^r(t), r > 0\}$ for each fixed $t \in \mathbb{R}_+$. Hence condition (iii) holds. □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Thus, we have verified all of the hypotheses of Theorem 5.4 of [13] and the desired result, Theorem 5.1, follows.

7. Discussion of other policies. In this section we elaborate on some possible directions for further research. Shah and Wischik [19] studied switched networks under some generalizations of the MWM policy considered in this paper. In particular, they established a multiplicative state space collapse result (Theorem 7.1 of [19]) for input-queued switches operating under a MW-α policy for $\alpha \in (0, \infty)$, where the MW-α policy chooses the matching π in the time interval $[n, n+1)$ that maximizes the weight

$$\omega_\pi(Q(n)^\alpha) = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij} (Q_{ij}(n))^\alpha. \tag{111}$$

(The MWM policy considered in this paper corresponds to $\alpha = 1$). The set of invariant states for input-queued switches operating under a MW-α policy can then be characterized using the unique solution $\Delta^\alpha(w)$ for $w \in \mathbb{R}_+^{2N-1}$ to the following optimization problem:

$$\begin{aligned} &\text{minimize} && \sum_{i,j \in \mathcal{I}} x_{ij}^{1+\alpha} \\ &\text{subject to} && Bx \geq Cw, \\ &\text{over} && x \in \mathbb{R}_+^{N^2}. \end{aligned} \tag{112}$$

Via Lagrange multipliers, $\Delta^\alpha(w)$ can be expressed in the form:

$$\Delta^\alpha(w) = (B'p)^{1/\alpha}, \quad p \in \mathbb{R}_+^{2N}, \quad \langle p, (B\Delta(w) - Cw) \rangle = 0. \tag{113}$$

This leads to a natural conjecture for a diffusion approximation to the workload process for a heavily loaded input-queued switch operating under a MW-α policy. This proposed diffusion lives in the cone:

$$\mathcal{W}_\alpha = \left\{ A(B'p)^{1/\alpha} : \ p \in \mathbb{R}_+^{2N} \right\}. \tag{114}$$

Proving a rigorous heavy traffic limit theorem justifying such a diffusion approximation for $\alpha \neq 1$ is a natural research problem. Here we illustrate some of the challenges associated with proving such a result. These revolve around the fact that when $\alpha \neq 1$, although the state space for the proposed diffusion approximation for the workload is a cone, it is not a polyhedral cone. In fact, it has piecewise smooth curved boundary faces. The complexity of the geometry as N increases, and the current lack of a general existence

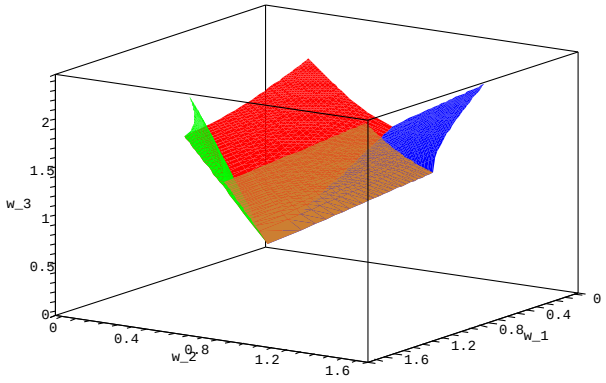


FIG 4. *A portion of the workload cone $\mathcal{W}_2$ is shown for a 2×2 input-queued switch with $\alpha = 2$.*

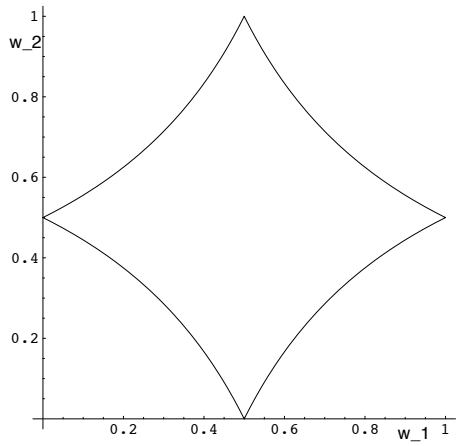


FIG 5. *A cross-section of the workload cone $\mathcal{W}_2$ depicted in Figure 4 taken at $w_3 = 1$.*

and uniqueness theory (and an associated invariance principle) for reflecting Brownian motions in such domains, are major obstacles to proving the conjecture for $\alpha \neq 1$. To illustrate the issues more concretely, for a 2×2 input-queued switch ($N = 2$), we depict the associated workload cone and a cross-section, for $\alpha = 2$ and $\alpha = 0.5$ in Figures 4–5 and 6–7, respectively.

For the case $\alpha = 2$, depicted in Figures 4–5 (which is representative of the case $\alpha > 1$), the workload cone has boundary faces that curve inwards. Away

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

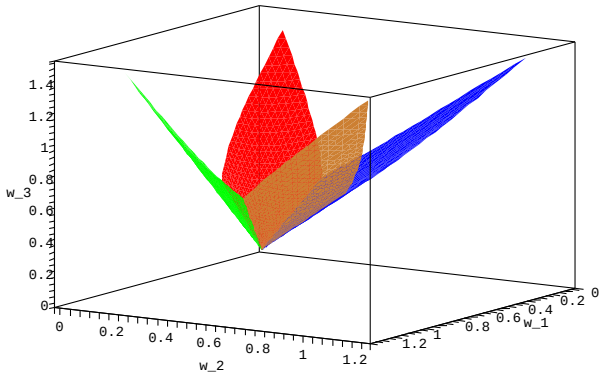


FIG 6. *A portion of the workload cone* $\mathcal{W}_{0.5}$ *is shown for a* 2×2 *input-queued switch with* $\alpha = 0.5$.

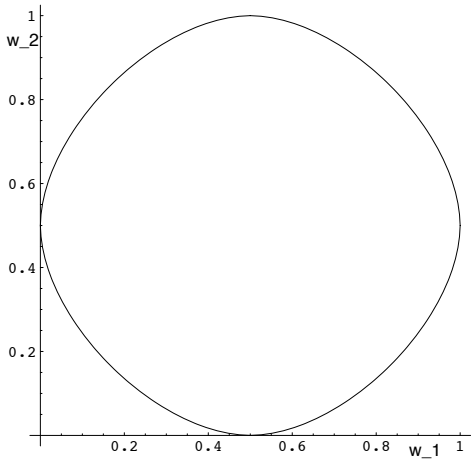


FIG 7. *A cross-section of the workload cone* $\mathcal{W}_{0.5}$ *depicted in Figure 6 taken at* $w_3 = 1$.

from the origin, the boundary and directions of reflection for the proposed diffusion approximation locally satisfy conditions required by the invariance principle given in [13]. However, the workload cone has a "singular point" at the origin where the conditions in [13] fail to be satisfied. However, since this is an isolated point, we believe that the invariance principle in [13] and the uniqueness result of [8] can be adapted to validate the diffusion ap-

proximation in this case. For higher dimensional analogues of this case, we anticipate that a valid diffusion approximation can be established. However, as N increases, it becomes more difficult to compute the inward normals to all boundary faces, and as yet we do not have a systematic way to characterize these geometric conditions.

For the case $\alpha = 0.5$, depicted in Figures 6–7 (which is representative of the case $\alpha \in (0, 1)$), the workload cone is convex and has boundary faces that curve outwards. In fact, in this case, the workload cone has a $\mathbb{C}^1$ boundary except at the origin and the direction of reflection is piecewise constant on the boundary. There is not yet an existence and uniqueness theory (nor an associated invariance principle) for the proposed diffusion process in this case. Furthermore, as for $\alpha > 1$, the geometry of the cone becomes difficult to compute as N increases.

REFERENCES

[1] Anderson, T., Owicki, S., Saxe, J. and Thacker, C. (1993). High speed switch scheduling for local area networks. *ACM Transactions on Computer Systems* **11** 319-351.

[2] Berman, A. and Plemmons, R. J. (1994). *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, PA. MR1298430

[3] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[4] Bramson, M. (1998). State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Syst. Theory Appl.* **30** 89-148. MR1663763

[5] Bertsekas, D. P. (1995). *Nonlinear Programming*. Athena Scientific, Belmont, MA.

[6] Dantzig, G. B., Folkman, J. and Shapiro, N. (1967). On the continuity of the minimum set of a continuous function. *Journal of Mathematical Analysis and Applications* **17** 519-548. MR0207426

[7] Dai, J. and Prabhakar, B. (2000). The throughput of switches with and without speed-up. *Proceedings of IEEE Infocom* 556-564.

[8] Dai, J. and Williams, R. J. (1995). Existence and uniqueness of semimartingale reflecting Brownian motion in convex polyhedrons. *Theory Probab. Appl.* **40** 1-40. [Correctional note: **50** (2006), 346-347.] MR2222685

[9] Ethier, S. N. and Kurtz, T. G. (1986). Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York. MR0838085

[10] Harrison, J. M. (2000). Brownian models of open processing networks: canonical representation of workload. *Annals of Applied Probability* **10** 75–103. Correction: **13** (2003), 390–393. MR1765204

[11] McKeown, N., Anantharam, V. and Walrand, J. (1996). Achieving 100% throughput in an input-queued switch. *Proceedings of IEEE Infocom* 296-302.

[12] Kang, W. N., Kelly, F. P., Lee, N. H. and Williams, R. J. (2009). State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy. *Annals of Applied Probability* **19** 1719-1780. MR2569806

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.

[13] Kang W. and Williams, R. J. (2007). An invariance principle for semimartingale reflecting Brownian motions in domains with piecewise smooth boundaries. *Annals of Applied Probability* **17** 741-779. MR2308342

[14] Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York. MR0238472

[15] Shah, D., Tsitsiklis, J. N. and Zhong, Y. (2010). Qualitative properties of α-weighted scheduling policies. *Proceedings of ACM SIGMETRICS 2010* 239-250.

[16] Shah, D., Tsitsiklis, J. N. and Zhong, Y. (2011). Optimal scaling of average queue sizes in an input-queued switch: an open problem. *Queueing Systems* **68** 375-384. MR2834209

[17] Shah, D., Tsitsiklis, J. N. and Zhong, Y. (2011). Qualitative properties of α-fair policies in bandwidth sharing networks, preprint. MR2834209

[18] Shah, D., Walton, N. S. and Zhong, Y. (2011). Optimal queue-size scaling in switched networks, preprint.

[19] Shah, D. and Wischik, D. (2012). Switched networks with maximum weight policies: fluid approximation and multiplicative state space collapse. *Annals of Applied Probability* **22** 70-127. MR2932543

[20] Stolyar, A. L. (2004). MaxWeight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *Annals of Applied Probability* **14** 1–53. MR2023015

[21] Tassiulas, L. and Ephremides, A. (1992). Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control* **37** 1936-1949. MR1200609

[22] Williams, R. J. (1998). An invariance principle for semimartingale reflecting Brownian motions in an orthant. Special Issue of *Queueing Systems: Theory and Applications* **30** 5–25. MR1663755

[23] Yao, D. D. and Ye, H.-Q. (2012) A stochastic network under proportional fair resource control–Diffusion limit with multiple bottlenecks. *Operations Research* **60** 716-738.

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD 21250, USA
E-mail: wkang@umbc.edu

Department of Mathematics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0112, USA
E-mail: williams@math.ucsd.edu

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:34 . For personal use only, all rights reserved.