

Energy-Efficient Transmission Scheduling with Strict Underflow Constraints

David I Shuman, Mingyan Liu, and Owen Q. Wu

Abstract

We consider a single source transmitting data to one or more receivers/users over a shared wireless channel. Due to random fading, the wireless channel conditions vary with time and from user to user. Each user has a buffer to store received packets before they are drained. At each time step, the source determines how much power to use for transmission to each user. The source's objective is to allocate power in a manner that minimizes an expected cost measure, while satisfying strict buffer underflow constraints and a total power constraint in each slot. The expected cost measure is composed of costs associated with power consumption from transmission and packet holding costs. The primary application motivating this problem is wireless media streaming. For this application, the buffer underflow constraints prevent the user buffers from emptying, so as to maintain playout quality. In the case of a single user with linear power-rate curves, we show that a modified base-stock policy is optimal under the finite horizon, infinite horizon discounted, and infinite horizon average expected cost criteria. For a single user with piecewise-linear convex power-rate curves, we show that a finite generalized base-stock policy is optimal under all three expected cost criteria. We also present the sequences of critical numbers that complete the characterization of the optimal control laws in each of these cases when some additional technical conditions are satisfied. We then analyze the structure of the optimal policy for the case of two users. We conclude with a discussion of methods to identify implementable near-optimal policies for the most general case of M users.

Index Terms

Wireless media streaming, underflow constraints, opportunistic scheduling, energy-delay tradeoff, resource allocation, dynamic programming, inventory theory, base-stock policy.

David I Shuman and Mingyan Liu are with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI. Email: {dishuman, mingyan}@umich.edu

Owen Q. Wu is with the Ross School of Business, University of Michigan, Ann Arbor, MI. Email: owenwu@bus.umich.edu

Part of the work reported here was presented at the *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, April 2008, Berlin, Germany.

I. INTRODUCTION

In this paper, we examine the problem of energy-efficient transmission scheduling over a wireless channel, subject to underflow constraints. We consider a single source transmitting to one or more receivers/users over a shared wireless channel. Each user has a buffer to store received packets before they are drained at a certain rate. The available data rate of the channel varies with time and from user to user, due to random fading. The transmitter's goal is to minimize total power consumption by exploiting the temporal and spatial variation of the channel, while preventing any user's buffer from emptying.

A. *Opportunistic Scheduling and Related Work*

This problem falls into the general class of opportunistic scheduling problems, where the common theme is to exploit the temporal and spatial variation of the channel.¹ At a high level, the idea of exploiting the *temporal diversity* of the channel via opportunistic scheduling can be explained as follows. Consider the case of a single sender transmitting to a single receiver with different linear power-rate curves for each possible channel condition. Consider one scheduling policy that transmits data in a just-in-time fashion, without regard to the condition of the time-varying channel. Over the long run, the total power consumption tends toward the power consumption per data packet under the average channel condition times the number of packets sent. If instead, the scheduler aims to send more data when the channel is in a "good" state (requiring less power per data packet), and less data when the channel is in a "bad" state, the total power consumption should be lower. Much of the challenge for the scheduler lies in determining how good or bad a channel condition is, and how much data to send accordingly.

Similarly, in the case of multiple receivers, the scheduler can exploit the *spatial diversity* of the channel by transmitting only to those receivers who have the best channel conditions in each time slot. The benefit of increasing system throughput and reducing total power consumption through such a joint resource allocation policy is commonly referred to as the *multiuser diversity*

¹Opportunistic scheduling problems are also referred to as multi-user variable channel scheduling problems [1].

gain [2]. It was introduced in the context of the analogous *uplink* problem where multiple sources transmit to a single destination (e.g., the base station) [3]. Since, there has been a wide range of literature on opportunistic scheduling problems in wireless networks.

Sending more data when the channel is in a good state can increase system throughput and/or reduce total energy consumption; however, in opportunistic scheduling problems, it is often the case that the transmission scheduler has competing quality of service (QoS) interests. For instance, one QoS interest commonly considered is *fairness*. If, when a single source is transmitting to multiple receivers, the scheduler only considers total throughput and energy consumption across all users, it may often be the case that it ends up transmitting to a single user or the same small group of users in every slot. This can happen, for instance, if a base station requires less power to send data to a nearby receiver, even when the nearby receiver's channel is in its worst possible condition and a farther away receiver's channel is in its best possible condition. Thus, fairness constraints are often imposed to ensure that the transmitter sends packets to all receivers.

A number of different fairness conditions have been examined in the literature. For example, [4] and [5] consider *temporal fairness*, where the scheduler must transmit to each receiver for some minimum fraction of the time over the long run. Under the *proportional fairness* considered by [2] and [6], the scheduler considers the current channel conditions relative to the average channel condition of each receiver. Reference [5] considers a more general *utilitarian fairness*, where the focus is on system performance from the receiver's perspective, rather than on resources consumed by each user. The authors of [7] incorporate fairness directly into the objective function by setting relative throughput target values for each receiver and maximizing the minimum relative long-run average throughput.

Another QoS consideration that is important in many applications is delay. Different notions of delay have been incorporated into opportunistic scheduling problems. One proxy for delay is the *stability* of all of the sender's queues for arriving packets awaiting transmission. The motivation for this criterion is that if none of these queues blows up, then the delay is not "too bad." With stability as an objective, it is common to restrict attention to *throughput optimal* policies, which

are scheduling policies that ensure the sender's queues are stable, as long as this is possible for the given arrival process and channel model. References [8]-[11] present such throughput optimal scheduling algorithms, and examine conditions guaranteeing stabilizability in different settings.

When an arriving packet model is used for the data, one can also define *end-to-end delay* as the time between a packet's arrival at the sender's buffer and its decoding by the receiver. A number of opportunistic scheduling studies have considered the *average* end-to-end delay of all packets over a long horizon. For instance, [12]-[23] all consider average delay, either as a constraint or by incorporating it directly into the objective function to be minimized. However, the average delay criterion allows for the possibility of long delays (albeit with small probability); thus, for many delay-sensitive applications, *strict* end-to-end delay is often a more appropriate consideration for studies with arriving packet models. In [24] and [25], Chen, Mitra, and Neely place strict constraints on the end-to-end delay of each packet in a point-to-point system, examine the optimal scheduling policy assuming all future channel conditions are known, and suggest heuristics based on this optimal *offline* scheduling policy for the more realistic *online* case where the scheduler only learns the channel conditions in a causal fashion. Rajan, Sabharwal, and Aazhang also consider strict constraints on the end-to-end delay in an arriving packet model in [16, Section IV].

A strict constraint on the end-to-end delay of each packet is one particular form of a *deadline* constraint, as each packet has a deadline by which it must be transmitted. This notion can be generalized to impose individual deadlines on each packet, whether the packets are arriving over time or are all in the sender's buffer from the beginning. References [26]-[31] consider point-to-point communication when a fixed amount of data is in the sender's buffer at the start of the time horizon and the individual deadlines coincide, so that all packets must be transmitted and received by a common deadline, the end of the time horizon under consideration. In [26, Section III-D] and [27, Section III-D], Fu, Modiano, and Tsitsiklis specify the optimal transmission policy when the power-rate curves under each channel condition are linear and the transmitter is subject to a per slot peak power constraint. In [28]-[31], Lee and Jindal model the power-rate curve under

each channel condition as convex, first of the form of the so-called Shannon cost function based on the capacity of the additive white Gaussian noise channel, and then as a convex monomial function.²

References [32] and [33] consider opportunistic scheduling problems with multiple receivers and a single deadline constraint at the end of a finite horizon. Packets arrive over time and the emphasis is on offline scheduling policies in [32], whereas [33] considers a fixed amount of data destined for each receiver, and assumes the data is already in the sender's buffers at the beginning of the horizon. The model of [33] is perhaps the closest to our general model for M receivers; however, two key differences are (i) the transmitter is not subject to a power constraint in [33]; and (ii) the transmitter can transmit to at most one receiver in each time slot in [33].

In our model, the strict underflow constraints serve as a notion of both fairness and delay. The notion of fairness is that none of the receivers' buffers are allowed to empty, guaranteeing the required level of service to all users. The underflow constraints also serve as a notion of delay, and can be seen as multiple deadline constraints - certain packets must arrive by the end of the first slot, another group by the end of the second slot, and so forth. Therefore, Sections III and IV of this paper aim to generalize the works of [26]-[27] and [28]-[31], respectively, by considering multiple deadlines in the point-to-point communication problem, rather than a single deadline at the end of the horizon. In addition to better representing some delay-sensitive applications, this extension of the model also allows us to consider infinite horizon problems. We compare related work in opportunistic scheduling problems with deadline constraints further in [34]. For more complete surveys of opportunistic scheduling studies in wireless networks, see [35] and [36].

B. Wireless Media Streaming and Related Work

The primary application we have in mind to motivate this problem is wireless media streaming. For this application, the data are audio/video sequences, and the packets are drained from the

²In our notation of Section II-A, these two cases correspond to power-rate curves of the form $c(z, s) = \frac{2^z - 1}{g_1(s)}$ and $c(z, s) = \frac{z^\zeta}{g_2(s)}$, respectively, where $c(z, s)$ is the power required to transmit z bits under channel condition s , $g_1(\cdot)$ and $g_2(\cdot)$ are known functions, and ζ is a fixed parameter.

receivers' buffers in order to be decoded and played. Enforcing the underflow constraints reduces playout interruptions to the end users. In order to make the presentation concrete, we use the above wireless media streaming terminology throughout the paper.

Transporting multimedia over wireless networks is a promising application that has seen recent advances [37]. At the same time, a number of resource allocation issues need to be addressed in order to provide high quality and efficient media over wireless. First, streaming is in general bandwidth-demanding. Second, streaming applications tend to have stringent QoS requirements (e.g., they can be delay and jitter intolerant). Third, it is desirable to operate the wireless system in an energy-efficient manner. This is obvious when the source of the media streaming (the sender) is a mobile. When the media comes from a base station that is not power-constrained, it is still desirable to conserve power in order to (i) limit potential interference to other base stations and their associated mobiles, and (ii) maximize the number of receivers the sender can support.

Of the related work in wireless media streaming, [38] has the closest setup to our model. The main differences are that [38] features a loose constraint on underflow (i.e., it is allowed, but at a cost), as opposed to our tight constraint, and the two studies adopt different wireless channel models. In the extension [39], the receiver may slow down its playout rate (at some cost) to avoid underflow. In this setting, the authors investigate the tradeoffs between power consumption and playout quality, and examine joint power/playout rate control policies. In our model, the receiver does not have the option to adjust the playout speeds. Our model also bears resemblance to [40]. The first difference here is that [40] aims to minimize transmission energy subject to a constant end-to-end delay constraint on each video frame. A second difference is that the controller in [40] must assign various source coding parameters such as quantization step size and coding mode, whereas our model assumes a fixed encoding/decoding scheme.

C. Summary of Contribution

In this paper, we formulate the task of energy-efficient transmission scheduling subject to strict underflow constraints as three different Markov decision problems (MDPs), with the finite

horizon discounted expected cost, infinite horizon discounted expected cost, and infinite horizon average expected cost criteria, respectively. These three MDPs feature a continuous component of the state space and a continuous action space at each state. Therefore, unlike *finite* MDPs, they cannot in general be solved exactly via dynamic programming, and suffer from the well-known *curse of dimensionality* [41], [42]. Our aim in this paper is to analyze the dynamic programming equations in order to (i) determine if there are circumstances under which we can analytically derive optimal solutions to the three problems; and (ii) leverage our mathematical analysis and results on the structures of the optimal scheduling policies to improve our intuitive understanding of the problems.

We begin by showing that in the case of a single receiver under linear power-rate curves, the optimal policy is an easily-implementable *modified base-stock* policy. In each time slot, it is optimal for the sender to transmit so as to bring the number of packets in the receiver’s buffer level after transmission as close as possible to a target level or critical number.³ The target level depends on the current channel condition, with a better channel condition corresponding to a higher target level. We also show that the strict underflow constraints may cause the scheduler to be less opportunistic than it otherwise would be, and transmit more packets under “medium” channel conditions in anticipation of deadline constraints in future time slots.

We then generalize this result in two different directions. First, we relax the assumption that the power-rate curves under each channel condition are linear, and model them as piecewise-linear convex to better approximate more realistic convex power-rate curves. Under piecewise-linear power-rate curves, we show the optimal policy is a *finite generalized base-stock* policy, and provide an intuitive explanation of this structure in terms of multiple target levels in each time slot. In addition to the structural results on the optimal policy for the case of a single receiver under either linear or piecewise-linear convex power-rate curves, we provide an efficient method to calculate the critical numbers that complete the characterization of the optimal policy when certain technical conditions are satisfied.

³We use the terms target level and critical number interchangeably throughout the paper.

The second generalization of the single receiver model under linear power-rate curves is to a single user transmitting to two receivers over a shared wireless channel. In this case, we state and prove the structure of the optimal policy, and show how the peak power constraint in each slot couples the optimal scheduling of the two receivers' packet streams.

In all three setups, we show the structure of the optimal policy in the finite horizon discounted expected cost problem extends to the infinite horizon discounted and average expected cost problems.

Throughout the analysis, we make a novel connection with inventory models that may prove useful in other wireless transmission scheduling problems. Because the inventory models corresponding to our wireless communication models have not been previously examined, our results also represent a contribution to the inventory theory literature.

The remainder of this paper is organized as follows. In the next section, we describe the system model, formulate finite and infinite horizon MDPs, and relate our model to models in inventory theory. In Section III, we consider the case of a single receiver under linear power-rate curves. While this case can be considered a special case of the models of Sections IV and V, we present it first in order to (i) state additional structural properties of the optimal transmission policy to a single user under linear power-rate curves that are not true in general for the cases discussed in Sections IV and V; (ii) highlight some intuitive takeaways that carry over to the generalized models, but are more transparent in the simpler model; and (iii) compare it to related problems in the wireless communications literature. We analyze the structure of the optimal scheduling policy for the finite horizon problem, provide a method to compute the critical numbers that complete the characterization of the optimal policy when some additional technical conditions are met, and provide sufficient conditions for this problem to be equivalent to a previously-studied single deadline problem. Section IV generalizes the analysis of Section III to the case of a single receiver under piecewise-linear convex power-rate curves, and also addresses the infinite horizon problems for the case of a single receiver. In Section V, we analyze the structure of the optimal policy when there are two receivers with linear power-rate curves. We discuss the relaxation of the strict underflow constraints and the extension to the general case of M receivers in Section

VI. Section VII concludes the paper.

II. PROBLEM DESCRIPTION

In this section, we present an abstraction of the transmission scheduling problem outlined in the previous section and formulate three optimization problems. While most of this paper focuses on the cases of one and two users, the formulation in this section is for the more general multi-user (multi-receiver) case, so that we can discuss this more general case in Section VI-B.

A. System Model and Assumptions

We consider a single source transmitting media sequences to M users/receivers over a shared wireless channel. The sender maintains a separate buffer for each receiver, and is assumed to always have data to transmit to each receiver.⁴ We consider a *fluid* packet model that allows packet to be split, with the receiver reassembling fractional packets. Each receiver has a playout buffer at the receiving end, assumed to be infinite. While in reality this cannot be the case, it is nevertheless a reasonable assumption considering the decreasing cost and size of memory, and the fact that our system model allows holding costs to be assessed on packets in the receiver buffers. See Figure 1 for a diagram of the system.

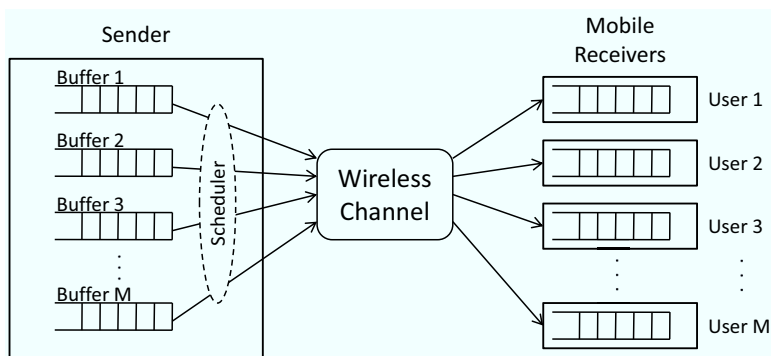


Fig. 1. System model.

⁴This assumption is commonly referred to as the infinite backlog assumption.

We consider time evolution in discrete steps, indexed backwards by $n = N, N - 1, \dots, 1$, with n representing the number of slots remaining in the time horizon. N is the length of the time horizon, and slot n refers to the time interval $[n, n - 1)$.

At the beginning of each time slot, the scheduler allocates some amount of power (possibly zero) for transmission to each user. The total power consumed in any one slot must not exceed the fixed power constraint, P . Following transmission and reception in each slot, a certain number of packets are removed/purged from each receiver buffer for playing. The transmitter (or scheduler) knows precisely the packet requirements of each receiver (i.e., the number of packets removed from the buffer) in each time slot. This is justified by the assumption that the transmitter knows the encoding and decoding schemes used. We assume that packets transmitted in slot n arrive in time to be used for playing in slot n , and that the users' consumption of packets in each slot is constant, denoted by $\mathbf{d} = (d^1, d^2, \dots, d^M)$. This latter assumption is less realistic, but may be justified if the receiving buffers are drained at a constant rate at the MAC layer, before packets are decoded by the media players at the application layer. It is also worth noting that the same techniques we use in this paper to analyze the constant drainage rate case can be used to examine the case of time-varying drainage rates. We discuss the extension to the case of time-varying drainage rates further in Section III-A. We also assume the receiver buffers are empty at the beginning of the time horizon, and that even when the channels are in their worst possible condition, the maximum power constraint P is sufficient to transmit enough packets to satisfy one time slot's packet requirements for every user. We discuss the relaxation of this assumption in Section VI-A.

In general, wireless channel conditions are time-varying. Adopting a block fading model, we assume that the slot duration is within the channel coherence time such that the channel conditions within a single slot are constant. User m 's channel condition in slot n is modeled as a random variable, S_n^m . We assume that the evolution of a given user's channel condition is independent of all other users' channel conditions and the transmitter's scheduling decisions. We also assume that the transmitter learns all the channel states through a feedback channel at the beginning of each time slot, prior to making the scheduling decisions.

We begin by modeling the evolution of each user's channel condition as a finite-state ergodic homogeneous Markov process, $\{S_n^m\}_{n=N, N-1, \dots, 1}$ with state space \mathcal{S}^m .⁵ Namely, conditioned on the channel state, S_n^m , at time n , user m 's channel states at *future* times ($n-1, n-2, \dots$) are independent of the channel states at *past* times ($n+1, n+2, \dots$). Note the somewhat unconventional notation that future times are indexed by lower epoch numbers, as n represents the number of slots remaining in the time horizon. Modeling time backwards facilitates the analysis of the infinite horizon problems, as will be seen for example in Section IV-C. It may also be the case that each user's channel condition is independent and identically distributed (IID) from slot to slot. When this is the case, we can often say more about the optimal transmission policy, as will be seen for example in Sections III-B and IV-B.

Associated with each channel condition for a given user is a power-rate function. If user m 's channel is in condition s^m , then the transmission of r units of data to user m incurs a power consumption of $c^m(r, s^m)$. This power-rate function $c^m(\cdot, s^m)$ is commonly assumed to be linear (in the low SNR regime) or convex (in the high SNR regime). In this paper, we consider power-rate functions that are linear or piecewise-linear convex, the latter of which can be used to approximate more general convex power-rate functions. We assume that sending data consumes a strictly positive amount of power, and therefore take the power-rate functions to be strictly increasing under all channel conditions.

The goal of this study is to characterize the control laws that minimize the transmission power and packet holding costs over a finite or infinite time horizon, subject to tight underflow constraints and a maximum power constraint in each time slot.

B. Notation

Before proceeding, we introduce some notation. We define $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{N} := \{1, 2, \dots\}$. A single dot, as in $a \cdot b$, represents scalar multiplication. We use bold font to denote column vectors, such as $\mathbf{w} = (w^1, w^2, \dots, w^M)$. We include a transpose superscript whenever a vector is

⁵Theorems 1, 3, 5, 7, 8, and 9 and their proofs remain valid as stated when each user's channel condition is given by a more general homogeneous Markov process that is not necessarily finite-state and ergodic.

meant to be a row vector, such as \mathbf{w}^T . The notations $\mathbf{w} \preceq \tilde{\mathbf{w}}$ and $\mathbf{w} \succeq \tilde{\mathbf{w}}$ denote component-wise inequalities; i.e., $w^m \leq$ (respectively, \geq) \tilde{w}^m , $\forall m$. Finally, we use the standard definitions of the meet and join of two vectors. Namely,

$$\begin{aligned} \mathbf{w} \wedge \tilde{\mathbf{w}} &= (w^1, w^2, \dots, w^M) \wedge (\tilde{w}^1, \tilde{w}^2, \dots, \tilde{w}^M) \\ &:= \left(\min \{w^1, \tilde{w}^1\}, \min \{w^2, \tilde{w}^2\}, \dots, \min \{w^M, \tilde{w}^M\} \right), \\ \text{and } \mathbf{w} \vee \tilde{\mathbf{w}} &= (w^1, w^2, \dots, w^M) \vee (\tilde{w}^1, \tilde{w}^2, \dots, \tilde{w}^M) \\ &:= \left(\max \{w^1, \tilde{w}^1\}, \max \{w^2, \tilde{w}^2\}, \dots, \max \{w^M, \tilde{w}^M\} \right). \end{aligned}$$

C. Problem Formulation

We consider three problems. Problem **(P1)** is the finite horizon discounted expected cost problem; Problem **(P2)** is the infinite horizon discounted expected cost problem; and Problem **(P3)** is the infinite horizon average expected cost problem. The three problems feature the same information state, action space, system dynamics, and cost structure, but different optimization criteria.

The information state at time n is the pair $(\mathbf{X}_n, \mathbf{S}_n)$, where the random vector $\mathbf{X}_n = (X_n^1, X_n^2, \dots, X_n^M)$ denotes the current receiver buffer queue lengths, and $\mathbf{S}_n = (S_n^1, S_n^2, \dots, S_n^M)$ denotes the channel conditions in slot n (recall that n is the number of steps remaining until the end of the horizon). The dynamics for the receivers' queues are governed by the simple equation $\mathbf{X}_{n-1} = \mathbf{X}_n + \mathbf{Z}_n - \mathbf{d}$ at all times $n = N, N-1, \dots, 1$, where \mathbf{Z}_n is a controlled random vector chosen by the scheduler at each time n that represents the number of packets transmitted to each user in the n^{th} slot. At each time n , \mathbf{Z}_n must be chosen to meet the peak power constraint:

$$\sum_{m=1}^M c^m(Z_n^m, S_n^m) \leq P,$$

and the underflow constraints:

$$X_n^m + Z_n^m \geq d^m, \quad \forall m \in \{1, 2, \dots, M\}.$$

Clearly, the scheduler cannot transmit a negative number of packets to any user, so it must also be true that $Z_n^m \geq 0$ for all m .

We now present the optimization criterion for each problem. In addition to the cost associated with power consumption from transmission, we introduce holding costs on packets stored in each user's playout buffer at the end of a time slot. The holding costs associated with user m in each slot are described by a convex, nonnegative, nondecreasing function, $h^m(\cdot)$, of the packets remaining in user m 's buffer following playout, with $\lim_{x \rightarrow \infty} h^m(x) = \infty$. We assume without loss of generality that $h^m(0) = 0$. Possible holding cost models include a linear model, $h^m(x) = \hat{h}^m \cdot x$ for some positive constant \hat{h}^m , or a barrier-type function such as:

$$h^m(x) := \begin{cases} 0, & \text{if } x \leq \mu \\ \kappa \cdot (x - \mu), & \text{if } x > \mu \quad (\kappa \text{ very large}) \end{cases},$$

which could represent a finite receiver buffer of length μ .⁶

In Problem **(P1)**, we wish to find a transmission policy π that minimizes $J_{N,\alpha}^\pi$, the finite horizon discounted expected cost under policy π , defined as:

$$J_{N,\alpha}^\pi := \mathbb{E}^\pi \left\{ \sum_{t=1}^N \sum_{m=1}^M \alpha^{N-t} \cdot \left\{ c^m(Z_t^m, S_t^m) + h^m(X_t^m + Z_t^m - d^m) \right\} \mid \mathcal{F}_N \right\},$$

where $0 \leq \alpha \leq 1$ is the discount factor and \mathcal{F}_N denotes all information available at the beginning of the time horizon. For Problem **(P2)**, the discount factor must satisfy $0 \leq \alpha < 1$, and the infinite horizon discounted expected cost function for minimization is defined as:

$$J_{\infty,\alpha}^\pi := \lim_{N \rightarrow \infty} J_{N,\alpha}^\pi,$$

For Problem **(P3)**, the average expected cost function for minimization is defined as:

$$J_{\infty,1}^\pi := \limsup_{N \rightarrow \infty} \frac{1}{N} J_{N,1}^\pi.$$

In all three cases, we allow the transmission policy π to be chosen from the set of all history-

⁶Taking μ to be greater than the time horizon N in the finite horizon expected cost problem is equivalent to not assessing any holding costs in Problem **(P1)**.

dependent randomized and deterministic control laws, Π (see, e.g., [43, Definition 2.2.3, pg. 15]).

Combining the constraints and criteria, we present the optimization formulations for Problem **(P1)** (or **(P2)** or **(P3)**):

$$\begin{aligned} & \inf_{\pi \in \Pi} J_{N,\alpha}^\pi \quad \left(\text{or } \inf_{\pi \in \Pi} J_{\infty,\alpha}^\pi \text{ or } \inf_{\pi \in \Pi} J_{\infty,1}^\pi \right) \\ \text{s.t. } & \sum_{m=1}^M c^m(Z_n^m, S_n^m) \leq P, \text{ w.p.1, } \forall n \\ & Z_n^m \geq \max\{0, d^m - X_n^m\}, \text{ w.p.1, } \forall n, \forall m \in \{1, 2, \dots, M\}. \end{aligned}$$

Problem **(P1)** may be solved using standard dynamic programming (see, e.g., [43], [44]). The recursive dynamic programming equations are given by:⁷

$$V_n(\mathbf{x}, \mathbf{s}) = \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \sum_{m=1}^M \{c^m(z^m, s^m) + h^m(x^m + z^m - d^m)\} \\ + \alpha \cdot \mathbb{E}[V_{n-1}(\mathbf{x} + \mathbf{z} - \mathbf{d}, \mathbf{S}_{n-1}) \mid \mathbf{S}_n = \mathbf{s}] \end{array} \right\} \quad (1)$$

$n = N, N-1, \dots, 1$

$$V_0(\mathbf{x}, \mathbf{s}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}_+^M, \forall \mathbf{s} \in \mathcal{S} := \mathcal{S}^1 \times \mathcal{S}^2 \times \dots \times \mathcal{S}^M,$$

where $V(\cdot, \cdot)$ is the value function or expected cost-to-go, and the action space is defined as:

$$\mathcal{A}^d(\mathbf{x}, \mathbf{s}) := \left\{ \mathbf{z} \in \mathbb{R}_+^M : \begin{array}{l} \mathbf{z} \succeq \max\{\mathbf{0}, \mathbf{d} - \mathbf{x}\} \text{ and} \\ \sum_{m=1}^M c^m(z^m, s^m) \leq P \end{array} \right\}, \quad \forall \mathbf{x} \in \mathbb{R}_+^M, \forall \mathbf{s} \in \mathcal{S}, \quad (2)$$

where the maximum in (2) is taken element-by-element (i.e., $z^m \geq \max\{0, d^m - x^m\} \forall m$). Note that our assumption that the maximum power constraint P is always sufficient to transmit enough packets to satisfy one time slot's packet requirements for every user (i.e., $\sum_{m=1}^M c^m(d^m, s^m) \leq P, \forall \mathbf{s} \in \mathcal{S}$) ensures that the action space $\mathcal{A}^d(\mathbf{x}, \mathbf{s})$ is always non-empty.

⁷As will be shown in the proofs of Theorems 6 and 10, our model satisfies the measurable selection condition 3.3.3 of [43, pg. 28], justifying the use of min rather than inf in the dynamic programming equations.

D. Relation to Inventory Theory

The model outlined in Section II-A corresponds closely to models used in inventory theory. Borrowing that field's terminology, our abstraction is a multi-period, single-echelon, multi-item, discrete-time inventory model with random (linear or piecewise-linear convex) ordering costs, a budget constraint, and deterministic demands. The items correspond to the streams of data packets, the random ordering costs to the random channel conditions, the budget constraint to the power available in each time slot, and the deterministic demands to the packet requirements for playout.

To the best of our knowledge, this particular problem has not been studied in the context of inventory theory, but similar problems have been examined, and some of the techniques from the inventory theory literature are useful in analyzing our model. References [45]-[52] all consider single-item inventory models with linear ordering costs and random prices. The key result for the case of deterministic demand of a single item with no resource constraint is that the optimal policy is a base-stock policy with different target stock levels for each price. Specifically, for each possible ordering price (translates into channel condition in our context), there exists a critical number such that the optimal policy is to fill the inventory (receiver buffer) up to that critical number if the current level is lower than the critical number, and not to order (transmit) anything if the current level is above the critical number. Of the prior work, Kingsman [47], [48] is the only author to consider a resource constraint, and he imposes a maximum on the number of items that may be ordered in each slot. The resource constraint we consider is of a different nature in that we limit the amount of power available in each slot. This is equivalent to a limit on the per slot budget (regardless of the stochastic price realization), rather than a limit on the number of items that can be ordered.

Of the related work on single-item inventory models with deterministic linear ordering costs and stochastic demand, [53] and [54] are the most relevant; in those studies, however, the resource constraint also amounts to a limit on the number of items that can be ordered in each slot, and is constant over time. References [55]-[57] consider single-item inventory models with deterministic piecewise-linear convex ordering costs and stochastic demand. The key result

in this setup is that the optimal inventory level after ordering is a piecewise-linear nondecreasing function of the current inventory level (i.e., there are a finite number of target stock levels), and the optimal ordering quantity is a piecewise-linear nonincreasing function of the current inventory level. Porteus [58] refers to policies of this form as *finite generalized base-stock policies*, to distinguish them from the superclass of *generalized base-stock policies*, which are optimal when the deterministic ordering costs are convex (but not necessarily piecewise-linear), as first studied in [59]. Under a generalized base-stock policy, the optimal inventory level after ordering is a nondecreasing function of the current inventory level, and the optimal ordering quantity is a nonincreasing function of the current inventory level.

References [60]-[63] consider multi-item inventory systems under deterministic ordering costs, stochastic demand, and resource constraints. We discuss related results from these studies in more detail in Section V.

We are not aware of any prior work on (i) single-item inventory models with random piecewise-linear convex ordering costs; (ii) exact computation of the critical numbers in any sort of finite generalized base-stock policy; or (iii) multi-item inventory models with random ordering costs and joint resource constraints. Therefore, not only is this connection between wireless transmission scheduling problems and inventory models novel, but the results we present in this paper also represent a contribution to the inventory theory literature.

III. SINGLE RECEIVER WITH LINEAR POWER-RATE CURVES

In this section, we analyze the finite horizon discounted expected cost problem when there is only a single receiver ($M = 1$), and the power-rate functions under different channel conditions are linear. One such family of power-rate functions is shown in Figure 2, where there are three possible channel conditions, and a different linear power-rate function associated with each channel condition. Note that due to the power constraint P in each slot, the effective power-rate function is a two-segment piecewise-linear convex function under all channel conditions. We subsequently simplify our notation and use c_s to denote the power consumption per unit of data transmitted when the channel condition is in state s . Because there is just a single receiver, we

also drop the dependence of the functions and random variables on m . We defer the infinite horizon expected cost problems for this case until Section IV-C.

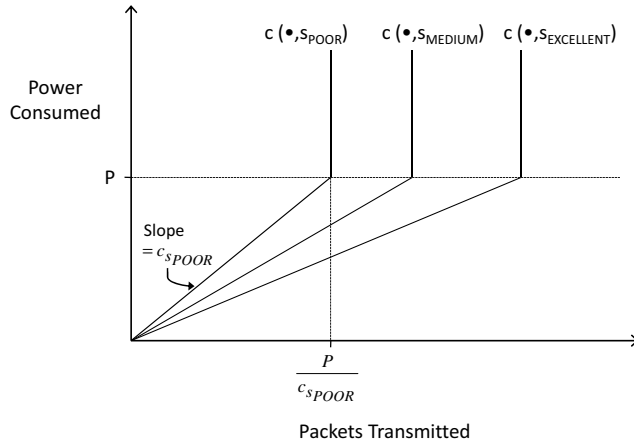


Fig. 2. A family of linear power-rate functions. Due to the power constraint, the effective power-rate function, shown above for each of the three channel conditions, is a two-segment piecewise-linear convex function. When the channel condition is s , the slope of the first segment is c_s .

We denote the “best” and “worst” channel conditions by s_{best} and s_{worst} , respectively, and denote the slopes of the power-rate functions under these respective conditions by c_{\min} and c_{\max} . That is,

$$0 < c_{s_{\text{best}}} = c_{\min} := \min_{s \in \mathcal{S}} \{c_s\} \leq \max_{s \in \mathcal{S}} \{c_s\} =: c_{\max} = c_{s_{\text{worst}}} \leq \frac{P}{d}.$$

With these notations in place, the dynamic program (1) for Problem **(P1)** becomes:

$$V_n(x, s) = \min_{\max(0, d-x) \leq z \leq \frac{P}{c_s}} \left\{ \begin{array}{l} c_s \cdot z + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_{n-1}(x + z - d, S_{n-1}) \mid S_n = s] \end{array} \right\} \quad (3)$$

$$= \min_{\max(x, d) \leq y \leq x + \frac{P}{c_s}} \left\{ \begin{array}{l} c_s \cdot (y - x) + h(y - d) \\ + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1}) \mid S_n = s] \end{array} \right\} \quad (4)$$

$$= -c_s \cdot x + \min_{\max(x, d) \leq y \leq x + \frac{P}{c_s}} \{g_n(y, s)\}, \quad n = N, N-1, \dots, 1,$$

$$V_0(x, s) = 0, \quad \forall x \in \mathbb{R}_+, \forall s \in \mathcal{S},$$

where $g_n(y, s) := c_s \cdot y + h(y - d) + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1}) \mid S_n = s]$. Here, the transition from (3)

to (4) is done by a change of variable in the action space from Z_n to Y_n , where $Y_n = X_n + Z_n$. The controlled random variable Y_n represents the queue length of the receiver buffer *after* transmission takes place in the n^{th} slot, but *before* playout takes place (i.e., before d packets are removed from the buffer). The restrictions on the action space, $\max(x, d) \leq y \leq x + \frac{P}{c_s}$, ensure: (i) a nonnegative number of packets is transmitted; (ii) there are at least d packets in the receiver buffer following transmission, in order to satisfy the underflow constraint; and (iii) the power constraint is satisfied.

A. Structure of Optimal Policy

With the above change of variable in the the action space, the expected cost-to-go at time n , $V_n(x, s)$, depends on the current buffer level, x , only through the fixed term $-c_s \cdot x$ and the action space; i.e., the function g_n does not depend on x . This separation allows us to leverage the inventory theory techniques of showing “single critical number” or “base-stock” policies, which date as far back as [64]. The following theorem gives the structure of the optimal transmission policy for the finite horizon discounted expected cost problem.

Theorem 1. *For every $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$, define the critical number*

$$b_n(s) := \min \left\{ \hat{y} \in [d, \infty) : g_n(\hat{y}, s) = \min_{y \in [d, \infty)} g_n(y, s) \right\} .$$

Then, for Problem (P1) in the case of a single receiver with linear power-rate curves, the optimal buffer level after transmission with n slots remaining is given by:

$$y_n^*(x, s) := \begin{cases} x, & \text{if } x \geq b_n(s) \\ b_n(s), & \text{if } b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ x + \frac{P}{c_s}, & \text{if } x < b_n(s) - \frac{P}{c_s} \end{cases} , \quad (5)$$

or, equivalently, the optimal number of packets to transmit in slot n is given by:

$$z_n^*(x, s) := \begin{cases} 0, & \text{if } x \geq b_n(s) \\ b_n(s) - x, & \text{if } b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ \frac{P}{c_s}, & \text{if } x < b_n(s) - \frac{P}{c_s} \end{cases} . \quad (6)$$

Furthermore, for a fixed s , $b_n(s)$ is nondecreasing in n :

$$N \cdot d \geq b_N(s) \geq b_{N-1}(s) \geq \dots \geq b_1(s) = d . \quad (7)$$

If, in addition, the channel condition is independent and identically distributed from slot to slot, then for a fixed n , $b_n(s)$ is nonincreasing in c_s ; i.e., for arbitrary $s^1, s^2 \in \mathcal{S}$ with $c_{s^1} \leq c_{s^2}$, we have:

$$n \cdot d \geq b_n(s_{\text{best}}) \geq b_n(s^1) \geq b_n(s^2) \geq b_n(s_{\text{worst}}) = d . \quad (8)$$

The optimal transmission policy in Theorem 1 is a modified base-stock policy. At time n , for each possible channel condition realization s , the critical number $b_n(s)$ describes the target number of packets to have in the user's buffer after transmission in the n^{th} slot. If that number of packets is already in the buffer, then it is optimal to not transmit any packets; if there are fewer than the target and the available power is enough to transmit the difference, then it is optimal to do so; and if there are fewer than the target and the available power is not enough to transmit the difference, then the sender should use the maximum power to transmit. See Figure 3 for diagrams of the optimal policy.

Details of the proof of Theorem 1 are included in Appendix A. The key realization is that for all n and all s , $g_n(\cdot, s) : [d, \infty) \rightarrow \mathbb{R}_+$ is a convex function in y , with $\lim_{y \rightarrow \infty} g_n(y, s) = \infty$. Thus, for all n and all s , $g_n(\cdot, s)$ has a global minimum $b_n(s)$, the target number of packets to have in the buffer following transmission in the n^{th} slot. The key idea to show (7) is to fix $s \in \mathcal{S}$, view $g_n(y, s)$ as a function of y and n , say $f(y, n)$, and show that the function $f(\cdot, \cdot)$ is submodular. From the proof, one can also see that if we relax the stationary (time-invariant) deterministic demand assumption to a nonstationary (time-varying) deterministic

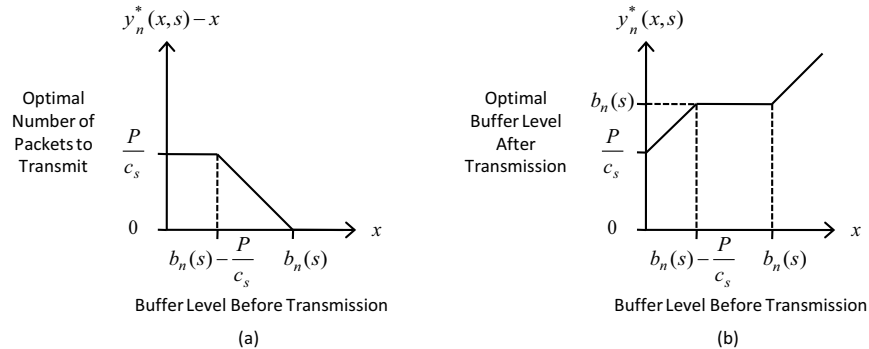


Fig. 3. Optimal policy in slot n when the state is (x, s) . (a) depicts the optimal transmission quantity, and (b) depicts the resulting number of packets available for playout in slot n .

demand sequence, $\{d_N, d_{N-1}, \dots, d_1\}$ (with $d_n \leq \frac{P}{c_{\max}}$ for all n), then the structure of the optimal policy is still as stated in (5). If the channel is IID, then the following statement, analogous to (8), is true for arbitrary $s^1, s^2 \in \mathcal{S}$ with $c_{s^1} \leq c_{s^2}$:

$$\sum_{i=1}^n d_i \geq b_n(s_{\text{best}}) \geq b_n(c_{s^1}) \geq b_n(c_{s^2}) \geq b_n(s_{\text{worst}}) = d_n, \quad \forall n \in \{1, 2, \dots, N\}. \quad (9)$$

However, (7), the monotonicity of critical numbers over time for a fixed channel condition, is not true in general under nonstationary deterministic demand. As one counterexample, (9) says that under an IID channel, the critical numbers for the worst possible channel condition are equal to the single period demands. Therefore, if the demand sequence is not monotonic, the sequence of critical numbers, $\{b_n(s_{\text{worst}})\}_{n=1,2,\dots,N}$, is not monotonic.

B. Computation of the Critical Numbers

In this section, we consider the special case where the channel condition is independent and identically distributed from slot to slot, the holding cost function is linear (i.e., $h(x) = h \cdot x$ for some $h \geq 0$), and the following technical condition is satisfied: for each possible channel condition s , $\frac{P}{c_s} = l \cdot d$ for some $l \in \mathbb{N}$; i.e., the maximum number of packets that can be transmitted in any slot covers exactly the playout requirements of some integer number of slots. Under these three assumptions, we can completely characterize the optimal transmission policy.

Theorem 2. Define the threshold $\gamma_{n,j}$ for $n \in \{1, 2, \dots, N\}$ and $j \in \mathbb{N}$ recursively, as follows:

- (i) If $j = 1$, $\gamma_{n,j} = \infty$;
- (ii) If $j > n$, $\gamma_{n,j} = 0$;
- (iii) If $2 \leq j \leq n$,

$$\gamma_{n,j} = -h + \alpha \cdot \left(\begin{array}{l} \sum_{s: c_s \geq \gamma_{n-1,j-1}} p(s) \cdot \gamma_{n-1,j-1} + \sum_{s: c_s < \gamma_{n-1,j-1}} p(s) \cdot c_s \\ + \sum_{s: c_s < \gamma_{n-1,j-1+L(s)}} p(s) \cdot [\gamma_{n-1,j-1+L(s)} - c_s] \end{array} \right), \quad (10)$$

where $p(s)$ is the probability of the channel being in state s in a time slot, and $L(s) := \frac{P}{d \cdot c_s}$. For each $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$, if $\gamma_{n,j+1} \leq c_s < \gamma_{n,j}$, define $b_n(s) := j \cdot d$. The optimal control strategy for Problem **(P1)** is then given by $\pi^* = \{y_N^*, y_{N-1}^*, \dots, y_1^*\}$, where

$$y_n^*(x, s) := \begin{cases} x, & \text{if } x \geq b_n(s) \\ b_n(s), & \text{if } b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ x + \frac{P}{c_s}, & \text{if } x < b_n(s) - \frac{P}{c_s} \end{cases} . \quad (11)$$

Note that with n slots remaining, $0 = \gamma_{n,n+1} \leq \gamma_{n,n} \leq \gamma_{n,n-1} \leq \dots \leq \gamma_{n,2} \leq \gamma_{n,1} = \infty$, so $b_n(s)$ is well-defined.

Compared to using standard numerical techniques to *approximately* solve the dynamic program and find a near-optimal policy, the above result not only sheds more insight on the structural properties of the problem and its *exactly*-optimal solution, but also offers a computationally simpler method. In particular, the optimal policy is completely characterized by the thresholds $\{\gamma_{n,j}\}_{n \in \{1,2,\dots,N\}, j \in \mathbb{N}}$. Calculating these thresholds recursively, as described in Theorem 2, requires $O(N^2 |\mathcal{S}|)$ operations, which is considerably simpler from a computational standpoint than approximately solving the dynamic program [41], [42].

To prove Theorem 2, we show by backwards induction that it is worse to transmit either fewer or more packets than the number suggested by the policy π^* . The detailed proof is omitted, as Theorem 2 is a special case of Theorem 4; however, we discuss some intuition behind the proof and the thresholds here.

The reason for the technical condition regarding the maximum number of packets that can be transmitted in any slot is as follows. The optimal action at all times (in general, without the technical condition) is either to transmit enough packets to fill the buffer up to a level satisfying the playout requirements of some number of future slots, or to transmit at maximum power. When the technical condition is satisfied, transmitting at maximum power also results in filling the buffer up to a level satisfying the playout requirements of some number of future slots. Thus, under the optimal policy, all realizations result in the buffer level at the end of every time slot being some integer multiple of the demand, d . This fact makes it easier to compute the thresholds $\{\gamma_{n,j}\}_{n \in \{1,2,\dots,N\}, j \in \mathcal{N}}$.

An intuitive explanation of the recursion (10) is as follows. The threshold $\gamma_{n,j}$ may be interpreted as the per packet power cost at which, with n slots remaining in the horizon, the expected cost-to-go of transmitting packets to cover the user's playout requirements for the next $j - 1$ slots is the same as the expected cost-to-go of transmitting packets to cover the user's requirements for the next j slots. That is, $\gamma_{n,j}$ should satisfy:

$$\alpha \cdot \mathbb{E} \left[V_{n-1} \left((j-1) \cdot d, S_{n-1} \right) \right] + \gamma_{n,j} \cdot d + h \cdot d = \alpha \cdot \mathbb{E} \left[V_{n-1} \left((j-2) \cdot d, S_{n-1} \right) \right],$$

which is equivalent to:

$$\begin{aligned} & \gamma_{n,j} \\ &= -h + \frac{\alpha}{d} \cdot \mathbb{E} \left[V_{n-1} \left((j-2) \cdot d, S_{n-1} \right) - V_{n-1} \left((j-1) \cdot d, S_{n-1} \right) \right] \\ &= -h + \frac{\alpha}{d} \cdot \sum_{s \in \mathcal{S}} p(s) \cdot \left[V_{n-1} \left((j-2) \cdot d, s \right) - V_{n-1} \left((j-1) \cdot d, s \right) \right] \end{aligned} \quad (12)$$

$$= -h + \frac{\alpha}{d} \cdot \left\{ \begin{aligned} & \sum_{s: b_{n-1}(s) \leq (j-2) \cdot d} p(s) \cdot \left\{ -h \cdot d + \alpha \cdot \mathbb{E} \left[\begin{array}{l} V_{n-2}((j-3) \cdot d, S_{n-2}) \\ -V_{n-2}((j-2) \cdot d, S_{n-2}) \end{array} \right] \right\} \\ & + \sum_{s: (j-2) \cdot d < b_{n-1}(s) \leq (j-2+L(s)) \cdot d} p(s) \cdot c_s \cdot d \\ & + \sum_{s: b_{n-1}(s) > (j-2+L(s)) \cdot d} p(s) \cdot \left\{ -h \cdot d + \alpha \cdot \mathbb{E} \left[\begin{array}{l} V_{n-2}((j-3+L(s)) \cdot d, S_{n-2}) \\ -V_{n-2}((j-2+L(s)) \cdot d, S_{n-2}) \end{array} \right] \right\} \end{aligned} \right\} \quad (13)$$

$$= -h + \alpha \cdot \left\{ \begin{aligned} & \sum_{s: b_{n-1}(s) \leq (j-2) \cdot d} p(s) \cdot \gamma_{n-1, j-1} \\ & + \sum_{s: (j-2) \cdot d < b_{n-1}(s) \leq (j-2+L(s)) \cdot d} p(s) \cdot c_s \\ & + \sum_{s: b_{n-1}(s) > (j-2+L(s)) \cdot d} p(s) \cdot \gamma_{n-1, j-1+L(s)} \end{aligned} \right\} \quad (14)$$

$$= -h + \alpha \cdot \left\{ \begin{aligned} & \sum_{s: c_s \geq \gamma_{n-1, j-1}} p(s) \cdot \gamma_{n-1, j-1} \\ & + \sum_{s: \gamma_{n-1, j-1+L(s)} \leq c_s < \gamma_{n-1, j-1}} p(s) \cdot c_s \\ & + \sum_{s: c_s < \gamma_{n-1, j-1+L(s)}} p(s) \cdot \gamma_{n-1, j-1+L(s)} \end{aligned} \right\}. \quad (15)$$

Here, (13) follows from the structure of the optimal control action (5). If the channel condition s in the $(n-1)^{st}$ slot is such that $b_{n-1}(s) \leq (j-2) \cdot d$, then no packets are transmitted when the starting buffer level is either $(j-2) \cdot d$ or $(j-1) \cdot d$, and the respective buffer levels at the beginning of slot $n-2$ are $(j-3) \cdot d$ and $(j-2) \cdot d$. The instantaneous costs resulting from the two starting buffer levels differ by $-h \cdot d$. When $(j-2) \cdot d < b_{n-1}(s) \leq (j-2+L(s)) \cdot d$, the power constraint is not tight starting from $(j-1) \cdot d$, so the buffer level after transmission is the same starting from $(j-2) \cdot d$ or $(j-1) \cdot d$. The instantaneous costs resulting from the two starting buffer levels differ by $c_s \cdot d$, as an extra d packets are transmitted if the starting buffer is $(j-2) \cdot d$. Finally, when $b_{n-1}(s) > (j-2+L(s)) \cdot d$, the power constraint is tight starting from both $(j-2) \cdot d$ and $(j-1) \cdot d$. Therefore, the instantaneous cost difference is $-h \cdot d$, and the respective buffer levels at the beginning of slot $n-2$ are $(j-3+L(s)) \cdot d$ and $(j-2+L(s)) \cdot d$. Equation (14) follows from (12), with $n-1, j-1$ substituted for n, j , and (15) follows from

the definition that $b_n(s) = j \cdot d$ if $\gamma_{n,j+1} \leq c_s < \gamma_{n,j}$.

Comparing the threshold $\gamma_{n,j}$ defined in (10) to the corresponding threshold in the unrestricted (no power constraint) single user problem [47], [52], the only difference is the third term of the right-hand side of (10):

$$\alpha \cdot \sum_{\{s: c_s < \gamma_{n-1,j-1+L(s)}\}} p(s) \cdot [\gamma_{n-1,j-1+L(s)} - c_s],$$

which is absent in the unrestricted case. For all $n \in \{1, 2, \dots, N\}$ and $j \in \mathbb{N}$, this term is nonnegative. Thus, for a fixed n and j , the threshold in the restricted case is at least as high as the corresponding threshold in the unrestricted case. It follows that the optimal stock-up level $b_n(s)$ is also at least as high in the restricted case for all $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$. The intuition behind this difference is that the sender should transmit more packets under the same (medium) conditions, because it is not able to take advantage of the best channel conditions to the same extent due to the power constraint.

C. Sufficient Conditions for Equivalence with the Single Deadline Problem

In [27, Section III-D], Fu, Modiano, and Tsitsiklis consider the related single user problem of transmitting a given amount of data with minimum energy by a fixed deadline. They also represent the fading channel by a linear power-rate function with a different slope in each channel condition, and consider a power constraint P in each slot. There is just a single explicit underflow constraint (the deadline) in their problem; however, because the terminal cost is set to ∞ if all the data is not transmitted by the deadline, the scheduler must transmit enough data in each slot so that it can still complete the job if the channel is in the worst possible condition in all subsequent slots. Thus, if d_{total} is the total amount of data that must be sent by the deadline and d_{worst} is the amount that can be sent in a slot under the worst channel condition, the transmitter must have sent at least $d_{\text{total}} - d_{\text{worst}}$ packets by the beginning of the last slot, at least $d_{\text{total}} - 2 \cdot d_{\text{worst}}$ packets by the beginning of the second to last slot, and so forth.⁸ So

⁸An unstated assumption in the formulation in [27, Section III-D] is that d_{worst} times the horizon length must be at least as large as d_{total} .

there are in fact implicit constraints on how much data must be transmitted by the end of slots $N - \left\lceil \frac{d_{\text{total}}}{d_{\text{worst}}} \right\rceil + 1, N - \left\lceil \frac{d_{\text{total}}}{d_{\text{worst}}} \right\rceil + 2, \dots, N - 2, N - 1$. With this interpretation, we believe that our Theorem 2 is equivalent to Theorem 3 and its corollary in [27] in the special case that, in addition to the hypotheses of our Theorem 2, $\alpha = 1$, $h = 0$, and $L(s_{\text{worst}}) = 1$. For, when these conditions are met, the implicit constraints in [27] coincide exactly with the explicit underflow constraints in our problem. Of course, when these three conditions are not satisfied, the two problems are quite different. For a more detailed comparison of these two problems, see [34].

D. Intuitive Takeaways on the Role of the Strict Underflow Constraints

As mentioned earlier, the main idea of energy-efficient communication over a fading channel via opportunistic scheduling is to minimize power consumption by transmitting more data when the channel is in a “good” state, and less data when the channel is in a “bad” state. However, in order to comply with the underflow or deadline constraints, the transmitter may be forced to send data under poor channel conditions.

One intuitive takeaway from the analysis is that it is better to anticipate the need to comply with these constraints in future slots by sending more packets (than one would without the deadlines) under “medium” channel conditions in earlier slots. Doing so is a way to manage the risk of being stuck sending a large amount of data over a poor channel to meet an imminent deadline constraint. Another intuitive takeaway is that the closer the deadlines and the more deadlines it faces, the less “opportunistic” the scheduler can afford to be. In summary, both the underflow constraints and the power constraints shift the definition of what constitutes a “good” channel, and how much data to send accordingly. For more detailed comparisons of single-receiver opportunistic scheduling problems highlighting the role of the deadline constraints, see [34].

IV. SINGLE RECEIVER WITH PIECEWISE-LINEAR CONVEX POWER-RATE CURVES

In this section, we analyze Problems **(P1)**, **(P2)**, and **(P3)** when there is only a single receiver ($M = 1$), and the power-rate functions under different channel conditions are piecewise-linear convex. Note that this is a generalization of the case considered in Section III.

We assume without loss of generality that under each channel condition s , the power-rate function has $K + 1$ segments, and thus the power consumed in transmitting z packets under channel condition s can be represented as follows:

$$c(z, s) = z \cdot \tilde{c}_0(s) + \sum_{k=0}^{K-1} \left\{ (\tilde{c}_{k+1}(s) - \tilde{c}_k(s)) \cdot \max\{z - \tilde{z}_k(s), 0\} \right\}, \text{ where}$$

$$0 < \tilde{c}_0(s) \leq \tilde{c}_1(s) \leq \dots \leq \tilde{c}_K(s), \text{ and}$$

$$0 = \tilde{z}_{-1}(s) < \tilde{z}_0(s) < \tilde{z}_1(s) < \dots < \tilde{z}_{K-1}(s) < \tilde{z}_K(s) = \infty.$$

The terms $\{\tilde{c}_k(s)\}_{k \in \{0,1,\dots,K\}}$ represent the slopes of the segments of $c(\cdot, s)$, and the terms $\{\tilde{z}_k(s)\}_{k \in \{0,1,\dots,K-1\}}$ represent the points at which the slopes of $c(\cdot, s)$ change. An example of a family of such power-rate functions is shown in Figure 4. For each channel condition $s \in \mathcal{S}$, we define the maximum number of packets that can be transmitted without exceeding the per slot power constraint P as:

$$\tilde{z}_{\max}(s) := \{z : c(z, s) = P\}.$$

Note that $\tilde{z}_{\max}(s)$ is well-defined due to the strictly increasing nature of $c(\cdot, s)$. Recall that we assume $\tilde{z}_{\max}(s) \geq d, \forall s \in \mathcal{S}$. We also assume without loss of generality that $\tilde{z}_{\max}(s) > \tilde{z}_{K-1}(s), \forall s \in \mathcal{S}$.

In this case, the dynamic program (1) for Problem **(P1)** becomes:

$$\begin{aligned} V_n(x, s) &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_{n-1}(x + z - d, S_{n-1}) \mid S_n = s] \end{array} \right\} \\ &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ c(z, s) + \tilde{g}_n(x + z, s) \right\}, \quad n = N, N-1, \dots, 1 \quad (16) \\ V_0(x, s) &= 0, \quad \forall x \in \mathbb{R}_+, \forall s \in \mathcal{S}, \end{aligned}$$

where $\tilde{g}_n(y, s) := h(y - d) + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1}) \mid S_n = s]$.

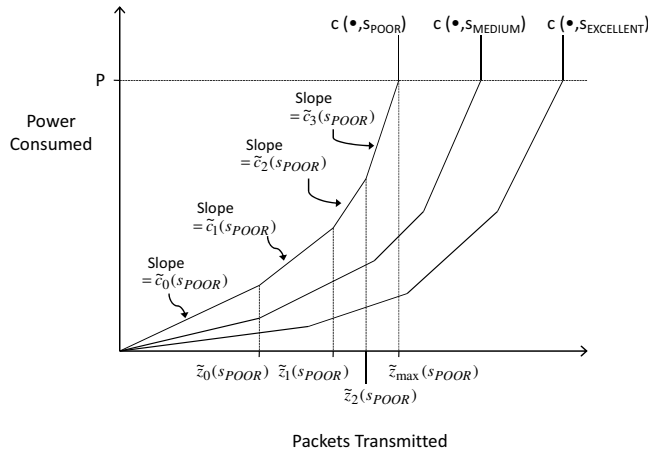


Fig. 4. A family of piecewise-linear convex power-rate functions. Like Figure 2, we incorporate the power constraint into each curve to show the effective power-rate curve. As an example, the power-rate function $c(\cdot, s_{POOR})$ is completely characterized by the sequence of slopes $\{\tilde{c}_k(s_{POOR})\}_{k \in \{0,1,2,3\}}$ and the sequence of points where the slopes change $\{\tilde{z}_k(s_{POOR})\}_{k \in \{0,1,2\}}$. The maximum number of packets that can be transmitted in a slot when the channel condition is s_{POOR} is $\tilde{z}_{\max}(s_{POOR})$.

A. Structure of Optimal Policy for the Finite Horizon Discounted Expected Cost Problem

We showed in Theorem 1 that the the optimal transmission policy to a single receiver in the case of linear power-rate curves is a modified base-stock policy characterized by a single critical level for each channel condition. In this section, we generalize this result to the case of piecewise-linear power-rate curves, and show that the optimal receiver buffer level after transmission (respectively, optimal number of packets to transmit) is no longer a three-segment piecewise-linear nondecreasing (respectively, nonincreasing) function of the starting buffer level as in Figure 3, but a more general piecewise-linear nondecreasing (respectively, nonincreasing) function.

Theorem 3. *In Problem (PI) with a single receiver under piecewise-linear convex power-rate curves, for every $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$, there exists a nonincreasing sequence of critical numbers $\{b_{n,k}(s)\}_{k \in \{0,1,\dots,K\}}$ such that the optimal number of packets to transmit with n slots*

remaining is given by:

$$z_n^*(x, s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } b_{n,k}(s) - \tilde{z}_{k-1}(s) < x \leq b_{n,k-1}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K\} \\ b_{n,k}(s) - x, & \text{if } b_{n,k}(s) - \tilde{z}_k(s) < x \leq b_{n,k}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K-1\} \\ b_{n,K}(s) - x, & \text{if } b_{n,K}(s) - \tilde{z}_{\max}(s) < x \leq b_{n,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } 0 \leq x \leq b_{n,K}(s) - \tilde{z}_{\max}(s) \end{cases}, \quad (17)$$

where $b_{n,-1}(s) := \infty$, $\forall s \in \mathcal{S}$. The optimal receiver buffer level after transmission is given by $y_n^*(x, s) = x + z_n^*(x, s)$.

The optimal transmission policy in Theorem 3 is a finite generalized base-stock policy. It can be interpreted as follows. Under each channel condition s , there is a target level or critical number associated with each segment of the associated piecewise-linear convex power-rate curve shown in Figure 4. If the starting buffer level is below the critical number associated with the first segment, $b_{n,0}(s)$, the scheduler should try to bring the buffer level as close as possible to the target, $b_{n,0}(s)$. If the maximum number of packets sent at this per packet power cost, $\tilde{z}_0(s)$, does not suffice to reach the critical number $b_{n,0}(s)$, then those $\tilde{z}_0(s)$ packets are scheduled, and the next segment of the power-rate curve is considered. This second segment has a slope of $\tilde{c}_1(s)$ and an associated critical number $b_{n,1}(s)$, which is no higher than $b_{n,0}(s)$, the first critical number. If the starting buffer level plus the $\tilde{z}_0(s)$ already-scheduled packets brings the buffer level above $b_{n,1}(s)$, then no more packets are scheduled for transmission. Otherwise, it is optimal to transmit so as to bring the buffer level as close as possible to $b_{n,1}(s)$, by transmitting up to $\tilde{z}_1(s) - \tilde{z}_0(s)$ additional packets at a cost of $\tilde{c}_1(s)$ power units per packet. This process continues with the sequential consideration of each segment of the power-rate curve. At each successive iteration, the target level is lower and the starting buffer level, updated to include already-scheduled packets, is higher. The process continues until the buffer level reaches or exceeds a critical number, or the full power P is consumed. Note that this sequential consideration is not actually done online, but

only meant to provide an intuitive explanation of the optimal policy. See Figure 5 for diagrams of the structure of the optimal finite generalized base-stock policy.

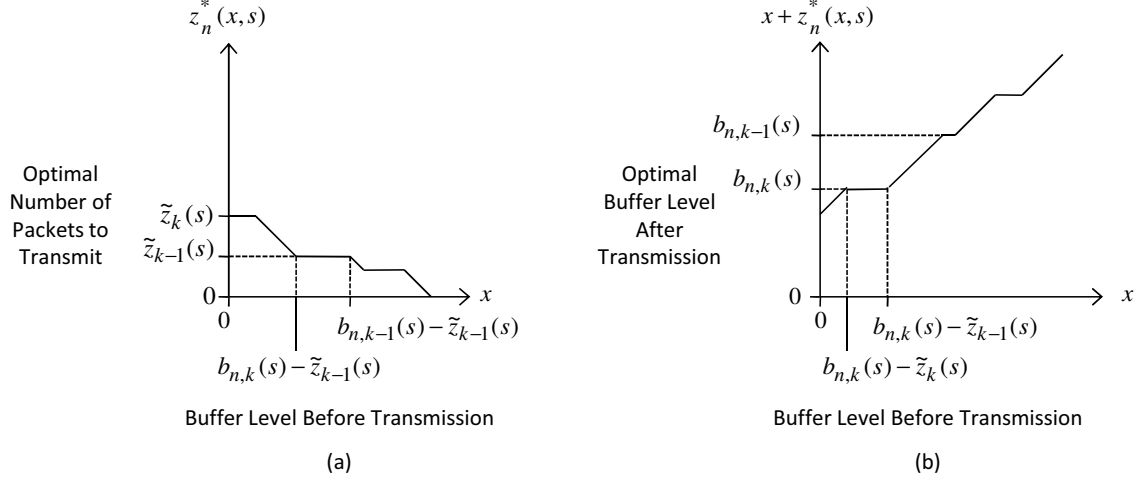


Fig. 5. Optimal transmission policy in slot n when the state is (x, s) . (a) depicts the optimal transmission quantity, and (b) depicts the resulting number of packets available for payout in slot n .

B. Computation of Critical Numbers

While finite generalized base-stock policies have been considered in the inventory literature for almost three decades, we are not aware of any previous studies that explicitly compute the critical numbers for any model where such a policy is optimal. In this section, we compute the critical numbers under each channel condition when technical conditions similar to those of Section III-B are satisfied. We consider the special case when the channel condition is independent and identically distributed from slot to slot; the holding cost function is linear (i.e., $h(x) = h \cdot x$); and the following technical condition on the power-rate functions is satisfied for each possible channel condition $s \in \mathcal{S}$: $\tilde{z}_{\max}(s) = \tilde{l}_{\max} \cdot d$ for some $\tilde{l}_{\max} \in \mathbb{N}$, and for every $k \in \{0, 1, \dots, K-1\}$, $\tilde{z}_k(s) = \tilde{l}_k \cdot d$ for some $\tilde{l}_k \in \mathbb{N}$; i.e., the slopes of the effective power-rate functions only change at integer multiples of the drainage rate d . Under these conditions, we can completely characterize the optimal transmission policy.

As in Theorem 2, we recursively define a set of thresholds, and use them to determine the critical numbers, $\{b_{n,k}(s)\}_{k \in \{-1, 0, \dots, K\}}$, for each channel condition, at each time.

Theorem 4. Define the thresholds $\tilde{\gamma}_{n,j}$ for $n \in \{1, 2, \dots, N\}$ and $j \in \mathbb{N}$ recursively, as follows:

- (i) If $j = 1$, $\tilde{\gamma}_{n,j} = \infty$;
- (ii) If $j > n$, $\tilde{\gamma}_{n,j} = 0$;
- (iii) If $2 \leq j \leq n$,

$$\tilde{\gamma}_{n,j} = -h + \alpha \cdot \left(\begin{array}{l} \sum_{s: \tilde{c}_0(s) \geq \tilde{\gamma}_{n-1,j-1}} p(s) \cdot \tilde{\gamma}_{n-1,j-1} \\ + \sum_{k=0}^{K-1} \left\{ \begin{array}{l} \sum_{s: \tilde{\gamma}_{n-1,j-1} + \tilde{L}_k(s) \leq \tilde{c}_k(s) < \tilde{\gamma}_{n-1,j-1} + \tilde{L}_{k-1}(s)} p(s) \cdot \tilde{c}_k(s) \\ + \sum_{s: \tilde{c}_k(s) < \tilde{\gamma}_{n-1,j-1} + \tilde{L}_k(s) \leq \tilde{c}_{k+1}(s)} p(s) \cdot \tilde{\gamma}_{n-1,j-1} + \tilde{L}_k(s) \end{array} \right\} \\ + \sum_{s: \tilde{\gamma}_{n-1,j-1} + \tilde{L}_{\max}(s) \leq \tilde{c}_K(s) < \tilde{\gamma}_{n-1,j-1} + \tilde{L}_{K-1}(s)} p(s) \cdot \tilde{c}_K(s) \\ + \sum_{s: \tilde{c}_K(s) < \tilde{\gamma}_{n-1,j-1} + \tilde{L}_{\max}(s)} p(s) \cdot \tilde{\gamma}_{n-1,j-1} + \tilde{L}_{\max}(s) \end{array} \right), \quad (18)$$

where $p(s)$ is the probability of the channel being in state s in a time slot, $\tilde{L}_k(s) := \frac{\tilde{z}_k(s)}{d}$ for all $s \in \mathcal{S}$ and $k \in \{0, 1, \dots, K-1\}$, and $\tilde{L}_{\max}(s) := \frac{\tilde{z}_{\max}(s)}{d}$ for all $s \in \mathcal{S}$. For each $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$, define $b_{n,-1}(s) := \infty$ and for all $k \in \{0, 1, \dots, K\}$, if $\tilde{\gamma}_{n,j+1} \leq \tilde{c}_k(s) < \tilde{\gamma}_{n,j}$, define $b_{n,k}(s) := j \cdot d$. The optimal control strategy for Problem (P1) is then given by $\boldsymbol{\pi}^* = \{z_N^*, z_{N-1}^*, \dots, z_1^*\}$, where for all $n \in \{N, N-1, \dots, 1\}$, $z_n^*(x, s)$ is given by (17).

It is straightforward to check that Theorem 4 is in fact a generalization of Theorem 2. To see this, set $K = 0$ so that the summation from $k = 0$ to $k = K - 1$ on the right-hand side of (18) drops out. Then $\tilde{\gamma}_{n,j}$ in (18) is the same as $\gamma_{n,j}$ in (10), $\tilde{c}_0(s)$ corresponds to c_s in (10), $b_{n,0}(s)$ corresponds to $b_n(s)$, $\tilde{z}_{\max}(s)$ corresponds to $\frac{P}{c_s}$, $\tilde{L}_{\max}(s)$ corresponds to $L(s)$, and $\tilde{L}_{K-1}(s) = 0$. The resulting optimal transmission policies are also the same.

In Theorem 4, the threshold $\tilde{\gamma}_{n,j}$ may again be interpreted as the per packet power cost at which, with n slots remaining in the horizon, the expected cost-to-go of transmitting packets to cover the user's playout requirements for the next $j - 1$ slots is the same as the expected cost-to-go of transmitting packets to cover the user's requirements for the next j slots. The intuition

behind the recursion (18) is similar to the detailed explanation given in Section III-B. Namely, we can start with equation (12) and expand out the right-hand side based on the known structure of the optimal policy, until, after a fair bit of algebra, the result is (18). A detailed proof of Theorem 4 is included in Appendix A.

C. Structure of the Optimal Policy for the Infinite Horizon Discounted Expected Cost Problems

In this section, we show that the optimal policy for the infinite horizon discounted expected cost problem is the natural extension of the optimal policy for the finite horizon discounted expected cost problem; namely, it is a finite generalized base-stock policy characterized by *time-invariant* sequences of critical numbers for each channel condition. These time-invariant sequences of critical numbers for the infinite horizon discounted expected cost problem are equal to the limit of the finite horizon sequences of critical numbers as the time horizon N goes to infinity.

Theorem 5.

- (a) For a fixed $x \in \mathbb{R}_+$ and $s \in \mathcal{S}$, $V_n(x, s)$ is nondecreasing in n . Moreover, $\lim_{n \rightarrow \infty} V_n(x, s)$ exists and is finite, $\forall x \in \mathbb{R}_+, \forall s \in \mathcal{S}$.
- (b) Define $V_\infty(x, s) := \lim_{n \rightarrow \infty} V_n(x, s)$. Then $V_\infty(x, s)$ is convex in x for any fixed $s \in \mathcal{S}$.
- (c) Define $\tilde{g}_\infty(y, s) := h(y - d) + \alpha \cdot \mathbb{E}[V_\infty(y - d, S') \mid S = s]$, where S' is the channel condition in the subsequent slot. Then $\tilde{g}_n(y, s)$ converges monotonically to $\tilde{g}_\infty(y, s)$, $\forall y \in [d, \infty)$, $\forall s \in \mathcal{S}$; $\tilde{g}_\infty(y, s)$ is convex in y for any fixed $s \in \mathcal{S}$; and $\lim_{y \rightarrow \infty} \tilde{g}_\infty(y, s) = \infty$, $\forall s \in \mathcal{S}$.
- (d) Define $b_{\infty, -1}(s) := \infty$ and

$$b_{\infty, k}(s) := \max \left\{ d, \inf \{ b \mid \tilde{g}'^+(b, s) \geq -\tilde{c}_k(s) \} \right\}, \quad \forall k \in \{0, 1, \dots, K\},$$

where $\tilde{g}'^+(b, s)$ represents the right derivative:

$$\tilde{g}'^+(b, s) := \lim_{y \downarrow b} \frac{\tilde{g}_\infty(y, s) - \tilde{g}_\infty(b, s)}{y - b}.$$

Then $b_{\infty, k}(s) = \lim_{n \rightarrow \infty} b_{n, k}(s)$ for all $k \in \{-1, 0, 1, \dots, K\}$.

(e) $V_\infty(x, s)$ satisfies the α -discounted cost optimality equation (α -DCOE):

$$\begin{aligned} V_\infty(x, s) &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_\infty(x + z - d, S') \mid S = s] \end{array} \right\} \\ &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ c(z, s) + \tilde{g}_\infty(x + z, s) \right\}, \quad \forall x \in \mathbb{R}_+, \forall s \in \mathcal{S}, \end{aligned} \quad (19)$$

and the minimum on the right hand side of (19) is achieved by:

$$z_\infty^*(x, s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } b_{\infty, k}(s) - \tilde{z}_{k-1}(s) < x \leq b_{\infty, k-1}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K\} \\ b_{\infty, k}(s) - x, & \text{if } b_{\infty, k}(s) - \tilde{z}_k(s) < x \leq b_{\infty, k}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K-1\} \\ b_{\infty, K}(s) - x, & \text{if } b_{\infty, K}(s) - \tilde{z}_{\max}(s) < x \leq b_{\infty, K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } 0 \leq x \leq b_{\infty, K}(s) - \tilde{z}_{\max}(s) \end{cases}$$

(f) The optimal stationary policy for Problem (P2) in the case of a single receiver with piecewise-linear convex power-rate curves is given by $\boldsymbol{\pi}_\infty^* = (z_\infty^*, z_\infty^*, \dots)$.

A detailed proof, which follows the logic conveyed in the statement of the theorem, is included in Appendix B. As a special case of Theorem 5, the optimal policy in Problem (P2) for the case discussed in Section III of a single receiver with linear power-rate curves is given by $\boldsymbol{\pi}_\infty^* = (z_\infty^*, z_\infty^*, \dots)$, where:

$$z_\infty^*(x, s) := \begin{cases} 0, & \text{if } x \geq b_\infty(s) \\ b_\infty(s) - x, & \text{if } b_\infty(s) - \frac{P}{c_s} \leq x < b_\infty(s) \\ \frac{P}{c_s}, & \text{if } x < b_\infty(s) - \frac{P}{c_s} \end{cases},$$

and $b_\infty(s) := \lim_{n \rightarrow \infty} b_n(s)$.

D. Structure of the Optimal Policy for the Infinite Horizon Average Expected Cost Problems

In this section we use the *vanishing discount approach* to show that the finite generalized base-stock structure is also optimal for the infinite horizon average expected cost problem, (P3).

We show that an optimal policy for the infinite horizon average expected cost problem exists and can be represented as the limit as the discount factor increases to one of optimal policies identified in Section IV-C for the infinite horizon discounted expected cost problem.

In Section IV-C, we suppressed the dependence of the value functions and optimal policies on the discount factor, α . Here, we make this dependence explicit by including the discount factor in the subscript labeling of the value functions and optimal policies for the infinite horizon discounted expected cost problem. For example, the value function defined in (b) of Theorem 5 is now denoted by $V_{\infty,\alpha}(x, s)$.

Theorem 6. *For all $\alpha \in [0, 1)$, define:*

$$\begin{aligned} m_{\infty,\alpha} &:= \inf_{\substack{x \in \mathbb{R}_+ \\ s \in \mathcal{S}}} V_{\infty,\alpha}(x, s), \\ \rho^* &:= \lim_{\alpha \nearrow 1} (1 - \alpha) \cdot m_{\infty,\alpha}, \text{ and} \\ w_{\infty,\alpha}(x, s) &:= V_{\infty,\alpha}(x, s) - m_{\infty,\alpha}, \quad \forall x \in \mathbb{R}_+, \quad \forall s \in \mathcal{S}. \end{aligned}$$

Then:

(a) *There exists a continuous function $w_{\infty,1}(\cdot, \cdot)$ and a selector $z_{\infty,1}^*(\cdot, \cdot)$ that satisfy the ACOE:*

$$\begin{aligned} \rho^* + w_{\infty,1}(x, s) &= \min_{\{\max(0, d-x) \leq z \leq \bar{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \mathbb{E}[w_{\infty,1}(x + z - d, S') \mid S = s] \end{array} \right\} \\ &= c(z_{\infty,1}^*(x, s), s) + h(x + z_{\infty,1}^*(x, s) - d) \\ &\quad + \mathbb{E}[w_{\infty,1}(x + z_{\infty,1}^*(x, s) - d, S') \mid S = s], \quad \forall x \in \mathbb{R}_+, \quad \forall s \in \mathcal{S}. \end{aligned}$$

(b) *The stationary policy $\pi_{\infty,1}^* = (z_{\infty,1}^*, z_{\infty,1}^*, \dots)$ is optimal for Problem (P3) in the case of a single receiver with piecewise-linear convex power-rate curves.*

(c) *The resulting optimal average cost beginning from any initial state $(x, s) \in \mathbb{R}_+ \times \mathcal{S}$ is ρ^* .*

(d) *For every increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, there exists*

a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ approaching 1 such that:

$$w_{\infty,1}(x, s) = \lim_{i \rightarrow \infty} w_{\infty, \alpha(l_i)}(x, s), \quad \forall x \in \mathbb{R}_+, \quad \forall s \in \mathcal{S}.$$

Therefore, for every $s \in \mathcal{S}$, $w_{\infty,1}(x, s)$ is convex in x .

- (e) For every $(x, s) \in \mathbb{R}_+ \times \mathcal{S}$ and increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ approaching 1 and a sequence $\{x(i)\}_{i=1,2,\dots}$ approaching x such that:

$$z_{\infty,1}^*(x, s) = \lim_{i \rightarrow \infty} z_{\infty, \alpha(l_i)}^*(x(i), s).$$

- (f) A stationary finite generalized base-stock policy is average cost optimal in the case of piecewise-linear convex power-rate curves, and a stationary modified base-stock policy is average cost optimal in the case of linear power-rate curves.

Thus, the structure of the optimal policy is the same for all three problems, **(P1)**, **(P2)**, and **(P3)**. The proof of Theorem 6 is discussed in Appendix C.

E. General Convex Power-Rate Curves

As mentioned in Section II-A, in general, the power-rate curve under each possible channel condition is convex. It can be shown that under convex power-rate curves at each time, the optimal number of packets to send is a nonincreasing function of the starting buffer level. However, without any further structure on the power-rate curves, it is not computationally tractable to compute such optimal policies, known as generalized base-stock policies (a superclass of the *finite* generalized base-stock policies discussed above). This is why we have chosen to analyze piecewise-linear convex power-rate curves, which can be used to approximate general convex power-rate curves. More specifically, our analysis suggests approximating the general convex power-rate curves by piecewise-linear convex power-rate curves where the slopes change at integer multiples of the demand d , in order to be able to apply Theorem 4 to compute the critical numbers in an extremely efficient manner. Doing so represents an approximation at the

modeling stage followed by an exact solution, as compared to modeling the power-rate curves as more general convex functions and having to approximate the solution. Finally, we note that increasing the number of segments used to model the piecewise-linear convex functions leads to a better approximation, but comes at the cost of some extra complexity in implementing the optimal policy, as the scheduler needs to store at least one critical number for each segment of each power-rate curve.

V. TWO RECEIVERS WITH LINEAR POWER-RATE CURVES

In this section, we analyze the finite and infinite horizon discounted expected cost problems when there are two receivers ($M = 2$), and the power-rate functions under different channel conditions are linear for each user. Each user m 's channel condition evolves as a homogeneous Markov process, $\{S_n^m\}_{n=N, N-1, \dots, 1}$. As discussed in Sections I and II, the time-varying channel conditions of the two users are independent of each other, and the transmission scheduler can exploit this spatial diversity. Like Section III, we denote the power consumption per unit of data transmitted to receiver m under channel condition s^m by c_s^m . The row vector of these per unit power consumptions is given by \mathbf{c}_s^T , so that the total power consumption in slot n is given by $\sum_{m=1}^2 c^m(Z_n^m, S_n^m) = \mathbf{c}_s^T \mathbf{Z}_n$. We denote the total holding costs $\sum_{m=1}^2 h^m(X_n^m + Z_n^m - d^m)$ by $h(\mathbf{X}_n + \mathbf{Z}_n - \mathbf{d})$.

With these notations, the dynamic program (1) for Problem **(P1)** becomes:

$$V_n(\mathbf{x}, \mathbf{s}) = \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T \mathbf{z} + h(\mathbf{x} + \mathbf{z} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_{n-1}(\mathbf{x} + \mathbf{z} - \mathbf{d}, \mathbf{S}_{n-1}) \mid \mathbf{S}_n = \mathbf{s}] \end{array} \right\} \quad (20)$$

$$= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T [\mathbf{y} - \mathbf{x}] + h(\mathbf{y} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_{n-1}(\mathbf{y} - \mathbf{d}, \mathbf{S}_{n-1}) \mid \mathbf{S}_n = \mathbf{s}] \end{array} \right\} \quad (21)$$

$$= -\mathbf{c}_s^T \mathbf{x} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ G_n(\mathbf{y}, \mathbf{s}) \right\} \quad n = N, N-1, \dots, 1,$$

$$V_0(\mathbf{x}, \mathbf{s}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S} := \mathcal{S}^1 \times \mathcal{S}^2,$$

where

$$G_n(\mathbf{y}, \mathbf{s}) := \mathbf{c}_s^T \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}[V_{n-1}(\mathbf{y} - \mathbf{d}, \mathbf{S}_{n-1}) \mid \mathbf{S}_n = \mathbf{s}],$$

$$\forall \mathbf{y} \in [d^1, \infty) \times [d^2, \infty), \forall \mathbf{s} \in \mathcal{S}, \text{ and}$$

$$\tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s}) := \left\{ \mathbf{y} \in \mathbb{R}_+^2 : \mathbf{y} \succeq \mathbf{d} \vee \mathbf{x} \text{ and } \mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] \leq P \right\}, \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S}. \quad (22)$$

The transition from (20) to (21) follows again from a change of variable in the action space from \mathbf{Z}_n to \mathbf{Y}_n , where $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}_n$. The controlled random vector \mathbf{Y}_n represents the queue lengths of the receiver buffers *after* transmission takes place in the n^{th} slot, but *before* playout takes place (i.e., before d^m packets are removed from user m 's buffer). The restrictions on the action space, $\mathbf{y} \succeq \mathbf{d} \vee \mathbf{x}$ and $\mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] \leq P$, ensure: (i) a nonnegative number of packets is transmitted to each user; (ii) there are at least d^m packets in user m 's receiver buffer following transmission, in order to satisfy the underflow constraint; and (iii) the power constraint is satisfied.

Without the per slot peak power constraint, this M -dimensional problem would be separable, and could be solved by solving M instances of the one-dimensional problem of Section III; however, the joint power constraint couples the queues.⁹ As a result, the optimal transmission quantity to one receiver depends on the other receivers' queue length, as the following example shows.

Example 1. *Assume receiver 1's channel is currently in a "poor" condition, receiver 2's channel is currently in a "medium" condition, and receiver 2's buffer contains enough packets to satisfy the demand for the next few slots. We consider two different scenarios for receiver 1's buffer level to show how the optimal transmission quantity to receiver 2 depends on receiver 1's buffer level. In Scenario 1, receiver 1's buffer already contains many packets. In this scenario, it may be beneficial for the scheduler to wait for receiver 2 to have a better channel condition, because it will be able to take full advantage of an "excellent" condition when it comes. In Scenario 2, receiver 1's queue only contains enough packets for playout in the current slot. It may be*

⁹This problem therefore falls into the class of *weakly coupled stochastic dynamic programs* [65], [66].

optimal to transmit some packets to receiver 2 in the current slot in this scenario. To see this, note that even if receiver 2 experiences the best possible channel condition in the next slot, the scheduler will need to allocate some power to receiver 1 in order to prevent receiver 1's buffer from emptying. Therefore, the scheduler anticipates not being able to take full advantage of receiver 2's "excellent" condition in the next slot, and may compensate by sending some packets in the current slot under the "medium" condition.

A. Structure of Optimal Policy for the Finite Horizon Discounted Expected Cost Problem

Before proceeding to the structure of the optimal transmission policy, we state some key properties of the value functions in the following theorem.

Theorem 7. *With two receivers and linear power-rate curves, the following statements are true for $n = 1, 2, \dots, N$, and for all $\mathbf{s} \in \mathcal{S}$:*

- (i) $V_{n-1}(\mathbf{x}, \mathbf{s})$ is convex in \mathbf{x} .
- (ii) $V_{n-1}(\mathbf{x}, \mathbf{s})$ is supermodular in \mathbf{x} ; i.e., for all $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in \mathbb{R}_+^2$,

$$V_{n-1}(\bar{\mathbf{x}}, \mathbf{s}) + V_{n-1}(\tilde{\mathbf{x}}, \mathbf{s}) \leq V_{n-1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + V_{n-1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) .$$

- (iii) $G_n(\mathbf{y}, \mathbf{s})$ is convex in \mathbf{y} .
- (iv) $G_n(\mathbf{y}, \mathbf{s})$ is supermodular in \mathbf{y} ; i.e., for all $\bar{\mathbf{y}}, \tilde{\mathbf{y}} \in [d^1, \infty) \times [d^2, \infty)$,

$$G_n(\bar{\mathbf{y}}, \mathbf{s}) + G_n(\tilde{\mathbf{y}}, \mathbf{s}) \leq G_n(\bar{\mathbf{y}} \wedge \tilde{\mathbf{y}}, \mathbf{s}) + G_n(\bar{\mathbf{y}} \vee \tilde{\mathbf{y}}, \mathbf{s}) .$$

- (v) $y_n^1 < \hat{y}_n^1$ implies:

$$\inf \left\{ \operatorname{argmin}_{y_n^2 \in [d^2, \infty)} \left\{ G_n(y_n^1, y_n^2, s^1, s^2) \right\} \right\} \geq \inf \left\{ \operatorname{argmin}_{y_n^2 \in [d^2, \infty)} \left\{ G_n(\hat{y}_n^1, y_n^2, s^1, s^2) \right\} \right\}$$

and $y_n^2 < \hat{y}_n^2$ implies:

$$\inf \left\{ \operatorname{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n(y_n^1, y_n^2, s^1, s^2) \right\} \right\} \geq \inf \left\{ \operatorname{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n(y_n^1, \hat{y}_n^2, s^1, s^2) \right\} \right\} .$$

A detailed proof is included in Appendix A. Because $-\mathbf{c}_s^T \mathbf{x}$ is supermodular in \mathbf{x} , the key part of the induction step in the proof of (ii) is to show that $\min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \{G_{n-1}(\mathbf{y}, \mathbf{s})\}$ is also supermodular in \mathbf{x} . Denoting $\operatorname{argmin}_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \{G_{n-1}(\mathbf{y}, \mathbf{s})\}$ by $\mathbf{y}^*(\mathbf{x}, \mathbf{s})$, we do this constructively by showing that for all $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in \mathbb{R}_+^2$:

$$\begin{aligned}
& \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})} \{G_{n-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{n-1}(\mathbf{y}, \mathbf{s})\} \\
& \leq G_{n-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{n-1}(\tilde{\mathbf{y}}, \mathbf{s}) \\
& \leq G_{n-1}(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) + G_{n-1}(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \tag{23} \\
& = \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\},
\end{aligned}$$

for a specific choice of $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$ and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})$. The difficulty is cleverly constructing $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$, depending on the relative locations of $\bar{\mathbf{x}}, \tilde{\mathbf{x}}, \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})$, and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}})$, so as to ensure (23) is true.

It follows from Theorem 7 that the structure of the optimal transmission policy for the finite horizon discounted expected cost problem is given by the following theorem.

Theorem 8. *For every $n \in \{1, 2, \dots, N\}$ and $\mathbf{s} \in \mathcal{S}^1 \times \mathcal{S}^2$, define the nonempty set of global minimizers of $G_n(\cdot, \mathbf{s})$:*

$$\mathcal{B}_n(\mathbf{s}) := \left\{ \hat{\mathbf{y}} \in [d^1, \infty) \times [d^2, \infty) : G_n(\hat{\mathbf{y}}, \mathbf{s}) = \min_{\mathbf{y} \in [d^1, \infty) \times [d^2, \infty)} G_n(\mathbf{y}, \mathbf{s}) \right\}.$$

Define also

$$b_n^1(\mathbf{s}) := \min \left\{ y^1 \in [d^1, \infty) : (y^1, y^2) \in \mathcal{B}_n(\mathbf{s}) \text{ for some } y^2 \in [d^2, \infty) \right\},$$

and

$$b_n^2(\mathbf{s}) := \min \left\{ y^2 \in [d^2, \infty) : (b_n^1(\mathbf{s}), y^2) \in \mathcal{B}_n(\mathbf{s}) \right\}.$$

Then the vector $\mathbf{b}_n(\mathbf{s}) = (b_n^1(\mathbf{s}), b_n^2(\mathbf{s})) \in \mathcal{B}_n(\mathbf{s})$ is a global minimizer of $G_n(\cdot, \mathbf{s})$. Define also

the functions:

$$f_n^1(x^2, \mathbf{s}) := \inf \left\{ \operatorname{argmin}_{y^1 \in [d^1, \infty)} \left\{ G_n(y^1, x^2, s^1, s^2) \right\} \right\}, \text{ for } x^2 \in [d^2, \infty), \text{ and}$$

$$f_n^2(x^1, \mathbf{s}) := \inf \left\{ \operatorname{argmin}_{y^2 \in [d^2, \infty)} \left\{ G_n(x^1, y^2, s^1, s^2) \right\} \right\}, \text{ for } x^1 \in [d^1, \infty).$$

Note that by construction, $f_n^1(b_n^2(\mathbf{s}), \mathbf{s}) = b_n^1(\mathbf{s})$ and $f_n^2(b_n^1(\mathbf{s}), \mathbf{s}) = b_n^2(\mathbf{s})$. Partition \mathbb{R}_+^2 into the following seven regions:

$$\begin{aligned} \mathcal{R}_I(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : \mathbf{x} \succeq (f_n^1(x^2, \mathbf{s}), f_n^2(x^1, \mathbf{s})) \text{ and } \mathbf{x} \neq \mathbf{b}_n(\mathbf{s}) \right\} \\ \mathcal{R}_{II}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : \mathbf{x} \preceq \mathbf{b}_n(\mathbf{s}) \text{ and } \mathbf{c}_s^\top [\mathbf{b}_n(\mathbf{s}) - \mathbf{x}] \leq P \right\} \\ \mathcal{R}_{III-A}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : x^2 > b_n^2(\mathbf{s}) \text{ and } f_n^1(x^2, \mathbf{s}) - \frac{P}{c_{s^1}} \leq x^1 < f_n^1(x^2, \mathbf{s}) \right\} \\ \mathcal{R}_{III-B}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : x^1 > b_n^1(\mathbf{s}) \text{ and } f_n^2(x^1, \mathbf{s}) - \frac{P}{c_{s^2}} \leq x^2 < f_n^2(x^1, \mathbf{s}) \right\} \\ \mathcal{R}_{IV-A}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : x^2 > b_n^2(\mathbf{s}) \text{ and } x^1 < f_n^1(x^2, \mathbf{s}) - \frac{P}{c_{s^1}} \right\} \\ \mathcal{R}_{IV-B}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : \mathbf{x} \preceq \mathbf{b}_n(\mathbf{s}) \text{ and } \mathbf{c}_s^\top [\mathbf{b}_n(\mathbf{s}) - \mathbf{x}] > P \right\} \\ \mathcal{R}_{IV-C}(n, \mathbf{s}) &:= \left\{ \mathbf{x} \in \mathbb{R}_+^2 : x^1 > b_n^1(\mathbf{s}) \text{ and } x^2 < f_n^2(x^1, \mathbf{s}) - \frac{P}{c_{s^2}} \right\}, \end{aligned}$$

and define $\mathcal{R}_{IV}(n, \mathbf{s}) := \mathcal{R}_{IV-A}(n, \mathbf{s}) \cup \mathcal{R}_{IV-B}(n, \mathbf{s}) \cup \mathcal{R}_{IV-C}(n, \mathbf{s})$.

Then for Problem (P1) in the case of two receivers with linear power-rate curves, for all $\mathbf{x} \notin \mathcal{R}_{IV}(n, \mathbf{s})$, an optimal control action with n slots remaining is given by:

$$\mathbf{y}_n^*(\mathbf{x}, \mathbf{s}) := \begin{cases} \mathbf{x}, & \text{if } \mathbf{x} \in \mathcal{R}_I(n, \mathbf{s}) \\ \mathbf{b}_n(\mathbf{s}), & \text{if } \mathbf{x} \in \mathcal{R}_{II}(n, \mathbf{s}) \\ (f_n^1(x^2, \mathbf{s}), x^2), & \text{if } \mathbf{x} \in \mathcal{R}_{III-A}(n, \mathbf{s}) \\ (x^1, f_n^2(x^1, \mathbf{s})), & \text{if } \mathbf{x} \in \mathcal{R}_{III-B}(n, \mathbf{s}) \end{cases}. \quad (24)$$

For all $\mathbf{x} \in \mathcal{R}_{IV}(n, \mathbf{s})$, there exists an optimal control action with n slots remaining, $\mathbf{y}_n^*(\mathbf{x}, \mathbf{s})$,

which satisfies:

$$\mathbf{c}_s^T [\mathbf{y}_n^*(\mathbf{x}, \mathbf{s}) - \mathbf{x}] = P. \quad (25)$$

A detailed proof is included in Appendix A. Equation (25) says that it is optimal for the transmitter to allocate the full power budget for transmission when the vector of receiver buffer levels at the beginning of slot n falls in region $\mathcal{R}_{IV}(n, \mathbf{s})$. We cannot say anything in general about the optimal allocation (split) of the full power budget between the two receivers when the starting buffer levels lie in region $\mathcal{R}_{IV}(n, \mathbf{s})$. Figure 6 shows the partition of \mathbb{R}_+^2 into the seven regions, and a diagram of the structure of the optimal transmission policy. Note that the figure shows the seven regions of the optimal policy for a *fixed realization* of the pair of channel conditions. Under different pairs of channel realizations, the seven regions have the same general form, but the targets $\mathbf{b}_n(\mathbf{s})$ are shifted and the boundary functions $f_n^1(x^2, \mathbf{s})$ and $f_n^2(x^1, \mathbf{s})$ are different.

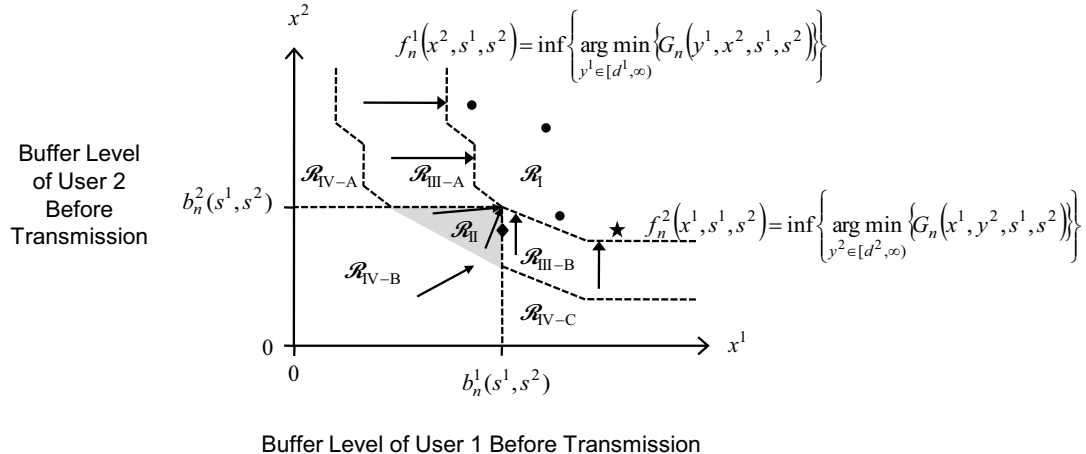


Fig. 6. Optimal transmission policy for the two receiver case in slot n when the state is (\mathbf{x}, \mathbf{s}) . The seven regions described in Theorem 8 are labeled. The tails of the arrows represent the vectors of the receiver buffer levels at the beginning of slot n , and the heads of the arrows represent the vectors of the receiver buffer levels after transmission but before playout in slot n under the optimal transmission policy. In region $\mathcal{R}_I(n, \mathbf{s})$, a single dot represents that it is optimal to not transmit any packets to either user. The \star and \blacklozenge represent possible starting buffer levels for Scenarios 1 and 2, respectively, in Example 1.

In some sense, the structure of the optimal policy outlined in Theorem 8 can be interpreted as an extension of the modified base-stock policy for the case of a single receiver outlined in

Theorem 1. Namely, under each channel condition at each time, there is a critical number for each receiver ($b_n^m(\mathbf{s})$) such that it is optimal to bring both receivers' buffer levels up to those critical numbers if it is possible to do so (region $\mathcal{R}_{II}(n, \mathbf{s})$), and it is optimal to not transmit any packets if both receivers' buffer levels start beyond their critical numbers (region $\mathcal{R}_I(n, \mathbf{s})$). However, this extended notion of the modified base-stock policy only captures the optimal behavior in two of the seven regions, and does not account for the coupling behavior between users that arises through the joint power constraint. For instance, possible starting buffer levels for Scenario 1 and Scenario 2 in Example 1 are illustrated in Figure 6 by the \star and \blacklozenge , respectively. Even though the buffer level of receiver 2 before transmission is the same under both scenarios, the optimal transmission quantity to receiver 2 is different under the two scenarios due to the different starting buffer levels of receiver 1.

B. Structure of the Optimal Policy for the Infinite Horizon Discounted Expected Cost Problems

In this section, we show that the structure of the optimal *stationary (or time-invariant)* policy for the infinite horizon discounted expected cost problem is the same as the structure of the optimal policy for the finite horizon discounted expected cost problem. Moreover, the boundaries of the seven regions of the finite horizon optimal policy shown in Figure 6 converge to the boundaries of the seven regions of the infinite horizon discounted expected cost optimal policy as the time horizon N goes to infinity.

Theorem 9. Define:

- (i) $V_\infty(\mathbf{x}, \mathbf{s}) := \lim_{n \rightarrow \infty} V_n(\mathbf{x}, \mathbf{s})$, for all $\mathbf{x} \in \mathbb{R}_+^2$ and $\mathbf{s} \in \mathcal{S}$ (this limit exists).
- (ii) $G_\infty(\mathbf{y}, \mathbf{s}) := \mathbf{c}_s^\top \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}[V_\infty(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}]$, for all $\mathbf{y} \in [d^1, \infty) \times [d^2, \infty)$ and $\mathbf{s} \in \mathcal{S}$.
- (iii) $\mathcal{B}_\infty(\mathbf{s}) := \left\{ \hat{\mathbf{y}} \in [d^1, \infty) \times [d^2, \infty) : G_\infty(\hat{\mathbf{y}}, \mathbf{s}) = \min_{\mathbf{y} \in [d^1, \infty) \times [d^2, \infty)} G_\infty(\mathbf{y}, \mathbf{s}) \right\}$.
- (iv) $b_\infty^1(\mathbf{s}) := \min \left\{ y^1 \in [d^1, \infty) : (y^1, y^2) \in \mathcal{B}_\infty(\mathbf{s}) \text{ for some } y^2 \in [d^2, \infty) \right\}$.
- (v) $b_\infty^2(\mathbf{s}) := \min \left\{ y^2 \in [d^2, \infty) : (b_\infty^1(\mathbf{s}), y^2) \in \mathcal{B}_\infty(\mathbf{s}) \right\}$.
- (vi) $\mathbf{b}_\infty(\mathbf{s}) := (b_\infty^1(\mathbf{s}), b_\infty^2(\mathbf{s}))$.

(vii) *The functions*

$$f_{\infty}^1(x^2, \mathbf{s}) := \inf \left\{ \operatorname{argmin}_{y^1 \in [d^1, \infty)} \left\{ G_{\infty}(y^1, x^2, s^1, s^2) \right\} \right\}, \text{ for } x^2 \in [d^2, \infty), \text{ and}$$

$$f_{\infty}^2(x^1, \mathbf{s}) := \inf \left\{ \operatorname{argmin}_{y^2 \in [d^2, \infty)} \left\{ G_{\infty}(x^1, y^2, s^1, s^2) \right\} \right\}, \text{ for } x^1 \in [d^1, \infty).$$

(viii) *The seven regions $\mathcal{R}_I(\infty, \mathbf{s}) - \mathcal{R}_{IV-C}(\infty, \mathbf{s})$, defined in the same way as in Theorem 8, with n replaced by ∞ .*

Then

(a) $V_{\infty}(\mathbf{x}, \mathbf{s})$ satisfies the α -discounted optimality equation (α -DCOE):

$$V_{\infty}(\mathbf{x}, \mathbf{s}) = \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T [\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_{\infty}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \end{array} \right\}, \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S}. \quad (26)$$

(b) *An optimal stationary policy for Problem (P2) in the case of two receivers with linear power-rate curves is given by $\boldsymbol{\pi}_{\infty}^* = (\mathbf{y}_{\infty}^*, \mathbf{y}_{\infty}^*, \dots)$, where*

$$\mathbf{y}_{\infty}^*(\mathbf{x}, \mathbf{s}) := \begin{cases} \mathbf{x}, & \text{if } \mathbf{x} \in \mathcal{R}_I(\infty, \mathbf{s}) \\ \mathbf{b}_{\infty}(\mathbf{s}), & \text{if } \mathbf{x} \in \mathcal{R}_{II}(\infty, \mathbf{s}) \\ (f_{\infty}^1(x^2, \mathbf{s}), x^2), & \text{if } \mathbf{x} \in \mathcal{R}_{III-A}(\infty, \mathbf{s}) \\ (x^1, f_{\infty}^2(x^1, \mathbf{s})), & \text{if } \mathbf{x} \in \mathcal{R}_{III-B}(\infty, \mathbf{s}) \end{cases},$$

and for all $\mathbf{x} \in \mathcal{R}_{IV}(\infty, \mathbf{s})$, there exists an optimal control action, $\mathbf{y}_{\infty}^(\mathbf{x}, \mathbf{s})$, which satisfies:*

$$\mathbf{c}_s^T [\mathbf{y}_{\infty}^*(\mathbf{x}, \mathbf{s}) - \mathbf{x}] = P.$$

- (c) $\lim_{n \rightarrow \infty} \mathbf{b}_n(\mathbf{s}) = \mathbf{b}_{\infty}(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}$.
- (d) $\lim_{n \rightarrow \infty} f_n^1(x^2, \mathbf{s}) = f_{\infty}^1(x^2, \mathbf{s})$ for all $x^2 \in [d^2, \infty)$ and $\mathbf{s} \in \mathcal{S}$.
- (e) $\lim_{n \rightarrow \infty} f_n^2(x^1, \mathbf{s}) = f_{\infty}^2(x^1, \mathbf{s})$ for all $x^1 \in [d^1, \infty)$ and $\mathbf{s} \in \mathcal{S}$.

A detailed proof of Theorem 9 is included in Appendix B.

C. Structure of the Optimal Policy for the Infinite Horizon Average Expected Cost Problems

In this section, we again use the vanishing discount approach to show that the structure of the optimal policy for the finite horizon expected cost and infinite horizon discounted expected cost problems extends to the infinite horizon average expected cost problem. As in Section IV-D, we make explicit the dependence of the value functions and optimal policies from the corresponding infinite horizon discounted expected cost problem on the discount factor, α .

Theorem 10. *For all $\alpha \in [0, 1)$, define:*

$$m_{\infty, \alpha} := \inf_{\substack{\mathbf{x} \in \mathbb{R}_+^2 \\ \mathbf{s} \in \mathcal{S}}} V_{\infty, \alpha}(\mathbf{x}, \mathbf{s}), \quad (27)$$

$$\rho^* := \lim_{\alpha \nearrow 1} (1 - \alpha) \cdot m_{\infty, \alpha}, \text{ and} \quad (28)$$

$$w_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) := V_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) - m_{\infty, \alpha}, \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \quad \forall \mathbf{s} \in \mathcal{S}. \quad (29)$$

Then:

(a) *There exists a continuous function $w_{\infty, 1}(\cdot, \cdot)$ and a selector $\mathbf{y}_{\infty, 1}^*(\cdot, \cdot)$ that satisfy the ACOE:*

$$\begin{aligned} \rho^* + w_{\infty, 1}(\mathbf{x}, \mathbf{s}) &= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^\top [\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) \\ + \mathbb{E}[w_{\infty, 1}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \end{array} \right\} \\ &= \mathbf{c}_s^\top [\mathbf{y}_{\infty, 1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{x}] + \mathbf{h}(\mathbf{y}_{\infty, 1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{d}) \\ &\quad + \mathbb{E}[w_{\infty, 1}(\mathbf{y}_{\infty, 1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}], \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \quad \forall \mathbf{s} \in \mathcal{S}. \end{aligned} \quad (30)$$

(b) *The stationary policy $\boldsymbol{\pi}_{\infty, 1}^* = (\mathbf{y}_{\infty, 1}^*, \mathbf{y}_{\infty, 1}^*, \dots)$ is optimal for Problem (P3) in the case of two receivers with linear power-rate curves.*

(c) *The resulting optimal average cost beginning from any initial state $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}_+^2 \times \mathcal{S}$ is ρ^* .*

(d) *For every increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ approaching 1 such that:*

$$w_{\infty, 1}(\mathbf{x}, \mathbf{s}) = \lim_{i \rightarrow \infty} w_{\infty, \alpha(l_i)}(\mathbf{x}, \mathbf{s}), \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \quad \forall \mathbf{s} \in \mathcal{S}.$$

Therefore, for every $\mathbf{s} \in \mathcal{S}$, $w_{\infty,1}(\mathbf{x}, \mathbf{s})$ is convex and supermodular in \mathbf{x} .

- (e) For every $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}_+^2 \times \mathcal{S}$ and increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ approaching 1 and a sequence $\{\mathbf{x}(i)\}_{i=1,2,\dots}$ approaching \mathbf{x} such that:

$$\mathbf{y}_{\infty,1}^*(\mathbf{x}, \mathbf{s}) = \lim_{i \rightarrow \infty} \mathbf{y}_{\infty, \alpha(l_i)}^*(\mathbf{x}(i), \mathbf{s}) .$$

- (f) There exists an optimal stationary policy with the same structure as statement (b) in Theorem 9.

A detailed proof of Theorem 10 is included in Appendix C.

D. Discussion

At first glance, the structure of the optimal policy described in Theorem 8 may also seem analogous to the structure of the optimal policy for the two-item resource-constrained inventory problem with *deterministic* prices and *stochastic* demands (i.e., the reverse of our problem), originally studied by Evans in [60], and revisited in [61]-[63]. The structure of the optimal control action at each time for that problem can also be described in terms of seven regions that look essentially the same as those shown in Figure 6.¹⁰ However, there are two fundamental differences that distinguish these two problems.

First, the function $\tilde{G}_n(\cdot)$ in the deterministic price and stochastic demand inventory problem that corresponds to our function $G_n(\cdot, \mathbf{s})$ has an additional structural property that Chen calls *μ -difference monotone* [62]. This property is equivalent to the function $\tilde{G}_n(\cdot)$ not only being supermodular, but also submodular with respect to a partial order introduced by Antoniadou in [67], [68] called the *direct value order* (see [69] for further details). This functional property leads to two additional structural results on the optimal control action: (i) when the initial vector of inventories (corresponds to the vector of receivers' buffer levels in our problem) is in region $\mathcal{R}_{IV-B}(n)$, there exists an optimal control action such that $\mathbf{y}_n^*(\mathbf{x}) \preceq \mathbf{b}_n$; and (ii) when the initial

¹⁰In the case of deterministic prices and stochastic demands, the boundaries of the regions do not depend on the ordering price (corresponding to the channel conditions \mathbf{s} in our case), because the vector of ordering prices is deterministic.

vector of inventories is in region $\mathcal{R}_{IV-A}(n)$ (respectively, $\mathcal{R}_{IV-C}(n)$), there exists an optimal control action that includes not ordering any of item 2 (respectively, item 1), corresponding to not transmitting any packets to user 2 (respectively, user 1) in our problem. Due to the time-varying channel conditions, this property does not hold for our function $G_n(\cdot, \mathbf{s})$, and these two additional statements on the structure of the optimal policy are not true in general for our problem, as shown by the following example.

Example 2. Consider a single sender transmitting to two statistically identical receivers, whose channel conditions are IID over time and independent of each other. The power-rate curves are linear, and the possible per packet power costs are 1.750 (best possible channel condition), 2.000, 2.001, and 2.100 (worst possible channel condition). The associated probabilities of each user experiencing these channel conditions are 0.4, 0.4, 0.1, and 0.1, respectively. The total power constraint in each slot is $P = 4.2$, and 1 packet is removed from each receiver's buffer at the end of each time slot (i.e., $\mathbf{d} = (1, 1)$). We consider a finite horizon problem with the discount rate $\alpha = 1$, and no holding costs. We are interested in the optimal control action with $T = 3$ time slots remaining, and the current channel conditions are such that it costs 2.000 units of power to transmit a packet to user 1, and 2.001 units of power to transmit a packet to user 2.

Exactly solving the dynamic program shows that the unique global minimizer of the function $G_3(\cdot, \cdot, \mathbf{s}_3)$ is the vector $(\frac{101}{75}, \frac{101}{75})$. However, if the vector of starting receiver buffer levels at time $T = 3$ is $\mathbf{x}_3 = (0.2, 0.2)$, the unique optimal scheduling decision in the slot is to transmit 0.8 packets to user 2, and use the remaining power for transmission to user 1, which results in 1.2996 packets being sent to user 1. A diagram of this optimal control action is shown in Figure 7. The interesting thing to note here is that despite being power-constrained (the vector of starting buffer levels is in Region \mathcal{R}_{IV-B}), the unique optimal scheduling decision calls for filling user 1's buffer beyond its critical number $b_3^1(\mathbf{s}_3) = \frac{101}{75}$. That is, the optimal scheduling decision brings the buffer levels from Region \mathcal{R}_{IV-B} to Region \mathcal{R}_{III-B} rather than Region \mathcal{R}_{II} .

The second fundamental difference is also a consequence of the time-varying channel con-

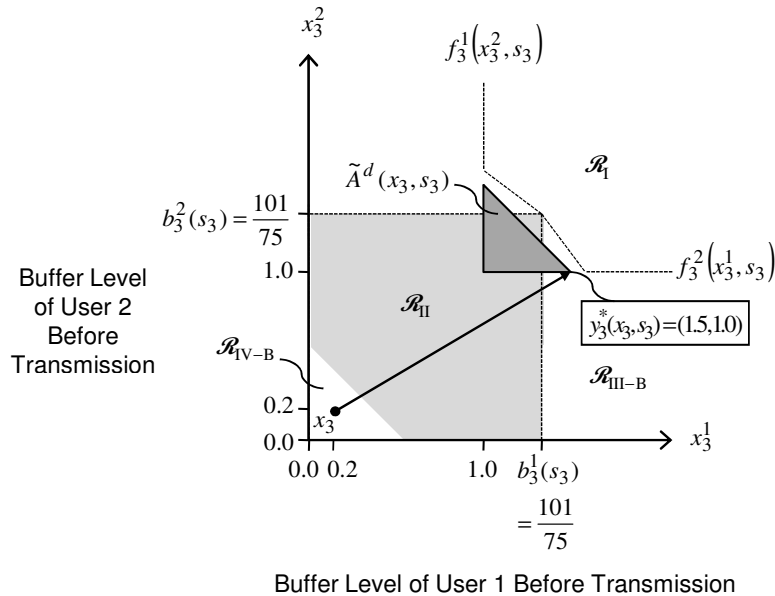


Fig. 7. Optimal scheduling decision with 3 slots remaining in Example 2. The action space is represented by the triangle $\tilde{A}^d(\mathbf{x}_3, \mathbf{s}_3)$. The critical vector $\mathbf{b}_3(\mathbf{s}_3)$ is not reachable from the starting buffer levels $\mathbf{x}_3 = (0.2, 0.2)$. The unique optimal control action is to choose $\mathbf{y}_3(\mathbf{x}_3, \mathbf{s}_3)$ (the buffer levels after transmission but before playout) to be $(1.5, 1.0)$. The interesting feature of the example is that even though $\mathbf{x}_3 \preceq \mathbf{b}_3(\mathbf{s}_3)$, we have $\mathbf{y}_3^*(\mathbf{x}_3, \mathbf{s}_3) \not\preceq \mathbf{b}_3(\mathbf{s}_3)$.

ditions in our model. In the infinite horizon version of the two-item inventory problem with deterministic prices and stochastic demands, the critical numbers are time-invariant. Combined with the above property that it is optimal to not order inventory so as to move out of regions \mathcal{R}_{II} and \mathcal{R}_{IV-B} , the time-invariant critical numbers mean that the region $\mathcal{R}_{II} \cup \mathcal{R}_{IV-B}$ (i.e., the lower-left square below the critical vector) is a “stability” region. Eventually, the vector of inventories enters this region under the optimal ordering policy, and once it does, it never leaves. This behavior both simplifies the analysis and opens the door for new mathematical techniques, such as analyzing *shortfall* to compute the critical numbers [54], [63]. In our Problems **(P2)** and **(P3)**, even though the boundaries of the seven regions for each possible channel condition are time-invariant, no such stability region exists, because the critical numbers vary over time due to the time-varying channel conditions. This makes it significantly more difficult to determine optimal and near-optimal policies.

VI. EXTENSIONS

In this section, we discuss the relaxation of the strict underflow constraints and the extension to the general case of M receivers.

A. Relaxation of the Strict Underflow Constraints

In some applications, it may not be the case that the peak power per slot is always sufficient to transmit one slot's worth of packets to each receiver, even under the worst channel conditions. In this case, a more appropriate model is to relax the strict underflow constraints, and allow underflow at a cost. One way to model this situation is to allow the receivers' queues to be negative, with a negative buffer level representing the number of packets that the playout process is behind. Then, in addition to the holding costs assessed on positive buffer levels, shortage costs are assessed on negative buffer levels. With some minor alterations to the proofs, it is straightforward to show that as long as the shortage cost function is a convex function of the negative buffer level, the structural results of Theorems 1, 3 and 8 are essentially unchanged by the relaxation of the strict underflow constraints to loose underflow constraints with penalties on underflow. This is not too surprising as the strict underflow constraint case we consider can be thought of as the limiting case as the penalties on underflow go to infinity.¹¹

B. Extension to the General Case of M Receivers

Our ongoing work includes examining the extension to the most general case of M receivers. It is unlikely that the structure of the optimal policy in this case has a simple, intuitive, and implementable form. Therefore, our approach is to find lower bounds on the value function and a feasible policy whose expected cost is as close as possible to these bounds. One simple lower bound to the value function can be found by relaxing the per slot peak power constraint of P units of total power allocated to all users, and allowing up to P units of power to be allocated

¹¹Tracking the number of packets that the playout process is behind in this manner corresponds to the *complete backlogging* assumption in inventory theory. An alternate model is to say that a packet is of no use once it misses its deadline, penalize missed packets, and keep the receiver queue length at zero. This model corresponds to the *lost sales* assumption in inventory theory.

to *each* receiver in a single slot (for a total of up to $M \cdot P$). The advantage of this technique is that it is easy to compute the lower bound, as the M -dimensional problem separates into M instances of the 1-dimensional problem we know how to solve from Section III. However, the resulting bound is likely to be loose. A second lower bounding method we are investigating is the information relaxation method of Brown, Smith, and Sun [70]. The main idea there is to assume the scheduler has access to future channel states (corresponding to the *non-causal* or *offline* model often considered in the literature), but penalize the scheduler for using this information. A clever choice of the penalty function often leads to tight lower bounds on the value function. A third method is the Lagrangian relaxation method discussed in [65], [66]. For our problem, this method is equivalent to relaxing the per slot peak power constraint to an average power constraint (i.e., the scheduler may allocate more than P units of power in some slots, but the average power consumed per slot over the duration of the horizon cannot exceed P). Like the first method we mentioned, the resulting relaxed problem under this method can be separated into M instances of a 1-dimensional problem, this time with an average power constraint of $\frac{P}{M}$ instead of a strict power constraint of P for each receiver. A fourth lower bounding method is the linear programming approach to approximate dynamic programming discussed in [66], [71], and [72]. The idea there is to formulate the dynamic program as a linear program, and approximate the value functions as linear combinations of a set of basis functions. For a more in-depth comparison of the Lagrangian relaxation and approximate linear programming approaches, see [66]. Once lower bounds to the value function are determined from any of these methods, feasible policies can be generated based on our structural results or via one-step greedy optimization with the lower bounds substituted into the right-hand side of the dynamic programming equation.

These same numerical techniques are most likely also the best way to approximate the boundaries of the seven regions of the two receiver optimal policy, and determine a near-optimal split of the power P between the two receivers when the vector of starting receiver buffer levels is in the power-constrained region $\mathcal{R}_{IV}(n, \mathbf{s})$.

The results we have presented in this paper are useful not only in terms of the intuition they provide, but also in generating feasible policies for the most general case of M receivers and

solving subproblems resulting from the relaxation methods described above.

VII. CONCLUSION

In this paper, we considered the problem of transmitting data to one or more receivers over a shared wireless channel in a manner that minimizes power consumption and prevents the receivers' buffers from emptying. We showed that under the finite horizon discounted expected cost, infinite horizon discounted expected cost, and infinite horizon average expected cost criteria, the optimal transmission policy to a single receiver under linear power-rate curves has a modified base-stock structure. When the power-rate curves are generalized to piecewise-linear power-rate curves, the optimal transmission policy to a single receiver has a finite generalized base-stock structure. For the special case when holding costs are linear, the stochastic process representing the channel condition evolution over time is IID, and the maximum number of packets that can be transmitted at any given marginal power cost in a slot is an integer multiple of the drainage rate of the receiver's buffer, we presented an efficient method to compute the critical numbers that fully characterize the modified base-stock and finite generalized base-stock policies.

We also analyzed the structure of the optimal transmission policy for the case of two receivers. In some sense, the structure of the optimal policy was shown to be an extension of the modified base-stock policy; however, the peak power constraint couples the optimal scheduling of the two data streams, and the time-varying channel conditions may result in counterintuitive optimal scheduling decisions that are not possible in the analogous inventory theory problems.

The extension to the most general case of M receivers is quite complex, and it is likely that numerical approximation techniques need to be used to develop further insights on the nature of the optimal policy. We presented a few possible approaches that constitute ongoing work in that regard.

VIII. APPENDIX A - FINITE HORIZON PROOFS

A. *Proof of Theorem 1*

Before proceeding to the proof of Theorem 1, we present a lemma due to Karush [73], which is presented in [74, pp. 237–238].

Lemma 1 (Karush, 1959). *Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ and that f is convex on \mathbb{R} . For $v \leq w$, define $\tilde{f}(v, w) := \min_{z \in [v, w]} f(z)$. Then it follows that:*

- (a) \tilde{f} can be expressed as $\tilde{f}(v, w) = F_1(v) + F_2(w)$, where F_1 is convex nondecreasing and F_2 is convex nonincreasing on \mathbb{R} .
- (b) Suppose that S is a minimizer of f over \mathbb{R} . Then \tilde{f} can be expressed as:

$$\tilde{f}(v, w) = \begin{cases} f(v), & \text{if } S \leq v \\ f(S), & \text{if } v \leq S \leq w \\ f(w), & \text{if } w \leq S \end{cases} .$$

Proof of Theorem 1: We present the proof in three parts.

Part I - Modified Base-Stock Structure: Recall the dynamic programming equation (4):

$$V_n(x, s) = -c_s \cdot x + \min_{\max(x, d) \leq y \leq x + \frac{P}{c_s}} \{g_n(y, s)\} , \quad n = N, N-1, \dots, 1 ,$$

where $g_n(y, s) := c_s \cdot y + h(y - d) + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1}) \mid S_n = s]$. We now show by induction on n that the following statements are true for every $n \in \{1, 2, \dots, N\}$ and all $s \in \mathcal{S}$:

- (i) $g_n(y, s)$ is convex in y on $[d, \infty)$.
- (ii) $\lim_{y \rightarrow \infty} g_n(y, s) = \infty$.
- (iii) $V_n(x, s)$ is convex in x on \mathbb{R}_+ .

Base Case: $n = 1$

Let $s_1 \in \mathcal{S}$ be arbitrary. We have $g_1(y, s_1) = c_{s_1} \cdot y + h(y - d)$, which clearly satisfies (i) and (ii). $y_1^*(x, s_1) = \max(x, d)$ and thus $V_1(x, s_1) = c_{s_1} \cdot (d - x)^+ + h((x - d)^+)$, which is convex in x . We conclude (i)-(iii) are true at time $n = 1$, for all $s \in \mathcal{S}$.

Induction Step: We now assume (i)-(iii) are true for $n = m - 1$ and all $s \in \mathcal{S}$, and show they hold for $n = m$ and an arbitrary $s_m \in \mathcal{S}$. Let $s_{m-1} \in \mathcal{S}$ also be arbitrary. $V_{m-1}(y - d, s_{m-1})$ is convex in y , so $g_m(y, s_m)$ is convex in y as it is the sum of an affine function, $c_{s_m} \cdot y$, a convex function, $h(y - d)$, and a nonnegative weighted sum/integral of convex functions, $\alpha \cdot \mathbb{E}[V_{m-1}(y - d, S_{m-1}) \mid$

$S_m = s_m]$ (see, e.g., [75, Section 3.2] for the relevant results on convexity-preserving operations).

To show (ii) for $n = m$, we have $\lim_{y \rightarrow \infty} g_m(y, s_m) \geq \lim_{y \rightarrow \infty} c_{s_m} \cdot y = \infty$, where the inequality follows from $V_{m-1}(x, s_{m-1}) \geq 0, \forall x \in \mathbb{R}_+, \forall s_{m-1} \in \mathcal{S}$ and $h(y - d) \geq 0$. Moving on to (iii), we have:

$$\begin{aligned} V_m(x, s_m) &= -c_{s_m} \cdot x + \min_{\max(x, d) \leq y \leq x + \frac{P}{c_{s_m}}} \{g_m(y, s_m)\} \\ &= -c_{s_m} \cdot x + F_1(\max(x, d)) + F_2(x + \frac{P}{c_{s_m}}), \end{aligned}$$

where, by Lemma 1, F_1 is convex nondecreasing and F_2 is convex nonincreasing. $F_1(\max(x, d))$ is also convex in x , as it is the composition of a convex increasing function with a convex function, and $V_m(x, s_m)$ is therefore convex in x . This concludes the induction step, and we conclude (i)-(iii) are true for all $n \in \{1, 2, \dots, N\}$.

Next, we define the critical numbers $b_n(s)$ for all $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$:

$$b_n(s) := \min \left\{ \hat{y} \in [d, \infty) : g_n(\hat{y}, s) = \min_{y \in [d, \infty)} g_n(y, s) \right\}.$$

Note that by properties (i) and (ii) from the above induction, the minimum of $g_n(\cdot, s)$ over $[d, \infty)$ is achieved, and the set of minimizers over $[d, \infty)$ is a nonempty closed, convex set. Thus, $b_n(s)$ is well-defined. The form of $y_n^*(x, s)$, (5), then follows from part (b) of Karush's result, Lemma 1, with $g_n(y, s)$ playing the role of f , $\max(x, d)$ the role of v , $x + \frac{P}{c_s}$ the role of w , and $b_n(s)$ the role of S .

Part II - Monotonicity of Thresholds in Time: In this section, we prove (7). We showed above that the optimal action with one time slot remaining is $y_1^*(x, s) = \max(x, d)$, for all $s \in \mathcal{S}$. This is precisely the policy suggested by (5) with $b_1(s) = d$, as $\frac{P}{c_s}$ is at least as great as d . Thus, we conclude the far right equality in (7) holds: $b_1(s) = d, \forall s \in \mathcal{S}$.

In order to show the far left inequality in (7), we claim more generally that $b_n(s) \leq n \cdot d$, for all n and s . This follows from a simple interchange argument, as all packets transmitted beyond $n \cdot d$ incur transmission costs and holding costs for the duration of the horizon; however, they do not satisfy the playout requirements in any remaining slot. Thus, a policy that transmits enough packets to fill the buffer up to $n \cdot d$ at time n is strictly superior to a policy that transmits more

packets.

Next, we prove:

$$b_{n+1}(s) \geq b_n(s), \quad \forall s \in \mathcal{S}, \quad \forall n \in \{1, 2, \dots, N-1\} . \quad (31)$$

By Topkis' Theorem 2.8.1 [76, pg. 76], in order to show (31), it suffices to show that for all $s \in \mathcal{S}$, $n \in \{1, 2, \dots, N-1\}$, and $y^1, y^2 \in [d, (n+1) \cdot d]$, $y^1 > y^2$ implies:

$$g_{n+1}(y^1, s) - g_n(y^1, s) \leq g_{n+1}(y^2, s) - g_n(y^2, s) . \quad (32)$$

We let $s \in \mathcal{S}$ be arbitrary, and proceed by induction on the time slot n .

Base Case: $n = 1$

For all $y \in [d, 2d]$,

$$\begin{aligned} g_2(y, s) - g_1(y, s) &= \alpha \cdot \mathbb{E}[V_1(y-d, S_1) \mid S_2 = s] \\ &= \alpha \cdot \mathbb{E}[c_{S_1} \mid S_2 = s] \cdot (2d - y) , \end{aligned}$$

which is decreasing in y as $\mathbb{E}[c_{S_1} \mid S_2 = s] > 0$.

Induction Step: We assume that (32) is true for all $n = 1, 2, \dots, m-1$ and $s \in \mathcal{S}$. We wish to show it is true for $n = m$. Let $y^1, y^2 \in [d, (m+1) \cdot d]$ be arbitrary, with $y^1 > y^2$. Also, let $\hat{s} \in \mathcal{S}$ be arbitrary. Define:

$$\begin{aligned} \beta_1 &:= \min \left\{ \operatorname{argmin}_{\max(y^1-d, d) \leq \hat{y} \leq y^1-d+\frac{P}{c_{\hat{s}}}} \{g_{m-1}(\hat{y}, \hat{s})\} \right\} \\ \text{and } \beta_2 &:= \min \left\{ \operatorname{argmin}_{\max(y^2-d, d) \leq \hat{y} \leq y^2-d+\frac{P}{c_{\hat{s}}}} \{g_m(\hat{y}, c_{\hat{s}})\} \right\} . \end{aligned}$$

Note that:

$$\max(y^1 - d, d) \leq \beta_1 \leq \beta_1 \vee \beta_2 \leq y^1 - d + \frac{P}{c_{\hat{s}}} , \quad \text{and} \quad (33)$$

$$\max(y^2 - d, d) \leq \beta_1 \wedge \beta_2 \leq \beta_2 \leq y^2 - d + \frac{P}{c_{\hat{s}}} . \quad (34)$$

Then we have:

$$\begin{aligned} & \min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_s}} \{g_m(\hat{y}, \hat{s})\} - \min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_s}} \{g_{m-1}(\hat{y}, \hat{s})\} \\ & \leq g_m(\beta_1 \vee \beta_2, \hat{s}) - g_{m-1}(\beta_1, \hat{s}) \end{aligned} \quad (35)$$

$$\leq g_m(\beta_2, \hat{s}) - g_{m-1}(\beta_1 \wedge \beta_2, \hat{s}) \quad (36)$$

$$\leq \min_{\max(y^2-d,d)\leq\hat{y}\leq y^2-d+\frac{P}{c_s}} \{g_m(\hat{y}, \hat{s})\} - \min_{\max(y^2-d,d)\leq\hat{y}\leq y^2-d+\frac{P}{c_s}} \{g_{m-1}(\hat{y}, \hat{s})\} . \quad (37)$$

Equation (35) follows from (33) and (37) follows from (34). If $\beta_2 \geq \beta_1$, (36) holds with equality. Otherwise, it follows from the induction hypothesis. Since \hat{s} was arbitrary, (37) holds for all $\hat{s} \in \mathcal{S}$. Therefore, combined with the fact that the Markov process $\{S_n\}_{n=N, N-1, \dots, 1}$ is homogeneous, (37) implies:

$$\begin{aligned} & \mathbb{E} \left[\min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_{S_m}}} \{g_m(\hat{y}, S_m)\} \mid S_{m+1} = s \right] \\ & - \mathbb{E} \left[\min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_{S_{m-1}}}} \{g_{m-1}(\hat{y}, S_{m-1})\} \mid S_m = s \right] \\ & \leq \mathbb{E} \left[\min_{\max(y^2-d,d)\leq\hat{y}\leq y^2-d+\frac{P}{c_{S_m}}} \{g_m(\hat{y}, S_m)\} \mid S_{m+1} = s \right] \\ & - \mathbb{E} \left[\min_{\max(y^2-d,d)\leq\hat{y}\leq y^2-d+\frac{P}{c_{S_{m-1}}}} \{g_{m-1}(\hat{y}, S_{m-1})\} \mid S_m = s \right] . \end{aligned} \quad (38)$$

Finally, we have:

$$\begin{aligned} & g_{m+1}(y^1, s) - g_m(y^1, s) \\ & = \alpha \cdot \mathbb{E} [V_m(y^1 - d, S_m) \mid S_{m+1} = s] - \alpha \cdot \mathbb{E} [V_{m-1}(y^1 - d, S_{m-1}) \mid S_m = s] \\ & = \alpha \cdot \mathbb{E} \left[\min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_{S_m}}} \{g_m(\hat{y}, S_m)\} \mid S_{m+1} = s \right] \\ & - \alpha \cdot \mathbb{E} \left[\min_{\max(y^1-d,d)\leq\hat{y}\leq y^1-d+\frac{P}{c_{S_{m-1}}}} \{g_{m-1}(\hat{y}, S_{m-1})\} \mid S_m = s \right] \end{aligned} \quad (39)$$

$$\begin{aligned} &\leq \alpha \cdot \mathbb{E} \left[\min_{\max(y^2-d, d) \leq \hat{y} \leq y^2-d+\frac{P}{c_{S_m}}} \{g_m(\hat{y}, S_m)\} \mid S_{m+1} = s \right] \\ &\quad - \alpha \cdot \mathbb{E} \left[\min_{\max(y^2-d, d) \leq \hat{y} \leq y^2-d+\frac{P}{c_{S_{m-1}}}} \{g_{m-1}(\hat{y}, S_{m-1})\} \mid S_m = s \right] \end{aligned} \quad (40)$$

$$\begin{aligned} &= \alpha \cdot \mathbb{E} [V_m(y^2 - d, S_m) \mid S_{m+1} = s] - \alpha \cdot \mathbb{E} [V_{m-1}(y^2 - d, S_{m-1}) \mid S_m = s] \\ &= g_{m+1}(y^2, s) - g_m(y^2, s). \end{aligned} \quad (41)$$

Here, (39) and (41) follow from the fact that $\mathbb{E}[c_{S_{m-1}} \mid S_m = s] = \mathbb{E}[c_{S_m} \mid S_{m+1} = s]$, and (40) follows from (38). This completes the induction step, and the proof of (7).

Part III - Monotonicity of Thresholds in the Channel Condition: Finally, we show (8), the monotonicity of the thresholds in the channel condition, when the channel condition process is IID. The far left inequality follows from the same interchange argument described above, showing $b_n(s) \leq n \cdot d$ for all s and n . We now show the far right equality of (8), $b_n(s_{\text{worst}}) = d$. To satisfy feasibility, we must have $b_n(s) \geq d$ for all $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$. To see that $b_n(s_{\text{worst}}) \leq d$, assume the channel condition at time n is s_{worst} , and consider two control policies satisfying (5), with the same critical numbers $b_m(s)$, for all times $m < n$. At time n , the first policy, π^1 , transmits according to (5), with critical number $b_n(s_{\text{worst}}) = d + \epsilon$ ($\epsilon > 0$), and the second, π^2 , transmits according to (5), with critical number $b_n(s_{\text{worst}}) = d$. These two strategies result in the same control action at time n if $x_n \geq d + \epsilon$, and we have already shown it is not optimal to fill the buffer beyond $n \cdot d$, so we only need to consider the case where $x_n < d + \epsilon$ and $\epsilon \leq (n-1) \cdot d$. Let $Z_n^1, Z_{n-1}^1, \dots, Z_1^1$ and $Z_n^2, Z_{n-1}^2, \dots, Z_1^2$ be random variables representing the number of packets transmitted at times $n, n-1, \dots, 1$ by π^1 and π^2 , respectively. If $d \leq x_n \leq d + \epsilon$, then $Z_n^2 = 0$ and $Z_n^1 - Z_n^2 = Z_n^1 = \min\left\{\frac{P}{c_{\max}}, d + \epsilon - x_n\right\}$. If $x_n < d$, then $Z_n^2 = d - x_n$, $Z_n^1 = \min\left\{\frac{P}{c_{\max}}, d + \epsilon - x_n\right\}$, and $Z_n^1 - Z_n^2 = \min\left\{\frac{P}{c_{\max}} - d + x_n, \epsilon\right\}$. Thus, for all $x_n < d + \epsilon$, we have $Z_n^1 - Z_n^2 \geq 0$. If $Z_n^1 - Z_n^2 = 0$, the two control policies result in the same actions for all remaining times, and therefore result in the same expected cost. So we only need to consider the case where $\lambda := Z_n^1 - Z_n^2 > 0$. Because the critical numbers at times $n-1, n-2, \dots, 1$ are the same for both policies, for any realization, ω , of the channel condition over future times,

we have $Z_m^1(\omega) \leq Z_m^2(\omega)$, $\forall m \in \{n-1, \dots, 1\}$. Moreover, because the scheduler must satisfy the playout requirements for the last n slots, we have $\sum_{m=1}^{n-1} (Z_m^2(\omega) - Z_m^1(\omega)) = \lambda$; i.e., over the remainder of the horizon, an extra λ packets are transmitted under the second policy. The total discounted holding costs from time n until the end of the horizon are therefore lower for π^2 than π^1 , because the number of packets remaining after transmission in each slot is never greater under policy π^2 . Furthermore, the total discounted transmission costs of the extra λ packets are also lower for π^2 as they are transmitted at the maximum cost c_{\max} under π^1 , and transmitted later (and therefore discounted more heavily) under π^2 . Thus, the total discounted transmission plus holding costs are lower for π^2 under all realizations, and the expected discounted cost of π^2 is lower than π^1 . We conclude $b_n(s_{\text{worst}}) = d$.

To show $c_{s^1} \leq c_{s^2}$ implies $b_n(s^1) \geq b_n(s^2)$, we follow Kalymon's methodology for the proof of Theorem 1.3 in [46]. For all $y \in [d, \infty)$, we have:

$$\begin{aligned}
g_n(y, s^2) &= c_{s^2} \cdot y + h(y - d) + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1})] \\
&= (c_{s^2} - c_{s^1}) \cdot y + c_{s^1} \cdot y + h(y - d) + \alpha \cdot \mathbb{E}[V_{n-1}(y - d, S_{n-1})] \\
&= (c_{s^2} - c_{s^1}) \cdot y + g_n(y, s^1) .
\end{aligned} \tag{42}$$

Assume $b_n(s^1) < b_n(s^2)$ for some $n \in \{1, 2, \dots, N\}$ and $s^1, s^2 \in \mathcal{S}$, with $c_{s^1} \leq c_{s^2}$. Substituting first $y = b_n(s^1)$ and then $y = b_n(s^2)$ into (42) yields:

$$\begin{aligned}
(c_{s^2} - c_{s^1}) \cdot b_n(s^1) + g_n(b_n(s^1), s^1) &= g_n(b_n(s^1), s^2) \\
&\geq g_n(b_n(s^2), s^2) \\
&= (c_{s^2} - c_{s^1}) \cdot b_n(s^2) + g_n(b_n(s^2), s^1) .
\end{aligned} \tag{43}$$

Yet, $c_{s^1} \leq c_{s^2}$ and $b_n(s^1) < b_n(s^2)$ imply:

$$(c_{s^2} - c_{s^1}) \cdot b_n(s^1) < (c_{s^2} - c_{s^1}) \cdot b_n(s^2) . \tag{44}$$

Equations (43) and (44) imply:

$$g_n(b_n(s^1), s^1) > g_n(b_n(s^2), s^1) ,$$

which clearly contradicts the fact that $b_n(s^1)$ is a global minimizer of $g_n(\cdot, s^1)$. We conclude that $c_{s^1} \leq c_{s^2}$ implies $b_n(s^1) \geq b_n(s^2)$, completing the proofs of (8) and Theorem 1. \square

B. Proof of Theorem 3

While the proof is similar in spirit to the proof of a finite generalized base-stock policy in [56, pp. 324–334], some key differences include the introduction of (i) stochastic channel conditions (ordering costs); (ii) the underflow constraint $x + z \geq d$; and (iii) the power constraint $z \leq \tilde{z}_{\max}(s)$.

We show by induction on n that the following two statements are true for every $n \in \{1, 2, \dots, N\}$ and $s \in \mathcal{S}$:

- (i) $V_n(x, s)$ is convex in x on \mathbb{R}_+ .
- (ii) There exists a nonincreasing sequence of critical numbers $\{b_{n,k}(s)\}_{k \in \{-1, 0, 1, \dots, K\}}$ such that the optimal control action with n slots remaining is given by:

$$z_n^*(x, s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } b_{n,k}(s) - \tilde{z}_{k-1}(s) \leq x < b_{n,k-1}(s) - \tilde{z}_{k-1}(s) , \\ & k \in \{0, 1, \dots, K\} \\ b_{n,k}(s) - x, & \text{if } b_{n,k}(s) - \tilde{z}_k(s) \leq x < b_{n,k}(s) - \tilde{z}_{k-1}(s) , \\ & k \in \{0, 1, \dots, K-1\} \\ b_{n,K}(s) - x, & \text{if } b_{n,K}(s) - \tilde{z}_{\max}(s) \leq x < b_{n,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } 0 \leq x < b_{n,K}(s) - \tilde{z}_{\max}(s) \end{cases} \quad (45)$$

Base Case: $n = 1$

$$\begin{aligned} V_1(x, s) &= \min_{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)} \{c(z, s) + h(x + z - d)\} \\ &= c(\max\{0, d-x\}, s) + h(\max\{0, x-d\}) , \end{aligned} \quad (46)$$

which is convex because $c(\cdot, s)$ and $h(\cdot)$ are both convex and nondecreasing functions, and

$\max\{0, d - x\}$ and $\max\{0, x - d\}$ are both convex functions (see, e.g., [75, Section 3.2] for the relevant results on convexity-preserving operations). Further, let $b_{1,-1}(s) = \infty$ and $b_{1,k}(s) = d$ for all $k \in \{0, 1, \dots, K\}$. Then (45) is equivalent to $z_1^*(x, s) = \max\{0, d - x\}$, which clearly achieves the minimum in (46).

Induction Step: We now assume (i)-(ii) are true for $n = m - 1$ and all $s \in \mathcal{S}$, and show they hold for $n = m$ and an arbitrary $s \in \mathcal{S}$. Let $\check{x}, \hat{x} \in \mathbb{R}_+$ and $\theta \in [0, 1]$ be arbitrary, and define $\bar{x} := \theta \cdot \check{x} + (1 - \theta) \cdot \hat{x}$. We have:

$$\begin{aligned}
& V_m(\theta \cdot \check{x} + (1 - \theta) \cdot \hat{x}, s) \\
&= V_m(\bar{x}, s) \\
&= \min_{\max\{0, d - \bar{x}\} \leq z \leq \tilde{z}_{\max}(s)} \left\{ c(z, s) + h(\bar{x} + z - d) + \alpha \cdot \mathbb{E}[V_{m-1}(\bar{x} + z - d, S_{m-1}) \mid S_m = s] \right\} \\
&\leq \min_{\substack{\max\{0, d - \check{x}\} \leq \check{z} \leq \tilde{z}_{\max}(s) \\ \max\{0, d - \hat{x}\} \leq \hat{z} \leq \tilde{z}_{\max}(s)}} \left\{ \begin{aligned} & c(\theta \cdot \check{z} + (1 - \theta) \cdot \hat{z}, s) + h(\bar{x} + \theta \cdot \check{z} + (1 - \theta) \cdot \hat{z} - d) \\ & + \alpha \cdot \mathbb{E}[V_{m-1}(\bar{x} + \theta \cdot \check{z} + (1 - \theta) \cdot \hat{z} - d, S_{m-1}) \mid S_m = s] \end{aligned} \right\} \quad (47) \\
&\leq \min_{\substack{\max\{0, d - \check{x}\} \leq \check{z} \leq \tilde{z}_{\max}(s) \\ \max\{0, d - \hat{x}\} \leq \hat{z} \leq \tilde{z}_{\max}(s)}} \left\{ \begin{aligned} & \theta \cdot c(\check{z}, s) + (1 - \theta) \cdot c(\hat{z}, s) + \\ & \theta \cdot h(\check{x} + \check{z} - d) + (1 - \theta) \cdot h(\hat{x} + \hat{z} - d) \\ & + \alpha \cdot \theta \cdot \mathbb{E}[V_{m-1}(\check{x} + \check{z} - d, S_{m-1}) \mid S_m = s] \\ & + \alpha \cdot (1 - \theta) \cdot \mathbb{E}[V_{m-1}(\hat{x} + \hat{z} - d, S_{m-1}) \mid S_m = s] \end{aligned} \right\} \quad (48) \\
&= \theta \cdot \min_{\max\{0, d - \check{x}\} \leq \check{z} \leq \tilde{z}_{\max}(s)} \left\{ \begin{aligned} & c(\check{z}, s) + h(\check{x} + \check{z} - d) \\ & + \alpha \cdot \mathbb{E}[V_{m-1}(\check{x} + \check{z} - d, S_{m-1}) \mid S_m = s] \end{aligned} \right\} \\
&\quad + (1 - \theta) \cdot \min_{\max\{0, d - \hat{x}\} \leq \hat{z} \leq \tilde{z}_{\max}(s)} \left\{ \begin{aligned} & c(\hat{z}, s) + h(\hat{x} + \hat{z} - d) \\ & + \alpha \cdot \mathbb{E}[V_{m-1}(\hat{x} + \hat{z} - d, S_{m-1}) \mid S_m = s] \end{aligned} \right\} \\
&= \theta \cdot V_m(\check{x}, s) + (1 - \theta) \cdot V_m(\hat{x}, s),
\end{aligned}$$

where (48) follows from the convexity of $c(\cdot, s)$, $h(\cdot)$, and $\mathbb{E}[V_{m-1}(\cdot, S_{m-1}) \mid S_m = s]$, the last of which follows from the induction hypothesis. Equation (47) follows from the fact that for every $\max\{0, d - \check{x}\} \leq \check{z} \leq \tilde{z}_{\max}(s)$ and $\max\{0, d - \hat{x}\} \leq \hat{z} \leq \tilde{z}_{\max}(s)$, there exists a

$\max\{0, d - \bar{x}\} \leq \bar{z} \leq \tilde{z}_{\max}(s)$ (namely, $\bar{z} := \theta \cdot \check{z} + (1 - \theta) \cdot \hat{z}$) such that:

$$\begin{aligned} & c(\bar{z}, s) + h(\bar{x} + \bar{z} - d) + \alpha \cdot \mathbb{E}[V_{m-1}(\bar{x} + \bar{z} - d, S_{m-1}) \mid S_m = s] \\ &= c(\theta \cdot \check{z} + (1 - \theta) \cdot \hat{z}, s) + h(\bar{x} + \theta \cdot \check{z} + (1 - \theta) \cdot \hat{z} - d) \\ &+ \alpha \cdot \mathbb{E}[V_{m-1}(\bar{x} + \theta \cdot \check{z} + (1 - \theta) \cdot \hat{z} - d, S_{m-1}) \mid S_m = s] . \end{aligned}$$

This concludes the induction step for (i) and we now proceed to (ii).

Note first that $\tilde{g}_m(y, s) = h(y - d) + \alpha \cdot \mathbb{E}[V_{m-1}(y - d, S_{m-1}) \mid S_m = s]$ is convex in y , as $h(\cdot)$ is convex, and $V_{m-1}(x, s)$ is convex in x for every $s \in \mathcal{S}$ by the induction hypothesis. Let $b_{m,-1}(s) := \infty$ and

$$b_{m,k}(s) := \max\left\{d, \inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_k(s)\}\right\}, \quad \forall k \in \{0, 1, \dots, K\},$$

where $\tilde{g}_m^+(b, s)$ represents the right derivative:

$$\tilde{g}_m^+(b, s) := \lim_{y \downarrow b} \frac{\tilde{g}_m(y, s) - \tilde{g}_m(b, s)}{y - b},$$

which is nondecreasing and continuous from the right, by the convexity of $\tilde{g}_m(\cdot, s)$ [77, Section 24]. Note that $\{b_{m,k}(s)\}_{k \in \{-1, 0, 1, \dots, K\}}$ is a nonincreasing sequence, because the sequence $\{\tilde{c}_k(s)\}_{k \in \{0, 1, \dots, K\}}$ is nondecreasing. We show the optimal control action $z_m^*(x, s)$ is then given by (45), by considering the four exhaustive cases.

Case 1: $b_{m,k}(s) - \tilde{z}_{k-1}(s) \leq x < b_{m,k-1}(s) - \tilde{z}_{k-1}(s)$, $k \in \{0, 1, \dots, K\}$

In order to show $z_m^*(x, s)$ is given by (45), it suffices to show:

$$c^+(z, s) + \tilde{g}_m^+(x + z, s) < 0, \quad \text{for } \max\{0, d - x\} \leq z < \tilde{z}_{k-1}(s), \quad \text{and} \quad (49)$$

$$c^+(z, s) + \tilde{g}_m^+(x + z, s) \geq 0, \quad \text{for } \tilde{z}_{k-1}(s) \leq z \leq \tilde{z}_{\max}(s). \quad (50)$$

First, let $z \in \left[\max\{0, d - x\}, \tilde{z}_{k-1}(s)\right)$ be arbitrary, and let $j \in \{0, 1, \dots, k - 1\}$ be such that $z \in \left[\tilde{z}_{j-1}(s), \tilde{z}_j(s)\right)$. If $b_{m,k-1}(s) = d$, then $b_{m,k}(s) = d$, as $d \leq b_{m,k}(s) \leq b_{m,k-1}(s) = d$. Yet, $b_{m,k}(s) = b_{m,k-1}(s) = d$ implies $d - \tilde{z}_{k-1}(s) \leq x < d - \tilde{z}_{k-1}(s)$, which is vacuous. Therefore,

we need only consider $b_{m,k-1}(s) = \inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_{k-1}(s)\}$. By the construction of the piecewise-linear function $c(\cdot, s)$, $z < \tilde{z}_{k-1}(s)$ implies:

$$c^+(z, s) \leq \tilde{c}_{k-1}(s) . \quad (51)$$

We also have:

$$x + z < x + \tilde{z}_{k-1}(s) < b_{m,k-1}(s) = \inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_{k-1}(s)\} ,$$

which implies:

$$\tilde{g}_m^+(x + z, s) < -\tilde{c}_{k-1}(s) . \quad (52)$$

Summing (51) and (52) yields (49).

Next, let $z \in [\tilde{z}_{k-1}(s), \tilde{z}_{\max}(s)]$ be arbitrary, so that by construction of $c(\cdot, s)$:

$$c^+(z, s) \geq \tilde{c}_k(s) . \quad (53)$$

We also have:

$$x + z \geq x + \tilde{z}_{k-1}(s) \geq b_{m,k}(s) \geq \inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_k(s)\} ,$$

which, in combination with the nondecreasing nature of $\tilde{g}_m^+(\cdot, s)$, implies:

$$\tilde{g}_m^+(x + z, s) \geq \tilde{g}_m^+\left(\inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_k(s)\}, s\right) . \quad (54)$$

Because $\tilde{g}_m^+(\cdot, s)$ is continuous from the right,

$$\tilde{g}_m^+\left(\inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_k(s)\}, s\right) \geq -\tilde{c}_k(s) . \quad (55)$$

Combining (54) and (55), and summing with (53) yields (50).

Case 2: $b_{m,k}(s) - \tilde{z}_k(s) \leq x < b_{m,k}(s) - \tilde{z}_{k-1}(s)$, $k \in \{0, 1, \dots, K - 1\}$

In order to show $z_m^*(x, s)$ is given by (45), it suffices to show:

$$c'^+(z, s) + \tilde{g}_m'^+(x + z, s) < 0, \text{ for } \max\{0, d - x\} \leq z < b_{m,k}(s) - x, \text{ and} \quad (56)$$

$$c'^+(z, s) + \tilde{g}_m'^+(x + z, s) \geq 0, \text{ for } b_{m,k}(s) - x \leq z \leq \tilde{z}_{\max}(s). \quad (57)$$

First, let $z \in [\max\{0, d - x\}, b_{m,k}(s) - x)$ be arbitrary. This case is vacuous if $b_{m,k}(s) = d$, so $b_{m,k}(s) = \inf\{b \mid \tilde{g}_m'^+(b, s) \geq -\tilde{c}_k(s)\}$. Thus, we have:

$$x + z < b_{m,k}(s) = \inf\{b \mid \tilde{g}_m'^+(b, s) \geq -\tilde{c}_k(s)\},$$

which implies:

$$\tilde{g}_m'^+(x + z, s) < -\tilde{c}_k(s). \quad (58)$$

Furthermore, from $z < b_{m,k}(s) - x \leq \tilde{z}_k(s)$ and the construction of the piecewise-linear function $c(\cdot, s)$,

$$c'^+(z, s) \leq \tilde{c}_k(s). \quad (59)$$

Summing (58) and (59) yields (56).

Next, let $z \in [b_{m,k}(s) - x, \tilde{z}_{\max}(s)]$ be arbitrary, so that $z \geq b_{m,k}(s) - x > \tilde{z}_{k-1}(s)$, which by the construction of the piecewise-linear function $c(\cdot, s)$ implies:

$$c'^+(z, s) \geq \tilde{c}_k(s). \quad (60)$$

We also have $x + z \geq b_{m,k}(s) \geq \inf\{b \mid \tilde{g}_m'^+(b, s) \geq -\tilde{c}_k(s)\}$. Therefore, because $\tilde{g}_m'^+(\cdot, s)$ is nondecreasing and continuous from the right,

$$\tilde{g}_m'^+(x + z, s) \geq \tilde{g}_m'^+(\inf\{b \mid \tilde{g}_m'^+(b, s) \geq -\tilde{c}_k(s)\}, s) \geq -\tilde{c}_k(s). \quad (61)$$

Summing (60) and (61) yields (57).

Case 3: $b_{m,K}(s) - \tilde{z}_{\max}(s) \leq x < b_{m,K}(s) - \tilde{z}_{K-1}(s)$

This case is the same as Case 2, with K in place of k , and $\tilde{z}_{\max}(s)$ in place of $\tilde{z}_k(s)$.

Case 4: $0 \leq x < b_{m,K}(s) - \tilde{z}_{\max}(s)$

Let $z \in [\max\{0, d-x\}, \tilde{z}_{\max}(s)]$ be arbitrary. $\tilde{z}_{\max}(s) \geq d$ by assumption, so this case is vacuous if $b_{m,K} = d$. Thus, we have $b_{m,K}(s) = \inf\{b \mid \tilde{g}_m^+(b, s) \geq -\tilde{c}_K(s)\}$, which, in combination with $x + z < x + \tilde{z}_{\max} < b_{m,K}(s)$, implies:

$$\tilde{g}_m^+(x + z, s) < -\tilde{c}_K(s) . \quad (62)$$

Additionally, $z < \tilde{z}_{\max}(s)$ implies:

$$c^+(z, s) \leq \tilde{c}_K(s) . \quad (63)$$

Summing (62) and (63) yields $c^+(z, s) + \tilde{g}_m^+(x + z, s) < 0$ for all $z \in [\max\{0, d-x\}, \tilde{z}_{\max}(s)]$, which implies $z_m^*(x, s) = \tilde{z}_{\max}(s)$. \square

C. Proof of Theorem 4

We proceed in a manner similar to [52], incorporating the per slot peak power constraints and the relaxing the linear ordering costs to piecewise-linear convex ordering costs. Before proving Theorem 2, we state and prove two lemmas. Let $\bar{\pi}$ be a strategy that prescribes transmitting according to (17).

Lemma 2. *If $\bar{\pi}$ is optimal for periods $m-1, m-2, \dots, 1$, then*

$$\alpha \cdot \mathbb{E} [V_{l-1}((r-1) \cdot d + \eta, S) - V_{l-1}((r-1) \cdot d, S)] \geq -\eta \cdot (\tilde{\gamma}_{l,r+1} + h), \quad (64)$$

for all $(l, r, \eta) \in \mathcal{Z}_1 := \{(l, r, \eta) \in \mathbb{N} \times \mathbb{N} \times [0, d] : 1 \leq l \leq m, 1 \leq r \leq l\}$.

Proof. We proceed by induction on l .

Base Case: $l = 1$

$l = 1$ implies $r = 1$, so we have:

$$\begin{aligned}
& \alpha \cdot \mathbb{E} [V_{l-1}((r-1) \cdot d + \eta, S) - V_{l-1}((r-1) \cdot d, S)] \\
&= \alpha \cdot \mathbb{E} [V_0(\eta, S) - V_0(0, S)] \\
&= 0 \\
&\geq -\eta \cdot h \\
&= -\eta \cdot (\tilde{\gamma}_{1,2} + h) ,
\end{aligned}$$

and we conclude (64) holds for $l = 1$.

Induction Step

Assume (64) is true for $l = 2, 3, \dots, t$ and all r and η such that $(l, r, \eta) \in \mathcal{Z}_1$. We show (64) is true for $l = t + 1$ by letting r and η be arbitrary such that $(t + 1, r, \eta) \in \mathcal{Z}_1$. Note that $(t + 1, r, \eta) \in \mathcal{Z}_1$ implies $t \leq m - 1$, so $\bar{\pi}$ is optimal at time t , and we have:

$$\begin{aligned}
& \alpha \cdot \mathbb{E} [V_t((r-1) \cdot d + \eta, S) - V_t((r-1) \cdot d, S)] \\
&= \sum_{\{s: b_{t,0}(s) \leq (r-1) \cdot d\}} \alpha \cdot p(s) \cdot \left[h \cdot \eta + \alpha \cdot \mathbb{E} [V_{t-1}((r-2) \cdot d + \eta, S) - V_{t-1}((r-2) \cdot d, S)] \right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{aligned} & \sum_{\{s: (r-1+\tilde{L}_{k-1}(s)) \cdot d < b_{t,k}(s) \leq (r-1+\tilde{L}_k(s)) \cdot d\}} -\alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_k(s)) \\ & + \sum_{\{s: b_{t,k+1}(s) \leq (r-1+\tilde{L}_k(s)) \cdot d < b_{t,k}(s)\}} \alpha \cdot p(s) \cdot \left[h \cdot \eta + \alpha \cdot \mathbb{E} \left[\begin{aligned} & V_{t-1}((r-2 + \tilde{L}_k(s)) \cdot d + \eta, S) \\ & - V_{t-1}((r-2 + \tilde{L}_k(s)) \cdot d, S) \end{aligned} \right] \right] \end{aligned} \right\} \\
&+ \sum_{\{s: (r-1+\tilde{L}_{K-1}(s)) \cdot d < b_{t,K}(s) \leq (r-1+\tilde{L}_{\max}(s)) \cdot d\}} -\alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_K(s)) \\
&+ \sum_{\{s: b_{t,K}(s) > (r-1+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot \left[h \cdot \eta + \alpha \cdot \mathbb{E} \left[\begin{aligned} & V_{t-1}((r-2 + \tilde{L}_{\max}(s)) \cdot d + \eta, S) \\ & - V_{t-1}((r-2 + \tilde{L}_{\max}(s)) \cdot d, S) \end{aligned} \right] \right]
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{\{s: b_{t,0}(s) \leq (r-1) \cdot d\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot \tilde{\gamma}_{t,r} \right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{aligned} &\sum_{\{s: (r-1+\tilde{L}_{k-1}(s)) \cdot d < b_{t,k}(s) \leq (r-1+\tilde{L}_k(s)) \cdot d\}} -\alpha \cdot p(s) \cdot \left(\eta \cdot \tilde{c}_k(s) \right) \\ &+ \sum_{\{s: b_{t,k+1}(s) \leq (r-1+\tilde{L}_k(s)) \cdot d < b_{t,k}(s)\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot \tilde{\gamma}_{t,r+\tilde{L}_k(s)} \right] \end{aligned} \right\} \\
&+ \sum_{\{s: (r-1+\tilde{L}_{K-1}(s)) \cdot d < b_{t,K}(s) \leq (r-1+\tilde{L}_{\max}(s)) \cdot d\}} -\alpha \cdot p(s) \cdot \left(\eta \cdot \tilde{c}_K(s) \right) \\
&+ \sum_{\{s: b_{t,K}(s) > (r-1+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)} \right] \\
&= -\alpha \cdot \eta \cdot \left\{ \begin{aligned} &\sum_{s: \tilde{c}_0(s) \geq \tilde{\gamma}_{t,r}} p(s) \cdot \tilde{\gamma}_{t,r} \\ &+ \sum_{k=0}^{K-1} \left\{ \begin{aligned} &\sum_{s: \tilde{\gamma}_{t,r+\tilde{L}_k(s)} \leq \tilde{c}_k(s) < \tilde{\gamma}_{t,r+\tilde{L}_{k-1}(s)}} p(s) \cdot \tilde{c}_k(s) \\ &+ \sum_{s: \tilde{c}_k(s) < \tilde{\gamma}_{t,r+\tilde{L}_k(s)} \leq \tilde{c}_{k+1}(s)} p(s) \cdot \tilde{\gamma}_{t,r+\tilde{L}_k(s)} \end{aligned} \right\} \\ &+ \sum_{s: \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)} \leq \tilde{c}_K(s) < \tilde{\gamma}_{t,r+\tilde{L}_{K-1}(s)}} p(s) \cdot \tilde{c}_K(s) \\ &+ \sum_{s: \tilde{c}_K(s) < \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)}} p(s) \cdot \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)} \end{aligned} \right\} \\
&= -\eta \cdot (\tilde{\gamma}_{t+1,r+1} + h) ,
\end{aligned}$$

where the inequality follows from the induction hypothesis, and the penultimate equality follows from the definition $b_{n,k} := j \cdot d$, if $\tilde{\gamma}_{n,j+1} \leq \tilde{c}_k(s) < \tilde{\gamma}_{n,j}$. This concludes the induction step, and the proof of Lemma 2. \square

Lemma 3. *If $\bar{\pi}$ is optimal for periods $m-1, m-2, \dots, 1$, then*

$$\alpha \cdot \mathbb{E} \left[V_{l-1}((r-1) \cdot d - \eta, S) - V_{l-1}((r-1) \cdot d, S) \right] \geq \eta \cdot (\tilde{\gamma}_{l,r} + h), \quad (65)$$

for all $(l, r, \eta) \in \mathcal{Z}_2 := \{(l, r, \eta) \in \mathbb{N} \times \mathbb{N} \times [0, d] : 2 \leq l \leq m, 2 \leq r \leq l\}$.

Proof. We proceed by induction on l .

Base Case: $l = 2$

$l = 2$ implies $r = 2$, so we have:

$$\begin{aligned} \alpha \cdot \mathbb{E} [V_{l-1}((r-1) \cdot d - \eta, S) - V_{l-1}((r-1) \cdot d, S)] \\ &= \alpha \cdot \mathbb{E} [V_1(d - \eta, S) - V_1(d, S)] \\ &= \alpha \cdot \mathbb{E} [c(\eta, S)] \\ &= \eta \cdot (\tilde{\gamma}_{2,2} + h), \end{aligned}$$

where the last equality follows from $\gamma_{2,2} = -h + \alpha \cdot \mathbb{E}[\tilde{c}_0(S)]$, and the fact that $\eta \leq \tilde{z}_0(s)$ for every $s \in \mathcal{S}$. So (65) holds with equality for $l = 2$.

Induction Step

Assume (65) is true for $l = 2, 3, \dots, t$ and all r and η such that $(l, r, \eta) \in \mathcal{Z}_2$. We show (65) is true for $l = t + 1$ by letting r and η be arbitrary such that $(t + 1, r, \eta) \in \mathcal{Z}_2$. Note that $(t + 1, r, \eta) \in \mathcal{Z}_2$ implies $t \leq m - 1$, so $\bar{\pi}$ is optimal at time t , and we have:

$$\begin{aligned}
& \alpha \cdot \mathbb{E} [V_t((r-1) \cdot d - \eta, S) - V_t((r-1) \cdot d, S)] \\
&= \sum_{\{s: b_{t,0}(s) \leq (r-2) \cdot d\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot h + \alpha \cdot \mathbb{E} [V_{t-1}((r-2) \cdot d - \eta, S) - V_{t-1}((r-2) \cdot d, S)] \right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{aligned} & \sum_{\{s: (r-2+\tilde{L}_{k-1}(s)) \cdot d < b_{t,k}(s) \leq (r-2+\tilde{L}_k(s)) \cdot d\}} \alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_k(s)) \\ & + \sum_{\{s: b_{t,k+1}(s) \leq (r-2+\tilde{L}_k(s)) \cdot d < b_{t,k}(s)\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot h + \alpha \cdot \mathbb{E} \left[\begin{aligned} & V_{t-1}((r-2+\tilde{L}_k(s)) \cdot d - \eta, S) \\ & - V_{t-1}((r-2+\tilde{L}_k(s)) \cdot d, S) \end{aligned} \right] \right] \end{aligned} \right\} \\
&+ \sum_{\{s: (r-2+\tilde{L}_{K-1}(s)) \cdot d < b_{t,K}(s) \leq (r-2+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_K(s)) \\
&+ \sum_{\{s: b_{t,K}(s) > (r-2+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot \left[-\eta \cdot h + \alpha \cdot \mathbb{E} \left[\begin{aligned} & V_{t-1}((r-2+\tilde{L}_{\max}(s)) \cdot d - \eta, S) \\ & - V_{t-1}((r-2+\tilde{L}_{\max}(s)) \cdot d, S) \end{aligned} \right] \right] \\
&\geq \sum_{\{s: b_{t,0}(s) \leq (r-2) \cdot d\}} \alpha \cdot p(s) \cdot \left[\eta \cdot \tilde{\gamma}_{t,r-1} \right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{aligned} & \sum_{\{s: (r-2+\tilde{L}_{k-1}(s)) \cdot d < b_{t,k}(s) \leq (r-2+\tilde{L}_k(s)) \cdot d\}} \alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_k(s)) \\ & + \sum_{\{s: b_{t,k+1}(s) \leq (r-2+\tilde{L}_k(s)) \cdot d < b_{t,k}(s)\}} \alpha \cdot p(s) \cdot \left[\eta \cdot \tilde{\gamma}_{t,r-1+\tilde{L}_k(s)} \right] \end{aligned} \right\} \\
&+ \sum_{\{s: (r-2+\tilde{L}_{K-1}(s)) \cdot d < b_{t,K}(s) \leq (r-2+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot (\eta \cdot \tilde{c}_K(s)) \\
&+ \sum_{\{s: b_{t,K}(s) > (r-2+\tilde{L}_{\max}(s)) \cdot d\}} \alpha \cdot p(s) \cdot \left[\eta \cdot \tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)} \right]
\end{aligned}$$

$$= \alpha \cdot \eta \cdot \left\{ \begin{aligned} & \sum_{s: \tilde{c}_0(s) \geq \tilde{\gamma}_{t,r-1}} p(s) \cdot \tilde{\gamma}_{t,r-1} \\ & + \sum_{k=0}^{K-1} \left\{ \begin{aligned} & \sum_{s: \tilde{\gamma}_{t,r-1+\tilde{L}_k(s)} \leq \tilde{c}_k(s) < \tilde{\gamma}_{t,r-1+\tilde{L}_{k-1}(s)}} p(s) \cdot \tilde{c}_k(s) \\ & + \sum_{s: \tilde{c}_k(s) < \tilde{\gamma}_{t,r-1+\tilde{L}_k(s)} \leq \tilde{c}_{k+1}(s)} p(s) \cdot \tilde{\gamma}_{t,r-1+\tilde{L}_k(s)} \end{aligned} \right\} \\ & + \sum_{s: \tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)} \leq \tilde{c}_K(s) < \tilde{\gamma}_{t,r-1+\tilde{L}_{K-1}(s)}} p(s) \cdot \tilde{c}_K(s) \\ & + \sum_{s: \tilde{c}_K(s) < \tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)}} p(s) \cdot \tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)} \end{aligned} \right\}$$

$$= \eta \cdot (\tilde{\gamma}_{t+1,r} + h) ,$$

where the inequality follows from the induction hypothesis, and the penultimate equality again follows from the definition of $b_{n,k}(s)$. This concludes the induction step, and the proof of Lemma 3. \square

We now return to the proof of Theorem 2. We first show by induction that $V_n^{\bar{\pi}}(x, s) = V_n(x, s), \forall n \in \{1, 2, \dots, N\}, \forall s \in \mathcal{S}$, and $\forall x \in \{0, d, 2d, 3d, \dots\}$.

Base Case: $n = 1$

With one slot remaining, we have:

$$\begin{aligned} V_1(x, s) &= \min_{\{ \max(0, d-x) \leq z_1 \leq \tilde{z}_{\max}(s) \}} \{c(z_1, s) + h(x + z_1 - d)\} \\ &= c(\max\{0, d-x\}, s) + h(\max\{0, (x-d)\}) , \end{aligned}$$

where the minimum is achieved by $z_1 = \max\{0, d-x\}$. $\tilde{\gamma}_{1,1} = \infty$ and $\tilde{\gamma}_{1,2} = 0$, so $b_{1,k}(s) = d$ for all $s \in \mathcal{S}$ and $k \in \{0, 1, \dots, K\}$. Thus, according to (17), $\bar{z}_1(x, s)$ is also equal to $\max\{0, d-x\}$, the optimal amount.

Induction Step

Assume that for $n = \{1, 2, \dots, m-1\}$, $V_n^{\bar{\pi}}(x, s) = V_n(x, s), \forall x \in \{0, d, 2d, 3d, \dots\}, \forall s \in \mathcal{S}$.

We show this is also true for $n = m$ by considering first any strategy that transmits more than $\bar{\pi}$ at time m , and then any strategy that transmits less than $\bar{\pi}$ at time m . Let $s \in \mathcal{S}$ be arbitrary. with $\tilde{\gamma}_{m,j_k+1} \leq \tilde{c}_k(s) < \tilde{\gamma}_{m,j_k}$ so that $\bar{\pi}$ prescribes $b_{m,k}(s) = j_k \cdot d$ for $k \in \{0, 1, \dots, K\}$. Let π^q be a strategy that at time m transmits enough to satisfy the demands of slots $m, m-1, m-2, \dots, q+1$, and q , and transmits optimally at times $m-1, m-2, \dots, 1$.

Part I: Do not transmit more than suggested by $\bar{\pi}$ at time m

Let $\pi'(\epsilon)$ be a feasible strategy with $z'_m = \bar{z}_m + \epsilon$, where $\epsilon > 0$, and the optimal transmission policy at times $m-1, m-2, \dots, 1$. We consider four cases for the current buffer level x .

Case (a): $j_k \cdot d - \tilde{z}_{k-1}(s) < x \leq j_{k-1} \cdot d - \tilde{z}_{k-1}(s)$, $k \in \{0, 1, \dots, K\}$

In this case, $\bar{z}_m = \tilde{z}_{k-1}(s)$. Let p be the integer such that $x + \tilde{z}_{k-1}(s) = p \cdot d$. Let q, η be such that $z'_m = \tilde{z}_{k-1}(s) + \epsilon = q \cdot d + \eta - x$ and $0 \leq \eta < d$ (i.e., $q = \lfloor \frac{z'_m + x}{d} \rfloor$ and $\eta = z'_m + x - q \cdot d$).

Thus, we have $q \geq p \geq j_k$.

Then we have:

$$\begin{aligned} V_m^{\pi'(\epsilon)}(x, s) - V_m^{\pi^q}(x, s) &= c(z'_m, s) - c(z'_m - \eta, s) \\ &\quad + \eta \cdot h \\ &\quad + \alpha \cdot \mathbb{E} \left[V_{m-1}((q-1) \cdot d + \eta, S) - V_{m-1}((q-1) \cdot d, S) \right] \\ &\geq c(z'_m, s) - c(z'_m - \eta, s) \\ &\quad - \eta \cdot \tilde{\gamma}_{m,q+1} \end{aligned} \tag{66}$$

$$\begin{aligned} &\geq c(z'_m, s) - c(z'_m - \eta, s) \\ &\quad - \eta \cdot \tilde{\gamma}_{m,j_k+1} \end{aligned} \tag{67}$$

$$\geq \eta \cdot (\tilde{c}_k(s) - \tilde{\gamma}_{m,j_k+1}) \tag{68}$$

$$\geq 0. \tag{69}$$

Equation (66) follows from Lemma 2, with $l = m, r = q$, and $\eta = \eta$. Equation (67) follows from $q+1 \geq j_k+1$, which implies $\tilde{\gamma}_{m,q+1} \leq \tilde{\gamma}_{m,j_k+1}$. Equation (68) follows from $z'_m - \eta \geq \tilde{z}_{k-1}(s)$

and the construction of $c(\cdot, s)$. Finally, (69) follows from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m, j_k+1}$, by construction of j_k , and we conclude:

$$V_m^{\pi'(\epsilon)}(x, s) \geq V_m^{\pi^q}(x, s) . \quad (70)$$

Now let $t \in \{q+1, q+2, \dots, m-p, m-p+1\}$ be arbitrary. We have:

$$\begin{aligned} V_m^{\pi^{t-1}}(x, s) - V_m^{\pi^t}(x, s) &= c\left((m-t+2) \cdot d - x, s\right) - c\left((m-t+1) \cdot d - x, s\right) \\ &\quad + d \cdot h \\ &\quad + \alpha \cdot \mathbb{E}\left[V_{m-1}\left((m-t+1) \cdot d, S\right) - V_{m-1}\left((m-t) \cdot d, S\right)\right] \\ &\geq c\left((m-t+2) \cdot d - x, s\right) - c\left((m-t+1) \cdot d - x, s\right) \\ &\quad - d \cdot \tilde{\gamma}_{m, m-t+2} \end{aligned} \quad (71)$$

$$\begin{aligned} &\geq c\left((m-t+2) \cdot d - x, s\right) - c\left((m-t+1) \cdot d - x, s\right) \\ &\quad - d \cdot \tilde{\gamma}_{m, j_k+1} \end{aligned} \quad (72)$$

$$\geq d \cdot \left(\tilde{c}_k(s) - \tilde{\gamma}_{m, j_k+1}\right) \quad (73)$$

$$\geq 0. \quad (74)$$

Equation (71) follows from Lemma 2, with $l = m$, $r = m-t+1 \leq m-q \leq m = l$, and $\eta = d$.

Equation (72) follows from:

$$t \leq m-p+1 \Leftrightarrow p+1 \leq m-t+2 \Rightarrow j_k+1 \leq m-t+2 \Rightarrow \tilde{\gamma}_{m, j_k+1} \geq \tilde{\gamma}_{m, m-t+2} .$$

Equation (73) follows from the construction of $c(\cdot, s)$ and the fact that:

$$(m-t+1) \cdot d - x \geq \left[m - (m-p+1) + 1\right] \cdot d - x = p \cdot d - x = \tilde{z}_{k-1}(s) .$$

Finally, (74) follows once again from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m, j_k+1}$, by construction of j_k . Rearranging (74) yields:

$$V_m^{\pi^{t-1}}(x, s) \geq V_m^{\pi^t}(x, s), \quad \forall t \in \{q+1, q+2, \dots, m-p, m-p+1\} . \quad (75)$$

Noting that $V_m^{\bar{\pi}}(x, s) = V_m^{\pi^{m-p+1}}(x, s)$, (70) and repeated application of (75) imply:

$$V_m^{\bar{\pi}}(x, s) = V_m^{\pi^{m-p+1}}(x, s) \leq V_m^{\pi^{m-p}}(x, s) \leq \dots \leq V_m^{\pi^{q+1}}(x, s) \leq V_m^{\pi^q}(x, s) \leq V_m^{\pi'(\epsilon)}(x, s),$$

and we conclude $\bar{\pi}$ is at least as good as $\pi'(\epsilon)$.

Case (b): $j_k \cdot d - \tilde{z}_k(s) < x \leq j_k \cdot d - \tilde{z}_{k-1}(s)$, $k \in \{0, 1, \dots, K-1\}$

Let q, η be such that $z'_m = (m-q+1) \cdot d + \eta - x$ and $0 \leq \eta < d$ (i.e., $q = m+1 - \lfloor \frac{z'_m+x}{d} \rfloor$ and $\eta = z'_m - (m-q+1) \cdot d - x$). Note that $m-q+1 \geq j_k$ by the assumption that $z'_m \geq \bar{z}_m = j_k \cdot d - x$.

Additionally, because $x \leq j_k \cdot d - \tilde{z}_{k-1}(s)$ and $m-q+1 \geq j_k$, we have:

$$(m-q+1) \cdot d - x \geq (m-q+1-j_k) \cdot d + \tilde{z}_{k-1}(s) \geq \tilde{z}_{k-1}(s),$$

which implies:

$$c\left((m-q+1) \cdot d + \eta - x, s\right) - c\left((m-q+1) \cdot d - x, s\right) \geq \eta \cdot \tilde{c}_k(s). \quad (76)$$

Then we have:

$$\begin{aligned} V_m^{\pi'(\epsilon)}(x, s) - V_m^{\pi^q}(x, s) &= c\left((m-q+1) \cdot d + \eta - x, s\right) - c\left((m-q+1) \cdot d - x, s\right) \\ &\quad + \eta \cdot h \\ &\quad + \alpha \cdot \mathbb{E}\left[V_{m-1}\left((m-q) \cdot d + \eta, S\right) - V_{m-1}\left((m-q) \cdot d, S\right)\right] \\ &\geq c\left((m-q+1) \cdot d + \eta - x, s\right) - c\left((m-q+1) \cdot d - x, s\right) \\ &\quad - \eta \cdot \tilde{\gamma}_{m, m-q+2} \end{aligned} \quad (77)$$

$$\begin{aligned} &\geq c\left((m-q+1) \cdot d + \eta - x, s\right) - c\left((m-q+1) \cdot d - x, s\right) \\ &\quad - \eta \cdot \tilde{\gamma}_{m, j_k+1} \end{aligned} \quad (78)$$

$$\geq \eta \cdot \left(\tilde{c}_k(s) - \tilde{\gamma}_{m, j_k+1}\right) \quad (79)$$

$$\geq 0. \quad (80)$$

Equation (77) follows from Lemma 2, with $l = m$, $r = m - q \leq m - 1$, and $\eta = \eta$. Equation

(78) follows from $m - q + 2 \geq j_k + 1$, which implies $\tilde{\gamma}_{m, m-q+2} \leq \tilde{\gamma}_{m, j_k+1}$. Equation (79) follows from (76). Finally, (80) follows from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m, j_k+1}$, by construction of j_k , and we conclude:

$$V_m^{\pi'(\epsilon)}(x, s) \geq V_m^{\pi^q}(x, s) . \quad (81)$$

Now let $t \in \{q + 1, q + 2, \dots, m - j_k, m - j_k + 1\}$ be arbitrary. We have:

$$\begin{aligned} V_m^{\pi^{t-1}}(x, s) - V_m^{\pi^t}(x, s) &= c\left((m - t + 2) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\ &\quad + d \cdot h \\ &\quad + \alpha \cdot \mathbb{E}\left[V_{m-1}\left((m - t + 1) \cdot d, S\right) - V_{m-1}\left((m - t) \cdot d, S\right)\right] \\ &\geq c\left((m - t + 2) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\ &\quad - d \cdot \tilde{\gamma}_{m, m-t+2} \end{aligned} \quad (82)$$

$$\begin{aligned} &\geq c\left((m - t + 2) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\ &\quad - d \cdot \tilde{\gamma}_{m, j_k+1} \end{aligned} \quad (83)$$

$$\geq d \cdot (\tilde{c}_k(s) - \tilde{\gamma}_{m, j_k+1}) \quad (84)$$

$$\geq 0. \quad (85)$$

Equation (82) follows from Lemma 2, with $l = m$, $r = m - t + 1 \leq m - q \leq m = l$, and $\eta = d$.

Equation (83) follows from:

$$t \leq m - j_k + 1 \Leftrightarrow j_k + 1 \leq m - t + 2 \Rightarrow \tilde{\gamma}_{m, j_k+1} \geq \tilde{\gamma}_{m, m-t+2} .$$

Similarly to (76), equation (84) follows from the fact that:

$$(m - t + 1) \cdot d - x \geq \left[m - (m - j_k + 1) + 1 \right] \cdot d - x = j_k \cdot d - x \geq \tilde{z}_{k-1}(s) .$$

Finally, (85) follows once again from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m, j_k+1}$, by construction of j_k . Rearranging (85) yields:

$$V_m^{\pi^{t-1}}(x, s) \geq V_m^{\pi^t}(x, s), \quad \forall t \in \{q + 1, q + 2, \dots, m - j_k, m - j_k + 1\} . \quad (86)$$

Noting that $V_m^{\bar{\pi}}(x, s) = V_m^{\pi^{m-j_k+1}}(x, s)$, (81) and repeated application of (86) imply:

$$V_m^{\bar{\pi}}(x, s) = V_m^{\pi^{m-j_k+1}}(x, s) \leq V_m^{\pi^{m-j_k}}(x, s) \leq \dots \leq V_m^{\pi^{q+1}}(x, s) \leq V_m^{\pi^q}(x, s) \leq V_m^{\pi'(\epsilon)}(x, s),$$

and we conclude $\bar{\pi}$ is at least as good as $\pi'(\epsilon)$.

Case (c): $j_K \cdot d - \tilde{z}_{\max}(s) < x \leq j_K \cdot d - \tilde{z}_{K-1}(s)$

Same as Case (b) with K replacing k .

Case (d): $0 \leq x \leq j_K \cdot d - \tilde{z}_{\max}(s)$

$\bar{z}_m(x, s) = \tilde{z}_{\max}(s)$, the upper bound of the action space, so it is not feasible to transmit more.

Part II: Do not transmit less than suggested by $\bar{\pi}$ at time m

Let $\pi''(\epsilon)$ be a feasible strategy with $z_m'' = \bar{z}_m - \epsilon$, where $\epsilon > 0$, and the optimal transmission policy at times $m-1, m-2, \dots, 1$. To satisfy feasibility, we require $\bar{z}_m - \epsilon \geq \max(0, d-x)$. Define $\eta := \epsilon - \lfloor \frac{\epsilon}{d} \rfloor \cdot d$, and note that $\eta \in [0, d)$. Let π_θ^l be a strategy that at time m satisfies the demands of periods $m, m-1, \dots, l$, except for θ units of the demand of period l , where $0 \leq \theta \leq d$, and behaves optimally in slots $m-1, m-2, \dots, 1$. We consider four exhaustive cases for the current buffer level x .

Case (a): $x > j_0 \cdot d$

$\bar{z}_m(x, s) = 0$, the lower bound of the action space, so it is not feasible to transmit less.

Case (b): $j_k \cdot d - \tilde{z}_k(s) < x \leq j_k \cdot d - \tilde{z}_{k-1}(s)$, $k \in \{0, 1, \dots, K\}$, where we define $\tilde{z}_K(s) := \tilde{z}_{\max}(s)$

Define $q := m - j_k + 1 + \lfloor \frac{\epsilon}{d} \rfloor$. By the feasibility of $\pi''(\epsilon)$ and $\epsilon > 0$, we have

$q \in \{m - j_k + 1, m - j_k + 2, \dots, m - 2, m - 1\}$. Furthermore, we have:

$$[m - q + 1] \cdot d - x = \left[m - \left(m - j_k + 1 + \lfloor \frac{\epsilon}{d} \rfloor \right) + 1 \right] \cdot d - x \leq j_k \cdot d - x \leq \tilde{z}_k(s),$$

which, by the construction of $c(\cdot, s)$, implies:

$$c\left((m - q + 1) \cdot d - \eta - x, s\right) - c\left((m - q + 1) \cdot d - x, s\right) \geq -\eta \cdot \tilde{c}_k(s). \quad (87)$$

We now compare π_η^q and π_0^q :

$$\begin{aligned}
V_m^{\pi_\eta^q}(x, s) - V_m^{\pi_0^q}(x, s) &= c\left((m - q + 1) \cdot d - \eta - x, s\right) - c\left((m - q + 1) \cdot d - x, s\right) \\
&\quad - h \cdot \eta \\
&\quad + \alpha \cdot \mathbb{E}\left[V_{m-1}\left((m - q) \cdot d - \eta, S\right) - V_{m-1}\left((m - q) \cdot d, S\right)\right] \\
&\geq c\left((m - q + 1) \cdot d - \eta - x, s\right) - c\left((m - q + 1) \cdot d - x, s\right) \\
&\quad + \eta \cdot \tilde{\gamma}_{m, m-q+1} \tag{88}
\end{aligned}$$

$$\begin{aligned}
&\geq c\left((m - q + 1) \cdot d - \eta - x, s\right) - c\left((m - q + 1) \cdot d - x, s\right) \\
&\quad + \eta \cdot \tilde{\gamma}_{m, j_k} \tag{89}
\end{aligned}$$

$$\geq \eta \cdot \left[\tilde{\gamma}_{m, j_k} - \tilde{c}_k(s)\right] \tag{90}$$

$$\geq 0. \tag{91}$$

Equation (88) follows from Lemma 3 with $r = m - q + 1 \leq m = l$ and $\eta = \eta$. Equation (89) follows from:

$$q \geq m - j_k + 1 \Leftrightarrow m - q + 1 \leq j_k \Rightarrow \tilde{\gamma}_{m, j_k} \leq \tilde{\gamma}_{m, m-q+1}.$$

Equation (90) follows from (87). Finally, (91) follows from $\tilde{c}_k(s) < \tilde{\gamma}_{m, j_k}$. Rearranging (91) yields:

$$V_m^{\pi_0^q}(x, s) \leq V_m^{\pi_\eta^q}(x, s). \tag{92}$$

Next, let $t \in \{m - j_k + 1, m - j_k + 2, \dots, m - 1\}$ be arbitrary. We have:

$$\begin{aligned}
V_m^{\pi_0^{t+1}}(x, s) - V_m^{\pi_0^t}(x, s) &= c\left((m - t) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\
&\quad - h \cdot d \\
&\quad + \alpha \cdot \mathbb{E}\left[V_{m-1}\left((m - t - 1) \cdot d, S\right) - V_{m-1}\left((m - t) \cdot d, S\right)\right] \\
&\geq c\left((m - t) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\
&\quad + d \cdot \tilde{\gamma}_{m, m-t+1} \tag{93}
\end{aligned}$$

$$\begin{aligned}
&\geq c\left((m - t) \cdot d - x, s\right) - c\left((m - t + 1) \cdot d - x, s\right) \\
&\quad + d \cdot \tilde{\gamma}_{m, j_k} \tag{94}
\end{aligned}$$

$$\geq d \cdot \left[\tilde{\gamma}_{m, j_k} - \tilde{c}_k(s)\right] \tag{95}$$

$$\geq 0. \tag{96}$$

Equation (93) follows from Lemma 3 with $r = m - t \leq m = l$ and $\eta = d$. Equation (94) follows from:

$$t \geq m - j_k + 1 \Leftrightarrow m - t + 1 \leq j_k \Rightarrow \tilde{\gamma}_{m, j_k} \leq \tilde{\gamma}_{m-t+1}.$$

Equation (95) follows from construction of $c(\cdot, s)$ and the fact that:

$$(m - t + 1) \cdot d - x \leq \left(m - (m - j_k + 1) + 1\right) \cdot d - x = j_k \cdot d - x \leq \tilde{z}_k(s).$$

Finally, (96) follows from $\tilde{c}_k(s) < \tilde{\gamma}_{m, j_k}$. Rearranging (96) yields:

$$V_m^{\pi_0^t}(x, s) \leq V_m^{\pi_0^{t+1}}(x, s) \quad \forall t \in \{m - j_k + 1, m - j_k + 2, \dots, m - 1\}. \tag{97}$$

Noting that $\bar{\pi} = \pi_0^{m-j_k+1}$, (92) and repeated application of (97) imply:

$$\begin{aligned}
V_m^{\bar{\pi}}(x, s) &= V_m^{\pi_0^{m-j_k+1}}(x, s) \\
&\leq V_m^{\pi_0^{m-j_k+2}}(x, s) \leq \dots \leq V_m^{\pi_0^q}(x, s) \leq V_m^{\pi_\eta^q}(x, s) = V_m^{\pi''(\epsilon)}(x, s), \tag{98}
\end{aligned}$$

and we conclude $\bar{\pi}$ is at least as good as $\pi''(\epsilon)$.

Case (c): $j_k \cdot d - \tilde{z}_{k-1}(s) < x \leq j_{k-1} \cdot d - \tilde{z}_{k-1}(s)$, $k \in \{1, \dots, K\}$

In this case, $\bar{\pi} = \pi_0^{m+1 - \frac{x + \tilde{z}_{k-1}(s)}{d}}$. Define $p := m + 1 - \frac{x + \tilde{z}_{k-1}(s)}{d}$, and $q := p + \lfloor \frac{\epsilon}{d} \rfloor$.

Again, we start by comparing π_η^q and π_0^q :

$$\begin{aligned} V_m^{\pi_\eta^q}(x, s) - V_m^{\pi_0^q}(x, s) &= c(z_m'' - \eta, s) - c(z_m'', s) \\ &\quad - h \cdot \eta \\ &\quad + \alpha \cdot \mathbb{E} \left[V_{m-1}((m-q) \cdot d - \eta, S) - V_{m-1}((m-q) \cdot d, S) \right] \\ &\geq c(z_m'' - \eta, s) - c(z_m'', s) \\ &\quad + \eta \cdot \tilde{\gamma}_{m, m-q+1} \end{aligned} \tag{99}$$

$$\begin{aligned} &\geq c(z_m'' - \eta, s) - c(z_m'', s) \\ &\quad + \eta \cdot \tilde{\gamma}_{m, j_{k-1}} \end{aligned} \tag{100}$$

$$\geq \eta \cdot \left[\tilde{\gamma}_{m, j_{k-1}} - \tilde{c}_{k-1}(s) \right] \tag{101}$$

$$\geq 0. \tag{102}$$

Equation (99) follows from Lemma 3 with $r = m - q + 1 \leq m = l$ and $\eta = \eta$. Equation (100) follows from:

$$m - q + 1 = m - \left(p + \left\lfloor \frac{\epsilon}{d} \right\rfloor \right) + 1 = \frac{x + \tilde{z}_{k-1}(s)}{d} - \left\lfloor \frac{\epsilon}{d} \right\rfloor \leq \frac{x + \tilde{z}_{k-1}(s)}{d} \leq j_{k-1},$$

which implies $\tilde{\gamma}_{m, j_{k-1}} \leq \tilde{\gamma}_{m, m-q+1}$. Equation (101) follows from $z_m'' < \tilde{z}_{k-1}(s)$ and the construction of $c(\cdot, s)$. Finally, (102) follows from $\tilde{c}_{k-1}(s) < \tilde{\gamma}_{m, j_{k-1}}$. Rearranging (102) yields:

$$V_m^{\pi_0^q}(x, s) \leq V_m^{\pi_\eta^q}(x, s). \tag{103}$$

Next, let $\hat{t} \in \{p, p+1, \dots, q-1\}$ be arbitrary. We have:

$$\begin{aligned}
V_m^{\pi_0^{\hat{t}+1}}(x, s) - V_m^{\pi_0^{\hat{t}}}(x, s) &= c\left((m - \hat{t}) \cdot d - x, s\right) - c\left((m - \hat{t} + 1) \cdot d - x, s\right) \\
&\quad - h \cdot d \\
&\quad + \alpha \cdot \mathbf{E}\left[V_{m-1}\left((m - \hat{t} - 1) \cdot d, S\right) - V_{m-1}\left((m - \hat{t}) \cdot d, S\right)\right] \\
&\geq c\left((m - \hat{t}) \cdot d - x, s\right) - c\left((m - \hat{t} + 1) \cdot d - x, s\right) \\
&\quad + d \cdot \tilde{\gamma}_{m, m-\hat{t}+1} \tag{104}
\end{aligned}$$

$$\begin{aligned}
&\geq c\left((m - \hat{t}) \cdot d - x, s\right) - c\left((m - \hat{t} + 1) \cdot d - x, s\right) \\
&\quad + d \cdot \tilde{\gamma}_{m, j_{k-1}} \tag{105}
\end{aligned}$$

$$\geq d \cdot \left[\tilde{\gamma}_{m, j_{k-1}} - \tilde{c}_{k-1}(s)\right] \tag{106}$$

$$\geq 0. \tag{107}$$

Equation (104) follows from Lemma 3 with $r = m - \hat{t} \leq m = l$ and $\eta = d$. Equation (105) follows from:

$$\hat{t} \geq p \Rightarrow m - \hat{t} + 1 \leq m - p + 1 = \frac{x + \tilde{z}_{k-1}(s)}{d} \leq j_{k-1} \Rightarrow \tilde{\gamma}_{m, j_{k-1}} \leq \tilde{\gamma}_{m-\hat{t}+1}.$$

Equation (106) follows from construction of $c(\cdot, s)$ and the fact that:

$$(m - \hat{t} + 1) \cdot d - x \leq (m - p + 1) \cdot d - x = \tilde{z}_{k-1}(s).$$

Finally, (107) follows from $\tilde{c}_{k-1}(s) < \tilde{\gamma}_{m, j_{k-1}}$. Rearranging (107) yields:

$$V_m^{\pi_0^{\hat{t}}}(x, s) \leq V_m^{\pi_0^{\hat{t}+1}}(x, s) \quad \forall \hat{t} \in \{p, p+1, \dots, q-1\}. \tag{108}$$

Then (103) and repeated application of (108) yield:

$$\begin{aligned}
V_m^{\bar{\pi}}(x, s) &= V_m^{\pi_0^p}(x, s) \\
&\leq V_m^{\pi_0^{p+1}}(x, s) \leq \dots \leq V_m^{\pi_0^{q-1}}(x, s) \leq V_m^{\pi_0^q}(x, s) \leq V_m^{\pi_q}(x, s) = V_m^{\pi''(\epsilon)}(x, s),
\end{aligned}$$

and we conclude $\bar{\pi}$ is at least as good as $\pi''(\epsilon)$.

Case (d): $0 \leq x \leq j_K \cdot d - \tilde{z}_{\max}$

The same argument as Case (c) applies with k replaced by $K + 1$ and $\tilde{z}_K(s) = \tilde{z}_{\max}(s)$. This completes Part II.

From Parts I and II, we conclude $\bar{\pi}$ is optimal if the starting queue level is an integer multiple of the demand d . By assumption, the starting queue level x at time N is zero. Thus, $\bar{\pi}$ is optimal at time N . $z_N^*(x, s) = \bar{z}_N(x, s)$ will also be an integer multiple of demand as $b_{N,k}(s)$, and $\{\tilde{z}_k(s)\}_{k=0,1,\dots,K}$ are all integer multiples of d . It follows that the queue level at the end of slot N (equal to the queue level at the beginning of slot $N - 1$), $z_N^*(x, s) - d$, will also be an integer multiple of d . Continuing this logic, if the strategy $\bar{\pi}$ is used, the queue level at the beginning of each subsequent time slot will be an integer multiple of demand. Thus, $\bar{\pi}$ is optimal. \square

D. Proof of Theorem 7

We prove statements (i)-(v) by joint induction on the time remaining, n .

Base Case: $n = 1$

$V_0(\mathbf{x}, \mathbf{s}_0) = 0$, for all \mathbf{s}_0 , so (i) and (ii) hold trivially. Let $\mathbf{s}_1 \in \mathcal{S}$ be arbitrary. $G_1(\mathbf{y}_1, \mathbf{s}_1) = \mathbf{c}_{\mathbf{s}_1}^T \mathbf{y}_1 + h(\mathbf{y}_1 - \mathbf{d})$, which is convex and supermodular. Thus, (iii) and (iv) are true. Additionally, $G_1(\mathbf{y}_1, \mathbf{s}_1) = \sum_{m=1}^2 \{c_s^m \cdot y_1^m + h^m(y_1^m - d^m)\}$, so $\inf \left\{ \operatorname{argmin}_{y_1^2 \in [d^2, \infty)} \left\{ G_1(y_1^1, y_1^2, s_1^1, s_1^2) \right\} \right\}$ is independent of y_1^1 , and vice versa. Thus, (v) is true for $n = 1$, completing the base case.

Induction Step

Assume statements (i)-(v) are true for $n = 2, 3, \dots, l - 1$. We want to show they are true for $n = l$. We let $\mathbf{s} \in \mathcal{S}$ be arbitrary, and proceed in order.

- (i) Consider two arbitrary points, $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in \mathbb{R}_+^2$. Let $\lambda \in [0, 1]$ be arbitrary, and define $\hat{\mathbf{x}} := \lambda \bar{\mathbf{x}} + (1 - \lambda) \tilde{\mathbf{x}}$. Let $\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s})$, $\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s})$, and $\mathbf{y}^*(\hat{\mathbf{x}}, \mathbf{s})$ be optimal buffer levels after transmission

in slot $l - 1$, for each of the respective starting points. We have:

$$\begin{aligned}
\lambda \cdot V_{l-1}(\bar{\mathbf{x}}, \mathbf{s}) + (1 - \lambda) \cdot V_{l-1}(\tilde{\mathbf{x}}, \mathbf{s}) &= -\mathbf{c}_s^T \hat{\mathbf{x}} + \lambda \cdot G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \\
&\quad + (1 - \lambda) \cdot G_{l-1}(\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \\
&\geq -\mathbf{c}_s^T \hat{\mathbf{x}} + G_{l-1}(\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1 - \lambda) \mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \\
&\geq -\mathbf{c}_s^T \hat{\mathbf{x}} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\hat{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \\
&= V_{l-1}(\hat{\mathbf{x}}, \mathbf{s}) = V_{l-1}(\lambda \bar{\mathbf{x}} + (1 - \lambda) \tilde{\mathbf{x}}, \mathbf{s}), \quad (109)
\end{aligned}$$

where the first inequality follows from the convexity of $G_{l-1}(\cdot, \mathbf{s})$ from the induction hypothesis. The second inequality follows from the following argument. $\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$ implies:

$$\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee \bar{\mathbf{x}} \quad \text{and} \quad \mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}}] \leq P. \quad (110)$$

Similarly, $\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})$ implies:

$$\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee \tilde{\mathbf{x}} \quad \text{and} \quad \mathbf{c}_s^T [\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}] \leq P. \quad (111)$$

Multiplying the equations in (110) by λ and the equations in (111) by $1 - \lambda$, and summing, we have:

$$\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1 - \lambda) \mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \succeq \lambda(\mathbf{d} \vee \bar{\mathbf{x}}) + (1 - \lambda)(\mathbf{d} \vee \tilde{\mathbf{x}}) \succeq \mathbf{d} \vee \hat{\mathbf{x}}, \quad (112)$$

and

$$\begin{aligned}
&\mathbf{c}_s^T [\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1 - \lambda) \mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \hat{\mathbf{x}}] \\
&= \lambda \mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}}] + (1 - \lambda) \mathbf{c}_s^T [\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}] \leq P. \quad (113)
\end{aligned}$$

From (112) and (113), we conclude $\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1 - \lambda) \mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\hat{\mathbf{x}}, \mathbf{s})$, as shown in Figure 8. Thus, the value of $G_{l-1}(\cdot, \mathbf{s})$ at this point is greater than or equal to the minimum of $G_l(\cdot, \mathbf{s})$ over the region $\tilde{\mathcal{A}}^d(\hat{\mathbf{x}}, \mathbf{s})$. From (109), we conclude $V_{l-1}(\cdot, \mathbf{s})$ is convex. This is

one such choice of $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$, and we have:

$$\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \succ \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}) . \quad (115)$$

Further, $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ implies $\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}] \leq P$, and thus:

$$\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \vee \tilde{\mathbf{x}}] \leq \mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}] \leq P . \quad (116)$$

Equations (115) and (116) imply $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, and thus:

$$G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) = \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \leq G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) . \quad (117)$$

However, we also have:

$$\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}) \succeq \mathbf{d} \vee (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) , \quad (118)$$

and

$$\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}] \leq \mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}] \leq P . \quad (119)$$

Equations (118) and (119) imply $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$, which, in combination with (117), implies it is optimal to move from $\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}$ to $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, contradicting the assumption that $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \succ \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ for all possible choices of $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$. \square

Now let $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ be arbitrary optimal actions such that $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \not\succeq \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. We show (114) by considering two exhaustive cases.

Case 1: $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$

We start with another lemma.

Lemma 5. *Let $f : [d^1, \infty) \times [d^2, \infty) \rightarrow \mathbb{R}$ be convex and supermodular, let $\sigma, \beta \in [0, 1]$ be arbitrary, and let $\mathbf{z} = (z_1, z_2) \preceq (\hat{z}_1, \hat{z}_2) = \hat{\mathbf{z}}$. Define $\mathbf{z}^{\lambda_1, \lambda_2} := (\lambda_1 \hat{z}_1 + (1 - \lambda_1)z_1, \lambda_2 \hat{z}_2 + (1 - \lambda_2)z_2)$. Then*

$$f(\mathbf{z}) + f(\hat{\mathbf{z}}) \geq f(\mathbf{z}^{\sigma, \beta}) + f(\mathbf{z}^{1-\sigma, 1-\beta}) . \quad (120)$$

Proof.

Step 1: Assume $\sigma, \beta \leq \frac{1}{2}$. Assume without loss of generality that $\sigma \leq \beta$. By the convexity of $f(\cdot)$, we have:

$$f(\mathbf{z}) + f(\hat{\mathbf{z}}) \geq f(\mathbf{z}^{\sigma, \sigma}) + f(\mathbf{z}^{1-\sigma, 1-\sigma}) , \quad (121)$$

and

$$f(\mathbf{z}^{1-\sigma, 1-\sigma}) + f(\mathbf{z}^{1-\sigma, \sigma}) \geq f(\mathbf{z}^{1-\sigma, \beta}) + f(\mathbf{z}^{1-\sigma, 1-\beta}) . \quad (122)$$

By the supermodularity of $f(\cdot)$, we have:

$$f(\mathbf{z}^{1-\sigma, \beta}) + f(\mathbf{z}^{\sigma, \sigma}) \geq f(\mathbf{z}^{\sigma, \beta}) + f(\mathbf{z}^{1-\sigma, \sigma}) . \quad (123)$$

Figure 9 shows these relationships. Combining (121)-(123), we have:

$$\begin{aligned} f(\mathbf{z}) + f(\hat{\mathbf{z}}) &\geq f(\mathbf{z}^{\sigma, \sigma}) + f(\mathbf{z}^{1-\sigma, 1-\sigma}) \\ &\geq f(\mathbf{z}^{\sigma, \sigma}) + f(\mathbf{z}^{1-\sigma, \beta}) - f(\mathbf{z}^{1-\sigma, \sigma}) + f(\mathbf{z}^{1-\sigma, 1-\beta}) \\ &\geq f(\mathbf{z}^{\sigma, \beta}) + f(\mathbf{z}^{1-\sigma, 1-\beta}) . \end{aligned}$$

Step 2: Now let $\sigma, \beta \in [0, 1]$, and define $\hat{\sigma} := \min\{\sigma, 1 - \sigma\}$ and $\hat{\beta} := \min\{\beta, 1 - \beta\}$.

Then $\hat{\sigma}, \hat{\beta} \leq \frac{1}{2}$, so by Step 1, we have:

$$f(\mathbf{z}) + f(\hat{\mathbf{z}}) \geq f(\mathbf{z}^{\hat{\sigma}, \hat{\beta}}) + f(\mathbf{z}^{1-\hat{\sigma}, 1-\hat{\beta}}) . \quad (124)$$

Note that $\mathbf{z}^{\sigma, \beta} \wedge \mathbf{z}^{1-\sigma, 1-\beta} = \mathbf{z}^{\hat{\sigma}, \hat{\beta}}$, and $\mathbf{z}^{\sigma, \beta} \vee \mathbf{z}^{1-\sigma, 1-\beta} = \mathbf{z}^{1-\hat{\sigma}, 1-\hat{\beta}}$, so by the supermodularity of $f(\cdot)$, we have:

$$f(\mathbf{z}^{\hat{\sigma}, \hat{\beta}}) + f(\mathbf{z}^{1-\hat{\sigma}, 1-\hat{\beta}}) \geq f(\mathbf{z}^{\sigma, \beta}) + f(\mathbf{z}^{1-\sigma, 1-\beta}) . \quad (125)$$

Combining (124) and (125) yields the desire result, (120). □

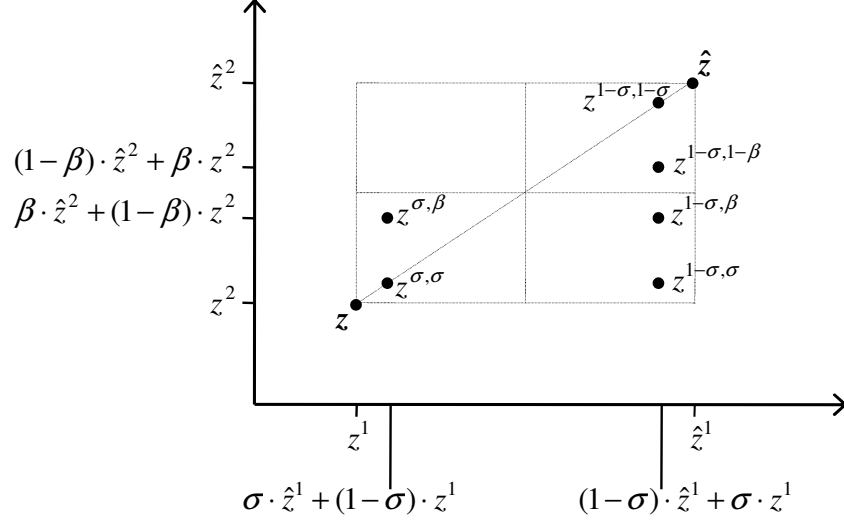


Fig. 9. Diagram of the points referred to in Step 1 of the proof of Lemma 5

Next, define the following points, shown in Figure 10:

$$\bar{\mathbf{y}} := \left(\begin{array}{c} \bar{x}^1 + \max \left\{ y^{*1} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^1, y^{*1} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^1 \right\}, \\ \bar{x}^2 + \min \left\{ y^{*2} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^2, y^{*2} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^2 \right\} \end{array} \right), \text{ and}$$

$$\tilde{\mathbf{y}} := \left(\begin{array}{c} \tilde{x}^1 + \min \left\{ y^{*1} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^1, y^{*1} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^1 \right\}, \\ \tilde{x}^2 + \max \left\{ y^{*2} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^2, y^{*2} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^2 \right\} \end{array} \right).$$

Note that $\bar{\mathbf{y}} \succeq \mathbf{d} \vee \bar{\mathbf{x}}$ and $\tilde{\mathbf{y}} \succeq \mathbf{d} \vee \tilde{\mathbf{x}}$. Furthermore, we have:

$$\begin{aligned} \mathbf{c}_s^T (\bar{\mathbf{y}} - \bar{\mathbf{x}}) &= \mathbf{c}_s^T \left(\begin{array}{c} \max \left\{ y^{*1} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^1, y^{*1} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^1 \right\}, \\ \min \left\{ y^{*2} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^2, y^{*2} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^2 \right\} \end{array} \right) \\ &\leq \max \left\{ \begin{array}{c} \mathbf{c}_s^T \left(y^{*1} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^1, y^{*2} (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^2 \right), \\ \mathbf{c}_s^T \left(y^{*1} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{x}^1, y^{*2} (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{x}^2 \right) \end{array} \right\} \\ &= \max \left\{ \mathbf{c}_s^T \left(\mathbf{y}^* (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) \right), \mathbf{c}_s^T \left(\mathbf{y}^* (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}) \right) \right\} \leq P. \end{aligned}$$

By a similar argument, $\mathbf{c}_s^T (\tilde{\mathbf{y}} - \tilde{\mathbf{x}}) \leq P$, and thus $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$, and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})$. So we

have:

$$\min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \leq G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) . \quad (126)$$

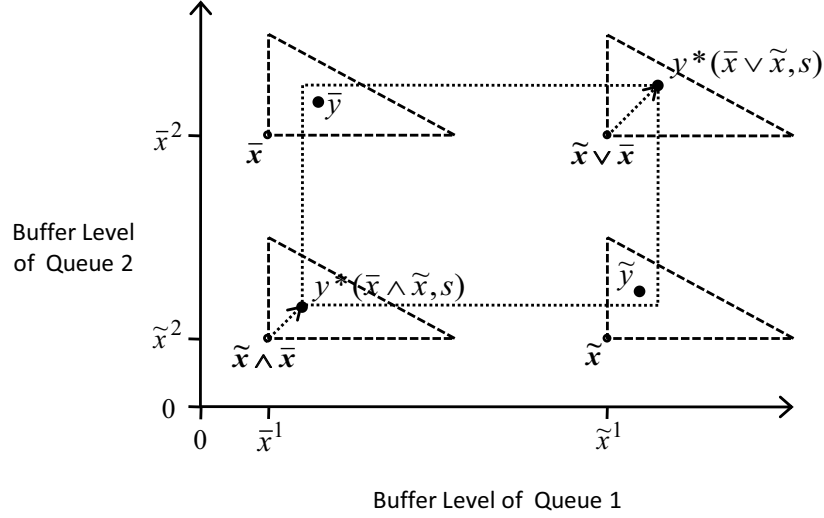


Fig. 10. Construction of feasible points $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ in Case 1 of the proof of supermodularity of $V_{l-1}(\cdot, \mathbf{s})$.

Now define¹²:

$$\sigma := \frac{y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{y}^1}{y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} , \text{ and}$$

$$\beta := \frac{y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{y}^2}{y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} .$$

Rearranging the definitions of σ and β yields:

$$\tilde{\mathbf{y}} = \left((1 - \sigma) \cdot y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) + \sigma \cdot y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), (1 - \beta) \cdot y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) + \beta \cdot y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \right) .$$

It is also straightforward to check that:

$$\bar{\mathbf{y}} = \left(\sigma \cdot y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) + (1 - \sigma) \cdot y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \beta \cdot y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) + (1 - \beta) \cdot y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \right) .$$

¹²If $y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) = 0$, let σ be arbitrary in $[0, 1]$. Similarly, if $y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) = 0$, let β be arbitrary in $[0, 1]$.

Note also that

$$\begin{aligned}
y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) &= \min \left\{ y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (\tilde{x}^1 - \tilde{x}^2) \right\} \\
&\leq \min \left\{ y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (\tilde{x}^1 - \tilde{x}^2) \right\} \\
&= \tilde{y}^1 \\
&\leq y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) ,
\end{aligned}$$

and thus, $\sigma \in [0, 1]$. Similarly, $y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq \tilde{y}^2 \leq y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, and thus, $\beta \in [0, 1]$.

Since $G_{l-1}(\cdot, \mathbf{s})$ is convex and supermodular, we can now apply Lemma 5, with $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ playing the role of \mathbf{z} ; $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ the role of $\hat{\mathbf{z}}$; $\bar{\mathbf{y}}$ the role of $\mathbf{z}^{\sigma, \beta}$; and $\tilde{\mathbf{y}}$ the role of $\mathbf{z}^{1-\sigma, 1-\beta}$, to get:

$$\begin{aligned}
G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) &\leq G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) + G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \\
&= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} . \quad (127)
\end{aligned}$$

Combining equations (126) and (127) yields the desired result, (114).

Case 2: $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \not\leq \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \not\leq \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$

There are two possibilities for this case. The first possibility is that $y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ and $y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. The second possibility is that $y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ and $y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. We show (114) under the first possibility, and a symmetric argument can be used to show (114) under the second possibility. We have:

$$y^{*1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \geq \max \{(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}})^1, d^1\} = \max \{\tilde{x}^1, d^1\} , \quad (128)$$

$$y^{*2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \geq \max \{(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})^2, d^2\} = \max \{\tilde{x}^2, d^2\} , \quad (129)$$

and

$$\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}] \leq \mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})] \leq P . \quad (130)$$

Equations (128), (129), and (130) imply $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})$. If it also happens that $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$, then we have:

$$\begin{aligned} & \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \\ & \leq G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) + G_{l-1}(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}) \\ & = \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\}. \end{aligned}$$

Otherwise, define:

$$\gamma := \frac{\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}}] - P}{\mathbf{c}_s^T [\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})]} .$$

From $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \notin \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$, we know:

$$\mathbf{c}_s^T \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) > \mathbf{c}_s^T \bar{\mathbf{x}} + P \geq \mathbf{c}_s^T (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) + P \geq \mathbf{c}_s^T \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) . \quad (131)$$

It is clear from (131) that the numerator and denominator of γ are positive, and $\gamma \in [0, 1]$.

Now define:

$$\begin{aligned} \bar{\mathbf{y}} & := \gamma \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (1 - \gamma) \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) , \text{ and} \\ \tilde{\mathbf{y}} & := (1 - \gamma) \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + \gamma \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) . \end{aligned}$$

It is somewhat tedious but straightforward to show that $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$, and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})$.

Thus, we have:

$$\min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \leq G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) . \quad (132)$$

In Figure 11, $\bar{\mathbf{y}}$ is the point where the line segment connecting $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ intersects the budget constraint (hypotenuse) of $\tilde{\mathcal{A}}^d(\bar{\mathbf{x}}, \mathbf{s})$, and $\tilde{\mathbf{y}}$ is a point along this line segment the same distance from $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ as $\bar{\mathbf{y}}$ is from $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. By the convexity of

$G_{l-1}(\cdot, \mathbf{s})$ along this line segment, we have:

$$\begin{aligned} G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) &\leq G_{l-1}\left(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\right) + G_{l-1}\left(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\right) \\ &= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\}. \end{aligned} \quad (133)$$

Combining (132) and (133) yields the desired result, (114).

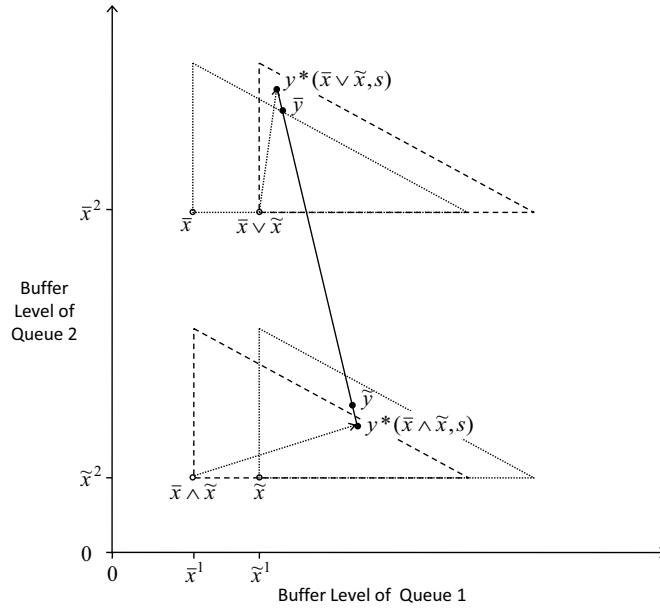


Fig. 11. Construction of feasible points $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ in Case 2 of the proof of supermodularity of $V_{l-1}(\cdot, \mathbf{s})$.

- (iii) $G_l(\mathbf{y}, \mathbf{s}) = \mathbf{c}_s^T \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}[V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{S})]$. By (i), for all \mathbf{s} , $V_{l-1}(\mathbf{x}, \mathbf{s})$ is convex in \mathbf{x} ; thus, $V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{s})$ is convex in \mathbf{y} as it is the composition of a convex function with an affine function. $\mathbb{E}[V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{S})]$ is also convex as it is the nonnegative weighted sum/integral of convex functions. It follows that $G_l(\mathbf{y}, \mathbf{s})$, the sum of convex functions, is convex in \mathbf{y} .
- (iv) Supermodularity of $G_l(\mathbf{y}, \mathbf{s})$ follows from the same series of arguments as (iii), because, like convexity, supermodularity is preserved under addition and scalar multiplication (Smith and McCardle refer to these as *closed convex cone properties* [78]).
- (v) This step basically follows from Topkis' Theorem 2.8.1 [76, pg. 76], but, for the reader's benefit, we reproduce the proof here with our notation. Let $y^2, \hat{y}^2 \in [d^2, \infty)$ be arbitrary with

$y^2 < \hat{y}^2$. Let $\bar{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1, \infty)} \{G_l(y^1, y^2, \mathbf{s})\}$ and $\tilde{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1, \infty)} \{G_l(y^1, \hat{y}^2, \mathbf{s})\}$ be arbitrary. We want to show:

$$\bar{y}^1 \wedge \tilde{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1, \infty)} \{G_l(y^1, \hat{y}^2, \mathbf{s})\} .$$

If $\tilde{y}^1 \leq \bar{y}^1$, this is trivial, so we check that it is true for $\tilde{y}^1 > \bar{y}^1$. Since \bar{y}^1 is a minimizer of $G_l(\cdot, y^2, \mathbf{s})$, we have:

$$G_l(\bar{y}^1, y^2, \mathbf{s}) \leq G_l(\tilde{y}^1, y^2, \mathbf{s}) , \quad (134)$$

and since \tilde{y}^1 is a minimizer of $G_l(\cdot, \hat{y}^2, \mathbf{s})$, we have:

$$G_l(\tilde{y}^1, \hat{y}^2, \mathbf{s}) \leq G_l(\bar{y}^1, \hat{y}^2, \mathbf{s}) . \quad (135)$$

By the supermodularity of $G_l(\cdot, \mathbf{s})$, we have:

$$\begin{aligned} G_l(\tilde{y}^1, y^2, \mathbf{s}) + G_l(\bar{y}^1, \hat{y}^2, \mathbf{s}) &\leq G_l(\tilde{y}^1 \wedge \bar{y}^1, y^2 \wedge \hat{y}^2, \mathbf{s}) + G_l(\tilde{y}^1 \vee \bar{y}^1, y^2 \vee \hat{y}^2, \mathbf{s}) \\ &= G_l(\bar{y}^1, y^2, \mathbf{s}) + G_l(\tilde{y}^1, \hat{y}^2, \mathbf{s}) , \end{aligned}$$

or, rearranging terms:

$$G_l(\tilde{y}^1, y^2, \mathbf{s}) - G_l(\bar{y}^1, y^2, \mathbf{s}) \leq G_l(\tilde{y}^1, \hat{y}^2, \mathbf{s}) - G_l(\bar{y}^1, \hat{y}^2, \mathbf{s}) . \quad (136)$$

Combining (134), (135), and (136) yields:

$$0 \leq G_l(\tilde{y}^1, y^2, \mathbf{s}) - G_l(\bar{y}^1, y^2, \mathbf{s}) \leq G_l(\tilde{y}^1, \hat{y}^2, \mathbf{s}) - G_l(\bar{y}^1, \hat{y}^2, \mathbf{s}) \leq 0 . \quad (137)$$

So (137) holds with equality throughout, implying $G_l(\tilde{y}^1, \hat{y}^2, \mathbf{s}) = G_l(\bar{y}^1, \hat{y}^2, \mathbf{s})$, and we conclude:

$$\tilde{y}^1 \wedge \bar{y}^1 = \bar{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1, \infty)} \{G_l(y^1, \hat{y}^2, \mathbf{s})\} .$$

Since \bar{y}^1 and \tilde{y}^1 were chosen arbitrarily, we have:

$$\inf \left\{ \operatorname{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n(y_n^1, y_n^2, s^1, s^2) \right\} \right\} \geq \inf \left\{ \operatorname{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n(y_n^1, \hat{y}_n^2, s^1, s^2) \right\} \right\}.$$

The first implication in (v) follows from a symmetric argument. \square

E. Proof of Theorem 8

Let $n \in \{1, 2, \dots, N\}$ and $\mathbf{s} \in \mathcal{S}$ be arbitrary. We start by proving (24). First, let $\mathbf{x} \in \mathcal{R}_I(n, \mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$ be arbitrary. We know from Theorem 7 that $G_n(\cdot, \mathbf{s})$ is convex on $[d^1, \infty) \times [d^2, \infty)$, which implies that $G_n(\cdot, \mathbf{s})$ is also convex on any line segment in $[d^1, \infty) \times [d^2, \infty)$ (see, e.g., [77, Theorem 4.1]). Specifically, by the convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^1 = \hat{y}^1$ and the fact that $\hat{y}^2 \geq x^2 \geq f_n^2(\hat{y}^1, \mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n((\hat{y}^1, x^2), \mathbf{s}) \geq G_n((\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})), \mathbf{s}). \quad (138)$$

Similarly, by the convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^2 = x^2$ and the fact that $\hat{y}^1 \geq x^1 \geq f_n^1(x^2, \mathbf{s})$, we have:

$$G_n((\hat{y}^1, x^2), \mathbf{s}) \geq G_n(\mathbf{x}, \mathbf{s}) \geq G_n((f_n^1(x^2, \mathbf{s}), x^2), \mathbf{s}). \quad (139)$$

Combining (138) and (139) yields:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n((\hat{y}^1, x^2), \mathbf{s}) \geq G_n(\mathbf{x}, \mathbf{s}),$$

and we conclude $G_n(\mathbf{x}, \mathbf{s}) = \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \{G_n(\mathbf{y}, \mathbf{s})\}$.

Second, let $\mathbf{x} \in \mathcal{R}_{II}(n, \mathbf{s})$ be arbitrary. Then $\mathbf{b}_n(\mathbf{s}) \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$ and $\mathbf{b}_n(\mathbf{s})$ is a global minimizer of $G_n(\cdot, \mathbf{s})$, so it is clearly optimal to transmit to bring the receivers' buffer levels up to $\mathbf{b}_n(\mathbf{s})$.

Next, let $\mathbf{x} \in \mathcal{R}_{III-A}(n, \mathbf{s})$ and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$ be arbitrary. By definition of $f_n^1(\cdot, \mathbf{s})$, we have:

$$G_n(\tilde{\mathbf{y}}, \mathbf{s}) \geq G_n((f_n^1(\tilde{y}^2, \mathbf{s}), \tilde{y}^2), \mathbf{s}). \quad (140)$$

Furthermore, the function $\min_{y^1 \in [d^1, \infty)} \{G_n((y^1, y^2), \mathbf{s})\}$ is convex in y^2 since $[d^1, \infty)$ is a

convex set (see, e.g., [75, pp. 101-102]). Thus, $\tilde{y}^2 \geq x^2 \geq b_n^2(\mathbf{s})$ implies:

$$\begin{aligned}
& G_n\left((f_n^1(\tilde{y}^2, \mathbf{s}), \tilde{y}^2), \mathbf{s}\right) \\
& \geq G_n\left((f_n^1(x^2, \mathbf{s}), x^2), \mathbf{s}\right) \\
& \geq G_n\left((f_n^1(b_n^2(\mathbf{s}), \mathbf{s}), b_n^2(\mathbf{s})), \mathbf{s}\right) \\
& = G_n(\mathbf{b}_n(\mathbf{s}), \mathbf{s}) .
\end{aligned} \tag{141}$$

Combining (140) and (141) yields:

$$G_n(\tilde{\mathbf{y}}, \mathbf{s}) \geq G_n\left((f_n^1(x^2, \mathbf{s}), x^2), \mathbf{s}\right) ,$$

and $\mathbf{x} \in \mathcal{R}_{III-A}(n, \mathbf{s})$ implies $(f_n^1(x^2, \mathbf{s}), x^2) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$. Since $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ was arbitrary, we conclude $\mathbf{y}_n^*(\mathbf{x}, \mathbf{s}) = (f_n^1(x^2, \mathbf{s}), x^2)$ is optimal.

The optimality of $\mathbf{y}_n^*(\mathbf{x}, \mathbf{s}) = (x^1, f_n^2(x^1, \mathbf{s}))$ for $\mathbf{x} \in \mathcal{R}_{III-B}(n, \mathbf{s})$ follows from a symmetric argument, using the convexity of $G_n(\cdot, \mathbf{s})$ along the curve $(x^1, f_n^2(x^1, \mathbf{s}))$.

Finally, we prove (25). Define:

$$\mathcal{H}^{\mathbf{d}}(\mathbf{x}, \mathbf{s}) := \left\{ \mathbf{y} \in [d^1, \infty) \times [d^2, \infty) : \mathbf{y} \succeq \mathbf{x} \text{ and } \mathbf{c}_s^{\mathbf{T}}[\mathbf{y} - \mathbf{x}] = P \right\} \subset \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s}) .$$

First, let $\mathbf{x} \in \mathcal{R}_{IV-B}(n, \mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ be arbitrary such that $\mathbf{c}_s^{\mathbf{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$. Define

$$\lambda_0 := \frac{\mathbf{c}_s^{\mathbf{T}}\mathbf{b}_n(\mathbf{s}) - \mathbf{c}_s^{\mathbf{T}}\mathbf{x} - P}{\mathbf{c}_s^{\mathbf{T}}\mathbf{b}_n(\mathbf{s}) - \mathbf{c}_s^{\mathbf{T}}\hat{\mathbf{y}}} .$$

Note that $\mathbf{c}_s^{\mathbf{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$ and $\mathbf{c}_s^{\mathbf{T}}[\mathbf{b}_n(\mathbf{s}) - \mathbf{x}] > P$ imply $\lambda_0 \in (0, 1)$. Then define:

$$\tilde{\mathbf{y}} := \lambda_0 \hat{\mathbf{y}} + (1 - \lambda_0)\mathbf{b}_n(\mathbf{s}) .$$

By the convexity of $G_n(\cdot, \mathbf{s})$ along the line segment from $\hat{\mathbf{y}}$ to $\mathbf{b}_n(\mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n(\tilde{\mathbf{y}}, \mathbf{s}) \geq G_n(\mathbf{b}_n(\mathbf{s}), \mathbf{s}) .$$

Since $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ was arbitrary, we conclude:

$$\min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})} \left\{ G_n(\mathbf{y}, \mathbf{s}) \right\} = \min_{\mathbf{y} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})} \left\{ G_n(\mathbf{y}, \mathbf{s}) \right\} .$$

Next, let $\mathbf{x} \in \mathcal{R}_{IV-C}(n, \mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ be arbitrary such that $\mathbf{c}_s^{\mathbf{T}} [\hat{\mathbf{y}} - \mathbf{x}] < P$. We consider two exhaustive cases, and for each case, we construct a $\tilde{\mathbf{y}} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ such that $G_n(\tilde{\mathbf{y}}, \mathbf{s}) \leq G_n(\hat{\mathbf{y}}, \mathbf{s})$.

Case 1: $\hat{y}^2 < f_n^2(\hat{y}^1, \mathbf{s})$ and $\bar{\mathbf{y}} := (\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})) \notin \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$

Let $\tilde{\mathbf{y}} := (\hat{y}^1, x^2 + \frac{P - c_{s1} \cdot [\hat{y}^1 - x^1]}{c_{s2}})$. Then, by the convexity of $G_n(\cdot, \mathbf{s})$ along $y^1 = \hat{y}^1$, the definition of $f_n^2(\hat{y}^1, \mathbf{s})$, and $\hat{y}^2 \leq \tilde{y}^2 \leq f_n^2(\hat{y}^1, \mathbf{s})$, we have:

$$G_n(\bar{\mathbf{y}}, \mathbf{s}) = G_n\left((\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})), \mathbf{s}\right) \leq G_n(\tilde{\mathbf{y}}, \mathbf{s}) \leq G_n(\hat{\mathbf{y}}, \mathbf{s}) .$$

It is also straightforward to check that $\tilde{\mathbf{y}} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$, as desired.

Case 2: All other $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ such that $\mathbf{c}_s^{\mathbf{T}} [\hat{\mathbf{y}} - \mathbf{x}] < P$

By the definition of $f_n^2(\hat{y}^1, \mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n\left((\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})), \mathbf{s}\right) . \quad (142)$$

Define:

$$\tilde{y}^1 := \sup \left\{ y^1 \in [x^1, \hat{y}^1] : \mathbf{c}_s^{\mathbf{T}} (y^1, f_n^2(y^1, \mathbf{s})) \geq \mathbf{c}_s^{\mathbf{T}} \mathbf{x} + P \right\}, \text{ and}$$

$$\tilde{y}^2 := \frac{P - c_{s1} \cdot [\tilde{y}^1 - x^1]}{c_{s2}} .$$

By the convexity of $G_n(\cdot, \mathbf{s})$ along $(y^1, f_n^2(y^1, \mathbf{s}))$, we have:

$$G_n\left((\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})), \mathbf{s}\right) \geq G_n\left((\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})), \mathbf{s}\right) . \quad (143)$$

Furthermore, we have:

$$G_n\left((\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})), \mathbf{s}\right) = G_n\left((\tilde{y}^1, \tilde{y}^2), \mathbf{s}\right) = G_n(\tilde{\mathbf{y}}, \mathbf{s}) . \quad (144)$$

If $\tilde{y}^2 = f_n^2(\tilde{y}^1, \mathbf{s})$, (144) is trivial. Otherwise, there is a discontinuity in $f_n^2(\cdot, \mathbf{s})$ at \tilde{y}^1 , and we have:

$$\lim_{y^1 \nearrow \tilde{y}^1} f_n^2(y^1, \mathbf{s}) \geq \tilde{y}^2 \geq \lim_{y^1 \searrow \tilde{y}^1} f_n^2(y^1, \mathbf{s}) , \quad (145)$$

with at least one of the inequalities being strict. Nonetheless, $G_n\left((y^1, f_n^2(y^1, \mathbf{s})), \mathbf{s}\right)$ is a continuous function of y^1 , and therefore:

$$G_n\left((\tilde{y}^1, \lim_{y^1 \nearrow \tilde{y}^1} f_n^2(y^1, \mathbf{s})), \mathbf{s}\right) = G_n\left((\tilde{y}^1, \lim_{y^1 \searrow \tilde{y}^1} f_n^2(y^1, \mathbf{s})), \mathbf{s}\right) = G_n\left((\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})), \mathbf{s}\right). \quad (146)$$

The convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^1 = \tilde{y}^1$ and (146) imply:

$$G_n\left((\tilde{y}^1, y^2), \mathbf{s}\right) = G_n\left((\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})), \mathbf{s}\right) , \quad \forall y^2 \in \left[\lim_{y^1 \searrow \tilde{y}^1} f_n^2(y^1, \mathbf{s}) , \lim_{y^1 \nearrow \tilde{y}^1} f_n^2(y^1, \mathbf{s}) \right] ,$$

which in combination with (145) implies (144). Combining (142)-(144) yields the desired result: $G_n(\tilde{\mathbf{y}}, \mathbf{s}) \leq G_n(\hat{\mathbf{y}}, \mathbf{s})$ for a $\tilde{\mathbf{y}} \in \mathcal{H}^d(\mathbf{x}, \mathbf{s})$.

The validity of (25) for $\mathbf{x} \in \mathcal{R}_{IV-A}(n, \mathbf{s})$ follows from a symmetric argument, completing the proof of (25) and Theorem 8. \square

IX. APPENDIX B - INFINITE HORIZON DISCOUNTED EXPECTED COST PROOFS

A. Proof of Theorem 5

Our line of analysis is similar in spirit to [53], [79], and [80, Chapter 8]. Let $x \in \mathbb{R}_+$ and $s \in \mathcal{S}$ be arbitrary. First, we show inductively that $V_1(x, s) \leq V_2(x, s) \leq \dots \leq V_n(x, s) \leq V_{n+1}(x, s) \leq \dots$

Base Case: $n = 1$

$$\begin{aligned} V_1(x, s) &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \{c(z, s) + h(x + z - d)\} \\ &\leq \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_1(x + z - d, S_1) \mid S_2 = s] \end{array} \right\} \\ &= V_2(x, s) , \end{aligned}$$

where the inequality follows from $V_1(x, s) \geq 0, \forall x, \forall s$.

Induction Step: Assume $V_n(x, s) \leq V_{n+1}(x, s)$ for $n = 1, 2, \dots, m-1$. We show it is true for $n = m$:

$$\begin{aligned} V_m(x, s) &= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_{m-1}(x + z - d, S_{m-1}) \mid S_m = s] \end{array} \right\} \\ &\leq \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_m(x + z - d, S_m) \mid S_{m+1} = s] \end{array} \right\} \\ &= V_{m+1}(x, s), \end{aligned}$$

where the inequality follows from the induction hypothesis and the homogeneity of the Markov process representing the channel condition. So, for every $x \in \mathbb{R}_+$ and $s \in \mathcal{S}$, $\{V_n(x, s)\}_{n=1,2,\dots}$ is a nondecreasing sequence.

Next, consider a policy π^d transmitting d packets in every slot, regardless of channel condition.

Define:

$$\tilde{c}_{\max} := \sup_{\substack{s \in \mathcal{S} \\ k \in \{0,1,\dots,K\}}} \{\tilde{c}_k(s)\} < \infty. \quad (147)$$

Then we have:

$$V_n(x, s) \leq V_n^{\pi^d}(x, s) \leq \left(\tilde{c}_{\max} \cdot d + h(x) \right) \frac{1 - \alpha^n}{1 - \alpha} \leq \left(\tilde{c}_{\max} \cdot d + h(x) \right) \frac{1}{1 - \alpha} < \infty,$$

so $\{V_n(x, s)\}_{n=1,2,\dots}$ is a bounded nondecreasing sequence, implying $\lim_{n \rightarrow \infty} V_n(x, s)$ exists and is finite, $\forall x \in \mathbb{R}_+, \forall s \in \mathcal{S}$.

We now move on to part (b). Recall from Section VIII-A that $V_n(x, s)$ is convex in x , for all n and all s . Define $V_\infty(x, s) := \lim_{n \rightarrow \infty} V_n(x, s)$. Let $s \in \mathcal{S}$ be arbitrary, but fixed. $V_\infty(x, s) = \sup_{n \in \mathbb{N}} V_n(x, s)$, so $V_\infty(x, s)$ is convex in x as it is the pointwise supremum of the convex functions $\{V_n(x, s)\}_{n=1,2,\dots}$.

Define $\tilde{g}_\infty : [d, \infty) \times \mathcal{S} \rightarrow \mathbb{R}_+$ by

$$\begin{aligned}
\tilde{g}_\infty(y, s) &:= h(y - d) + \alpha \cdot \mathbb{E} [V_\infty(y - d, S') \mid S = s] \\
&= h(y - d) + \alpha \cdot \mathbb{E} \left[\lim_{n \rightarrow \infty} V_n(y - d, S') \mid S = s \right] \\
&= h(y - d) + \alpha \cdot \lim_{n \rightarrow \infty} \mathbb{E} [V_n(y - d, S') \mid S = s] \\
&= \lim_{n \rightarrow \infty} \tilde{g}_n(y, s) ,
\end{aligned} \tag{148}$$

where (148) follows from the homogeneity of the Markov process representing the channel condition and the Monotone Convergence Theorem. Furthermore, for each $s \in \mathcal{S}$, $\tilde{g}_\infty(y, s)$ is convex in y and $\lim_{y \rightarrow \infty} \tilde{g}_\infty(y, s) \geq \lim_{y \rightarrow \infty} h(y - d) = \infty$. Thus, for every s , at least one finite number achieves the global minimum of $\tilde{g}_\infty(y, s)$.

Next, we proceed to part (d), and let $s \in \mathcal{S}$ be arbitrary. Define $b_{\infty, -1}(s) := \infty$ and

$$b_{\infty, k}(s) := \max \left\{ d, \inf \{ b \mid \tilde{g}'_+(b, s) \geq -\tilde{c}_k(s) \} \right\} , \quad \forall k \in \{0, 1, \dots, K\} .$$

Clearly, $b_{\infty, -1}(s) = \lim_{n \rightarrow \infty} b_{n, -1}(s)$, as $b_{n, -1}(s) := \infty$ for every n . Let $k \in \{0, 1, \dots, K\}$ be arbitrary. We want to show:

$$\begin{aligned}
\lim_{n \rightarrow \infty} b_{n, k}(s) &= \lim_{n \rightarrow \infty} \max \left\{ d, \inf \{ b \mid \tilde{g}'_n^+(b, s) \geq -\tilde{c}_k(s) \} \right\} \\
&= \max \left\{ d, \inf \{ b \mid \tilde{g}'_\infty^+(b, s) \geq -\tilde{c}_k(s) \} \right\} := b_{\infty, k}(s) .
\end{aligned}$$

By the continuity of $\max\{d, \cdot\}$, it suffices to show:

$$\lim_{n \rightarrow \infty} \left\{ \inf \{ b \mid \tilde{g}'_n^+(b, s) \geq -\tilde{c}_k(s) \} \right\} = \inf \{ b \mid \tilde{g}'_\infty^+(b, s) \geq -\tilde{c}_k(s) \} . \tag{149}$$

Before proceeding to show (149), we present a lemma due to Sobel [81, Lemma 3, pg. 732], which is also presented in [80, Lemma 8-5, pg. 425].

Lemma 6 (Sobel, 1971). *Let g, g_1, g_2, \dots be convex functions on an open convex subset X of \mathbb{R} such that $g_n(x) \rightarrow g(x)$ as $n \rightarrow \infty$ and $g_n(x) \leq g_{n+1}(x)$ for all n and x . Let $g'_n{}^-(x)$ and $g'^-(x)$ denote derivatives from the left and $g'_n{}^+(x)$ and $g'^+(x)$ denote derivatives from the right.*

Then for all $x \in X$:

$$g'^-(x) \leq \liminf_{n \rightarrow \infty} g_n'^-(x) \leq \limsup_{n \rightarrow \infty} g_n'^+(x) \leq g'^+(x) . \quad (150)$$

We now prove (149) by contradiction. Define:

$$\begin{aligned} \hat{b}_{n,k}(s) &:= \inf\{b \mid \tilde{g}_n'^+(b, s) \geq -\tilde{c}_k(s)\}, \text{ and} \\ \hat{b}_{\infty,k}(s) &:= \inf\{b \mid \tilde{g}_{\infty}'^+(b, s) \geq -\tilde{c}_k(s)\} . \end{aligned}$$

First, assume $\liminf_{n \rightarrow \infty} \hat{b}_{n,k}(s) < \hat{b}_{\infty,k}(s)$, so there exists an $x_0 \in \mathbb{R}_+$ such that $d < x_0 < \hat{b}_{\infty,k}(s)$, and a sequence $\{n_i\}_{i=1,2,\dots}$ such that $\lim_{i \rightarrow \infty} \hat{b}_{n_i,k}(s) = x_0$. Then we have:

$$-\tilde{c}_k(s) \leq \lim_{i \rightarrow \infty} \tilde{g}_{n_i}'^+(x_0, s) \quad (151)$$

$$\begin{aligned} &\leq \limsup_{n \rightarrow \infty} \tilde{g}_n'^+(x_0, s) \\ &\leq \tilde{g}_{\infty}'^+(x_0, s) . \end{aligned} \quad (152)$$

Here, (151) follows from $\lim_{i \rightarrow \infty} \hat{b}_{n_i,k}(s) = x_0$, and the fact that $\tilde{g}_n'^+(\cdot, s)$ is continuous from the right. Equation (152) follows from Lemma 6. Yet, $\tilde{g}_{\infty}'^+(x_0, s) \geq -\tilde{c}_k(s)$ implies $\hat{b}_{\infty,k}(s) \leq x_0$, which is a contradiction. We conclude:

$$\liminf_{n \rightarrow \infty} \hat{b}_{n,k}(s) \geq \hat{b}_{\infty,k}(s) . \quad (153)$$

Next, assume $\limsup_{n \rightarrow \infty} \hat{b}_{n,k}(s) > \hat{b}_{\infty,k}(s) \geq d$, and define:

$$x_1 := \frac{\limsup_{n \rightarrow \infty} \hat{b}_{n,k}(s) + \hat{b}_{\infty,k}(s)}{2} .$$

Then we have:

$$-\tilde{c}_k(s) \leq \tilde{g}'^+(\hat{b}_{\infty,k}(s), s) \quad (154)$$

$$\leq \tilde{g}'^-(x_1, s) \quad (155)$$

$$\leq \liminf_{n \rightarrow \infty} \tilde{g}'^-(x_1, s) \quad (156)$$

$$\leq \liminf_{n \rightarrow \infty} \tilde{g}'^+(x_1, s) \quad (157)$$

Here, (154) follows from the fact that $\tilde{g}'^+(\cdot, s)$ is continuous from the right; (156) follows from Lemma 6¹³; and (155) and (157) follow from the fact (see, e.g., [77, pg. 228]) that for a proper convex function f on \mathbb{R} , $z_1 < x < z_2$ implies:

$$f'^+(z_1) \leq f'^-(x) \leq f'^+(x) \leq f'^-(z_2) .$$

$\liminf_{n \rightarrow \infty} \tilde{g}'^+(x_1, s) \geq -\tilde{c}_k(s)$ implies that for every sequence $\{n_j\}_{j=1,2,\dots}$, we have:

$$\lim_{j \rightarrow \infty} \tilde{g}'^+(x_1, s) \geq -\tilde{c}_k(s) ,$$

and, in turn:

$$\lim_{j \rightarrow \infty} \hat{b}_{n_j,k}(s) \leq x_1 .$$

Therefore, $\limsup_{n \rightarrow \infty} \hat{b}_{n,k}(s) \leq x_1$, which is a contradiction. We conclude:

$$\limsup_{n \rightarrow \infty} \hat{b}_{n,k}(s) \leq \hat{b}_{\infty,k}(s) . \quad (158)$$

Equations (153) and (158) imply (149).

¹³One hypothesis of Lemma 6 is that all functions are defined on an *open* convex subset of \mathbb{R} . While our functions $\tilde{g}'_\infty(\cdot, s)$ and $\{\tilde{g}'_n(\cdot, s)\}_{n \in \mathbb{N}}$ are defined on $[d, \infty)$, we only apply Lemma 6 at the points $x_0, x_1 \in (d, \infty)$. Thus, equations (152) and (156) follow from the application of Lemma 6 to the restrictions of the functions $\tilde{g}'_\infty(\cdot, s)$ and $\{\tilde{g}'_n(\cdot, s)\}_{n \in \mathbb{N}}$ to the domain of (d, ∞) .

We are now ready to prove parts (e) and (f) of Theorem 5. Define

$$z_{\infty}^*(x, s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } b_{\infty,k}(s) - \tilde{z}_{k-1}(s) < x \leq b_{\infty,k-1}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K\} \\ b_{\infty,k}(s) - x, & \text{if } b_{\infty,k}(s) - \tilde{z}_k(s) < x \leq b_{\infty,k}(s) - \tilde{z}_{k-1}(s), \\ & k \in \{0, 1, \dots, K-1\} \\ b_{\infty,K}(s) - x, & \text{if } b_{\infty,K}(s) - \tilde{z}_{\max}(s) < x \leq b_{\infty,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } 0 \leq x \leq b_{\infty,K}(s) - \tilde{z}_{\max}(s) \end{cases}$$

Clearly, $\lim_{n \rightarrow \infty} b_{n,k}(s) = b_{\infty,k}(s)$ implies $\lim_{n \rightarrow \infty} z_n^*(x, s) = z_{\infty}^*(x, s)$, $\forall x \in \mathbb{R}^+$, $\forall s \in \mathcal{S}$. Furthermore, $\tilde{g}_n(y, s) \rightarrow \tilde{g}_{\infty}(y, s)$ and $z_n^*(x, s) \rightarrow z_{\infty}^*(x, s)$ as $n \rightarrow \infty$ imply:

$$\lim_{n \rightarrow \infty} \tilde{g}_n(x + z_n^*(x, s)) = \tilde{g}_{\infty}(x + z_{\infty}^*(x, s)), \quad \forall x \in \mathbb{R}^+, \quad \forall s \in \mathcal{S}. \quad (159)$$

So for all $x \in \mathbb{R}^+$ and $s \in \mathcal{S}$, we have:

$$\begin{aligned} V_{\infty}(x, s) &= \lim_{n \rightarrow \infty} V_n(x, s) \\ &= \lim_{n \rightarrow \infty} \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ c(z, s) + \tilde{g}_n(x + z, s) \right\} \\ &= \lim_{n \rightarrow \infty} \left\{ c(z_n^*(x, s), s) + \tilde{g}_n(x + z_n^*(x, s), s) \right\} \end{aligned} \quad (160)$$

$$= c(z_{\infty}^*(x, s), s) + \tilde{g}_{\infty}(x + z_{\infty}^*(x, s), s) \quad (161)$$

$$= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ c(z, s) + \tilde{g}_{\infty}(x + z, s) \right\} \quad (162)$$

$$= \min_{\{\max(0, d-x) \leq z \leq \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z, s) + h(x + z - d) \\ + \alpha \cdot \mathbb{E}[V_{\infty}(x + z - d, S') \mid S = s] \end{array} \right\}.$$

Equation (160) follows from Theorem 1, and (161) follows from (159) and the continuity of $c(\cdot, s)$. Equation (162) follows from the same line of analysis as part (ii) of the induction step in the proof of Theorem 3, with $\tilde{g}_{\infty}(\cdot, s)$, $b_{\infty,k}(s)$, and $z_{\infty}^*(\cdot, s)$ replacing $\tilde{g}_m(\cdot, s)$, $b_{m,k}(s)$, and $z_m^*(\cdot, s)$, respectively. Thus, $V_{\infty}(\cdot, \cdot)$, the limit of the finite horizon value functions, satisfies the α -DCOE (19) and is also equal to the infinite horizon discounted expected cost-to-go resulting

from the stationary policy $\pi_\infty^* := (z_\infty^*, z_\infty^*, \dots)$. We conclude π_∞^* , the natural extension of the finite horizon optimal policy, is optimal for the infinite horizon problem (see, for example, [44, Propositions 9.12 and 9.16]). \square

B. Proof of Theorem 9

We follow the same line of analysis as the proof of Theorem 5. Let $\mathbf{x} \in \mathbb{R}_+^2$ and $\mathbf{s} \in \mathcal{S}$ be arbitrary. First, we show inductively that $V_1(\mathbf{x}, \mathbf{s}) \leq V_2(\mathbf{x}, \mathbf{s}) \leq \dots \leq V_n(\mathbf{x}, \mathbf{s}) \leq V_{n+1}(\mathbf{x}, \mathbf{s}) \leq \dots$

Base Case: $n = 1$

$$\begin{aligned} V_1(\mathbf{x}, \mathbf{s}) &= \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \{ \mathbf{c}_s^T \mathbf{z} + h(\mathbf{x} + \mathbf{z} - \mathbf{d}) \} \\ &\leq \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T \mathbf{z} + h(\mathbf{x} + \mathbf{z} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_1(\mathbf{x} + \mathbf{z} - \mathbf{d}, \mathbf{S}_1) \mid \mathbf{S}_2 = \mathbf{s}] \end{array} \right\} \\ &= V_2(\mathbf{x}, \mathbf{s}), \end{aligned}$$

where the inequality follows from $V_1(\mathbf{x}, \mathbf{s}) \geq 0$, $\forall \mathbf{x}, \forall \mathbf{s}$.

Induction Step: Assume $V_n(\mathbf{x}, \mathbf{s}) \leq V_{n+1}(\mathbf{x}, \mathbf{s})$ for $n = 1, 2, \dots, m-1$. We show it is true for $n = m$:

$$\begin{aligned} V_m(\mathbf{x}, \mathbf{s}) &= \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T \mathbf{z} + h(\mathbf{x} + \mathbf{z} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_{m-1}(\mathbf{x} + \mathbf{z} - \mathbf{d}, \mathbf{S}_{m-1}) \mid \mathbf{S}_m = \mathbf{s}] \end{array} \right\} \\ &\leq \min_{\mathbf{z} \in \mathcal{A}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T \mathbf{z} + h(\mathbf{x} + \mathbf{z} - \mathbf{d}) \\ + \alpha \cdot \mathbb{E}[V_m(\mathbf{x} + \mathbf{z} - \mathbf{d}, \mathbf{S}_m) \mid \mathbf{S}_{m+1} = \mathbf{s}] \end{array} \right\} \\ &= V_{m+1}(\mathbf{x}, \mathbf{s}), \end{aligned}$$

where the inequality follows from the induction hypothesis and the homogeneity of the Markov process representing the channel condition. So, for every $\mathbf{x} \in \mathbb{R}_+^2$ and $\mathbf{s} \in \mathcal{S}$, $\{V_n(\mathbf{x}, \mathbf{s})\}_{n=1,2,\dots}$ is a nondecreasing sequence.

Next, consider a policy π^d transmitting d^1 packets to user 1 and d^2 packets to user 2 in every

slot, regardless of channel condition. Define:

$$\mathbf{c}_{\max}^{\text{T}} := (c_{\max}^1, c_{\max}^2)^{\text{T}}, \text{ where } c_{\max}^i := \sup_{s^i \in \mathcal{S}^i} \{c_{s^i}\} < \infty. \quad (163)$$

Then we have:

$$V_n(\mathbf{x}, \mathbf{s}) \leq V_n^{\pi^{\text{d}}}(\mathbf{x}, \mathbf{s}) \leq \left(\mathbf{c}_{\max}^{\text{T}} \mathbf{d} + h(\mathbf{x}) \right) \frac{1 - \alpha^n}{1 - \alpha} \leq \left(\mathbf{c}_{\max}^{\text{T}} \mathbf{d} + h(\mathbf{x}) \right) \frac{1}{1 - \alpha} < \infty,$$

so $\{V_n(\mathbf{x}, \mathbf{s})\}_{n=1,2,\dots}$ is a bounded nondecreasing sequence, implying $\lim_{n \rightarrow \infty} V_n(\mathbf{x}, \mathbf{s})$ exists and is finite, $\forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S}$.

Next, recall from Theorem 7 that $V_n(\mathbf{x}, \mathbf{s})$ is convex and supermodular in \mathbf{x} , for all n and all \mathbf{s} . Define $V_{\infty}(\mathbf{x}, \mathbf{s}) := \lim_{n \rightarrow \infty} V_n(\mathbf{x}, \mathbf{s})$. Let $\mathbf{s} \in \mathcal{S}$ be arbitrary, but fixed. $V_{\infty}(\mathbf{x}, \mathbf{s}) = \sup_{n \in \mathbb{N}} V_n(\mathbf{x}, \mathbf{s})$, so $V_{\infty}(\mathbf{x}, \mathbf{s})$ is convex in \mathbf{x} as it is the pointwise supremum of the convex functions $\{V_n(\mathbf{x}, \mathbf{s})\}_{n=1,2,\dots}$. Furthermore, the pointwise limit of supermodular functions is supermodular (see, e.g., [76, Lemma 2.6.1]), so $V_{\infty}(\mathbf{x}, \mathbf{s})$ is also supermodular in \mathbf{x} .

Define $G_{\infty} : [d^1, \infty) \times [d^2, \infty) \times \mathcal{S} \rightarrow \mathbb{R}_+$ by

$$\begin{aligned} G_{\infty}(\mathbf{y}, \mathbf{s}) &:= \mathbf{c}_{\mathbf{s}}^{\text{T}} \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}[V_{\infty}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \\ &= \mathbf{c}_{\mathbf{s}}^{\text{T}} \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}\left[\lim_{n \rightarrow \infty} V_n(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}\right] \\ &= \mathbf{c}_{\mathbf{s}}^{\text{T}} \mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \lim_{n \rightarrow \infty} \mathbb{E}[V_n(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \end{aligned} \quad (164)$$

$$= \lim_{n \rightarrow \infty} G_n(\mathbf{y}, \mathbf{s}), \quad (165)$$

where (164) follows from the homogeneity of the Markov process representing the channel condition and the Monotone Convergence Theorem. Furthermore, for each $\mathbf{s} \in \mathcal{S}$, $G_{\infty}(\mathbf{y}, \mathbf{s})$ is convex and supermodular in \mathbf{y} as it is the sum of an affine function of \mathbf{y} , a convex separable function of $\mathbf{y} - \mathbf{d}$ and a weighted sum of the convex supermodular functions $V_{\infty}(\mathbf{y} - \mathbf{d}, \mathbf{s}')$. Additionally, $\lim_{\|\mathbf{y}\| \rightarrow \infty} G_{\infty}(\mathbf{y}, \mathbf{s}) \geq \lim_{\|\mathbf{y}\| \rightarrow \infty} \mathbf{c}_{\mathbf{s}}^{\text{T}} \mathbf{y} = \infty$. Thus, for every \mathbf{s} , at least one finite vector achieves the global minimum of $G_{\infty}(\mathbf{y}, \mathbf{s})$; $\mathcal{B}_{\infty}(\mathbf{s})$ is a nonempty closed convex set; and $\mathbf{b}_{\infty}(\mathbf{s})$, $f_{\infty}^1(\cdot, \mathbf{s})$, and $f_{\infty}^2(\cdot, \mathbf{s})$ are well-defined. The structure of the optimal policy outlined in (b) then follows from the same line of analysis used to prove the the structure of the optimal policy in

the induction step of Theorem 8.

Moreover, since for a fixed $\mathbf{s} \in \mathcal{S}$ and $x^2 \in [d^2, \infty)$,

$$f_n^1(x^2, \mathbf{s}) := \inf \left\{ \operatorname{argmin}_{y^1 \in [d^1, \infty)} \left\{ G_n(y^1, x^2, s^1, s^2) \right\} \right\} = \inf \{ b^1 \mid G_n'^+(b^1, x^2, s^1, s^2) \geq 0 \} ,$$

the convergence of $f_n^1(x^2, \mathbf{s})$ to $f_\infty^1(x^2, \mathbf{s})$ follows from the same argument used to show (149).

The convergence of $f_n^2(x^1, \mathbf{s})$ to $f_\infty^2(x^1, \mathbf{s})$ follows from a symmetric argument.

For all $\mathbf{s} \in \mathcal{S}$ and $x^1 \in [d^1, \infty)$, define:

$$\Psi_n(x^1, \mathbf{s}) := \min_{x^2 \in [d^2, \infty)} \{ G_n(x^1, x^2, s^1, s^2) \} = G_n(x^1, f_n^2(x^1, \mathbf{s}), s^1, s^2) , \quad \forall n \in \mathbb{N},$$

and

$$\Psi_\infty(x^1, \mathbf{s}) := \min_{x^2 \in [d^2, \infty)} \{ G_\infty(x^1, x^2, s^1, s^2) \} = G_\infty(x^1, f_\infty^2(x^1, \mathbf{s}), s^1, s^2) .$$

For fixed but arbitrary x^1 and \mathbf{s} , $f_n^2(x^1, \mathbf{s})$ converges to $f_\infty^2(x^1, \mathbf{s})$, and, by Dini's Theorem, $G_n(x^1, \cdot, \mathbf{s})$ converges to $G_\infty(x^1, \cdot, \mathbf{s})$ uniformly on a compact interval containing $f_\infty^2(x^1, \mathbf{s})$.

Thus, $\Psi_n(x^1, \mathbf{s})$ converges pointwise to $\Psi_\infty(x^1, \mathbf{s})$. Moreover, for every \mathbf{s} , $\{\Psi_n(x^1, \mathbf{s})\}_{n \in \mathbb{N}}$ and $\Psi_\infty(x^1, \mathbf{s})$ are all convex in x^1 with the limit as x^1 approaches infinity equal to infinity. Therefore, by the same argument used to show (149), $b_n^1(\mathbf{s})$ converges pointwise to $b_\infty^1(\mathbf{s})$.

For all $\mathbf{s} \in \mathcal{S}$ and $x^2 \in [d^2, \infty)$, define:

$$\tilde{\Psi}_n(x^2, \mathbf{s}) := G_n(b_n^1(\mathbf{s}), x^2, s^1, s^2) , \quad \forall n \in \mathbb{N},$$

and

$$\tilde{\Psi}_\infty(x^2, \mathbf{s}) := G_\infty(b_\infty^1(\mathbf{s}), x^2, s^1, s^2) .$$

For fixed but arbitrary x^2 and \mathbf{s} , $b_n^1(\mathbf{s})$ converges to $b_\infty^1(\mathbf{s})$, and, by Dini's Theorem, $G_n(\cdot, x^2, \mathbf{s})$ converges to $G_\infty(\cdot, x^2, \mathbf{s})$ uniformly on a compact interval around $b_\infty^1(\mathbf{s})$. Thus, $\tilde{\Psi}_n(x^2, \mathbf{s})$ converges pointwise to $\tilde{\Psi}_\infty(x^2, \mathbf{s})$. Moreover, for every \mathbf{s} , $\{\tilde{\Psi}_n(x^2, \mathbf{s})\}_{n \in \mathbb{N}}$ and $\tilde{\Psi}_\infty(x^2, \mathbf{s})$ are all convex in x^2 with the limit as x^2 approaches infinity equal to infinity. Therefore, by the same argument used to show (149), $b_n^2(\mathbf{s})$ converges pointwise to $b_\infty^2(\mathbf{s})$, and we conclude $\mathbf{b}_\infty(\mathbf{s}) =$

$\lim_{n \rightarrow \infty} \mathbf{b}_n(\mathbf{s})$. \square

X. APPENDIX C - INFINITE HORIZON AVERAGE EXPECTED COST PROOFS

In this section, we prove Theorem 10 using the vanishing discount approach (see, e.g., [43]). The proof of Theorem 6 is nearly identical, and we note the few key differences.

A. Proof of Theorem 10

Substituting (27) and (29) into the α -DCOE (26) and rearranging yields:

$$\begin{aligned} & (1 - \alpha) \cdot m_{\infty, \alpha} + w_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) \\ &= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}[w_{\infty, \alpha}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \right\}, \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S}. \end{aligned} \quad (166)$$

The main idea of the vanishing discount approach is to take the limit as α goes to 1, and show that (166) converges to the ACOE (30).

We start by presenting five conditions from the literature on the vanishing discount approach.

Condition (G). $\rho := \inf_{\pi \in \Pi} \inf_{\mathbf{x} \in \mathbb{R}_+^2} \inf_{\mathbf{s} \in \mathcal{S}} \left\{ \limsup_{N \rightarrow \infty} \frac{1}{N} V_{N,1}^\pi(\mathbf{x}, \mathbf{s}) \right\} < \infty$.

Condition (W). (i) *The state space $\mathbb{R}_+^2 \times \mathcal{S}$ is a locally compact space with countable base.*

(ii) *The action space $\tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$ is a nonempty compact subset of the state space $\mathbb{R}_+^2 \times \mathcal{S}$, and the multifunction $\phi : (\mathbf{x}, \mathbf{s}) \mapsto \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$ is upper semicontinuous; that is, $\phi^{-1}(F)$ is closed in $\mathbb{R}_+^2 \times \mathcal{S}$ for every closed set $F \subset \mathbb{R}_+^2$.*

(iii) *The transition law is weakly continuous (see, e.g., [43, Appendix C]).*

(iv) *The one-stage cost $c(\mathbf{z}, \mathbf{s}) + h(\mathbf{x} + \mathbf{z} - \mathbf{d})$ is lower semicontinuous and nonnegative.*

Condition (B). $\sup_{\alpha < 1} w_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) < \infty$ for all $\mathbf{x} \in \mathbb{R}_+^2$ and $\mathbf{s} \in \mathcal{S}$.

Condition (B2). *There is a measurable function $\bar{\kappa} : \mathbb{R}_+^2 \times \mathcal{S} \rightarrow \mathbb{R}_+$ such that $\bar{\kappa} \geq w_{\infty, \alpha}$ for all $\alpha \in [0, 1)$, and:*

$$\mathbb{E}[\bar{\kappa}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] < \infty, \quad \forall (\mathbf{x}, \mathbf{s}) \in \mathbb{R}_+^2 \times \mathcal{S}, \quad \forall \mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s}). \quad (167)$$

Condition (E). For every increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, the sequence $\{w_{\infty,\alpha(l)}\}_{l=1,2,\dots}$ is equicontinuous.

We show below that our model satisfies these five conditions, but first we show how they lead to Theorem 10. Parts (b), (c), and (e) of Theorem 10 follow directly from the following theorem due to Schäl [82, Theorem 3.8] and adapted to our notation.

Theorem 11 (Schäl, 1993). *Suppose conditions (G), (W), and (B) hold. Then the minimum average cost $\rho^* = \inf_{\pi \in \Pi} \inf_{\mathbf{x} \in \mathbb{R}_+^2} \left\{ \limsup_{N \rightarrow \infty} \frac{1}{N} V_{N,1}^{\pi}(\mathbf{x}, \mathbf{s}) \right\} = \lim_{\alpha \nearrow 1} (1 - \alpha) \cdot m_{\infty,\alpha}$. Moreover, there exists an optimal selector $\mathbf{y}_{\infty,1}^*(\cdot, \cdot)$ such that:*

$$\begin{aligned} \rho^* + w_{\infty,1}(\mathbf{x}, \mathbf{s}) &\geq \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T [\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) \\ + \mathbb{E}[w_{\infty,1}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \end{array} \right\} \\ &= \mathbf{c}_s^T [\mathbf{y}_{\infty,1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{x}] + \mathbf{h}(\mathbf{y}_{\infty,1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{d}) \\ &\quad + \mathbb{E}[w_{\infty,1}(\mathbf{y}_{\infty,1}^*(\mathbf{x}, \mathbf{s}) - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}], \quad \forall \mathbf{x} \in \mathbb{R}_+^2, \quad \forall \mathbf{s} \in \mathcal{S}, \end{aligned} \quad (168)$$

where for every $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}_+^2 \times \mathcal{S}$ and any increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1,

$$w_{\infty,1}(\mathbf{x}, \mathbf{s}) := \liminf_{l \rightarrow \infty} w_{\infty,\alpha(l)}(\mathbf{x}, \mathbf{s}) . \quad (169)$$

Furthermore, for every $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}_+^2 \times \mathcal{S}$ and any increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\dots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ approaching 1 and a sequence $\{\mathbf{x}(i)\}_{i=1,2,\dots}$ approaching \mathbf{x} such that:

$$\mathbf{y}_{\infty,1}^*(\mathbf{x}, \mathbf{s}) = \lim_{i \rightarrow \infty} \mathbf{y}_{\infty,\alpha(l_i)}^*(\mathbf{x}(i), \mathbf{s}) .$$

To get the opposite inequality from (168), we use a method from [83] and [84, Theorem 4.1] (which is presented in [43, Section 5.5]). Namely, for every $\mathbf{x} \in \mathbb{R}_+^2$, $\mathbf{s} \in \mathcal{S}$, $\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})$, and

$\alpha(l)$ from (169), (166) implies:

$$\begin{aligned} & (1 - \alpha(l)) \cdot m_{\infty, \alpha(l)} + w_{\infty, \alpha(l)}(\mathbf{x}, \mathbf{s}) \\ & \leq \mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) + \alpha(l) \cdot \mathbb{E}[w_{\infty, \alpha(l)}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] . \end{aligned} \quad (170)$$

Furthermore, in combination with Conditions (B) and (E), the Arzelá-Ascoli Theorem implies there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\dots}$ of $\{\alpha(l)\}_{l=1,2,\dots}$ such that:

$$w_{\infty, 1}(\mathbf{x}, \mathbf{s}) = \lim_{i \rightarrow \infty} w_{\infty, \alpha(l_i)}(\mathbf{x}, \mathbf{s}) , \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S} . \quad (171)$$

Then, taking the limit of (170) as α goes to 1 along the sequence $\{\alpha(l_i)\}_{i=1,2,\dots}$, (28), (171), Condition (B2), and the Lebesgue Dominated Convergence Theorem imply:

$$\begin{aligned} \rho^* + w_{\infty, 1}(\mathbf{x}, \mathbf{s}) & \leq \mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) + \mathbb{E}[w_{\infty, 1}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] , \\ & \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S}, \forall \mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s}) , \end{aligned}$$

which implies:

$$\rho^* + w_{\infty, 1}(\mathbf{x}, \mathbf{s}) \leq \min_{\mathbf{y} \in \tilde{\mathcal{A}}^d(\mathbf{x}, \mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_s^T[\mathbf{y} - \mathbf{x}] + \mathbf{h}(\mathbf{y} - \mathbf{d}) \\ + \mathbb{E}[w_{\infty, 1}(\mathbf{y} - \mathbf{d}, \mathbf{S}') \mid \mathbf{S} = \mathbf{s}] \end{array} \right\} , \forall \mathbf{x} \in \mathbb{R}_+^2, \forall \mathbf{s} \in \mathcal{S} . \quad (172)$$

Equations (168) and (172) yield the ACOE (30). Moreover, from (171) and the fact that convexity and supermodularity are preserved under pointwise limits, we conclude that for every $\mathbf{s} \in \mathcal{S}$, $w_{\infty, 1}(\mathbf{x}, \mathbf{s})$ is convex and supermodular in \mathbf{x} . Then, by the same argument as the one used in Theorems 8 and 9, there exists an optimal stationary policy with the same structure as statement (b) in Theorem 9 that minimizes the right hand side of the ACOE.

Thus, it just remains to show our model satisfies the five conditions. We proceed in order, beginning with Condition (G). Consider again the policy π^d transmitting d^1 packets to user 1 and d^2 packets to user 2 in every slot, regardless of channel condition. Let the initial vector of buffer levels $\mathbf{x}_0 = (0, 0)$, and let the initial vector of channel conditions \mathbf{s}_0 be arbitrary. Then we

have:

$$\rho := \inf_{\boldsymbol{\pi} \in \Pi} \inf_{\substack{x \in \mathbb{R}_+^2 \\ \mathbf{s} \in \mathcal{S}}} \left\{ \limsup_{N \rightarrow \infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}}(\mathbf{x}, \mathbf{s}) \right\} \leq \limsup_{N \rightarrow \infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}^d}(\mathbf{x}_0, \mathbf{s}_0) \leq \mathbf{c}_{\max}^{\top} \mathbf{d} < \infty ,$$

where \mathbf{c}_{\max}^{\top} is defined in (163).¹⁴

The only nontrivial statement in Condition (W) is the weak continuity of the transition law. Let $\{\mathbf{x}_i\}_{i=1,2,\dots}$, $\{\mathbf{s}_i\}_{i=1,2,\dots}$, and $\{\mathbf{y}_i\}_{i=1,2,\dots}$ be sequences approaching \mathbf{x} , \mathbf{s} , and \mathbf{y} , respectively, and let Γ be a bounded, continuous function on $\mathbb{R}_+^2 \times \mathcal{S}$. We need to show:

$$\lim_{i \rightarrow \infty} \mathbb{E} [\Gamma(\mathbf{X}', \mathbf{S}') | \mathbf{X} = \mathbf{x}_i, \mathbf{S} = \mathbf{s}_i, \mathbf{Y} = \mathbf{y}_i] = \mathbb{E} [\Gamma(\mathbf{X}', \mathbf{S}') | \mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}, \mathbf{Y} = \mathbf{y}] .$$

This is true, as

$$\begin{aligned} & \lim_{i \rightarrow \infty} \mathbb{E} [\Gamma(\mathbf{X}', \mathbf{S}') | \mathbf{X} = \mathbf{x}_i, \mathbf{S} = \mathbf{s}_i, \mathbf{Y} = \mathbf{y}_i] \\ &= \lim_{i \rightarrow \infty} \sum_{\mathbf{s}' \in \mathcal{S}} \Pr(\mathbf{S}' = \mathbf{s}' | \mathbf{S} = \mathbf{s}_i) \cdot \Gamma(\mathbf{y}_i - \mathbf{d}, \mathbf{s}') \\ &= \sum_{\mathbf{s}' \in \mathcal{S}} \left[\lim_{i \rightarrow \infty} \Pr(\mathbf{S}' = \mathbf{s}' | \mathbf{S} = \mathbf{s}_i) \right] \cdot \left[\lim_{i \rightarrow \infty} \Gamma(\mathbf{y}_i - \mathbf{d}, \mathbf{s}') \right] \\ &= \sum_{\mathbf{s}' \in \mathcal{S}} \Pr(\mathbf{S}' = \mathbf{s}' | \mathbf{S} = \mathbf{s}) \cdot \Gamma(\mathbf{y} - \mathbf{d}, \mathbf{s}') \\ &= \mathbb{E} [\Gamma(\mathbf{X}', \mathbf{S}') | \mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}, \mathbf{Y} = \mathbf{y}] . \end{aligned}$$

Next, we prove Conditions (B) and (B2). Let $\alpha \in [0, 1)$ be arbitrary. For every $\mathbf{s} \in \mathcal{S}$, $V_{\infty, \alpha}(\mathbf{x}, \mathbf{s})$ is convex in \mathbf{x} , and

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} V_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) \geq \lim_{\|\mathbf{x}\| \rightarrow \infty} h(\mathbf{x} - \mathbf{d}) = \infty ,$$

so there exists an $\mathbf{x}^*(\mathbf{s}) \in \mathbb{R}_+^2$ such that:

$$\min_{\mathbf{x} \in \mathbb{R}_+^2} \{V_{\infty, \alpha}(\mathbf{x}, \mathbf{s})\} = V_{\infty, \alpha}(\mathbf{x}^*(\mathbf{s}), \mathbf{s}) .$$

¹⁴For the proof of Theorem 6, we use \tilde{c}_{\max} defined in (147) instead.

Define:

$$\mathbf{s}^* := \operatorname{argmin}_{\mathbf{s} \in \mathcal{S}} \{V_{\infty, \alpha}(\mathbf{x}^*(\mathbf{s}), \mathbf{s})\} ,$$

so that

$$m_{\infty, \alpha} = V_{\infty, \alpha}(\mathbf{x}^*(\mathbf{s}^*), \mathbf{s}^*) .$$

Define also the stationary policy $\check{\pi} = (\check{\mathbf{y}}, \check{\mathbf{y}}, \dots)$, where: $\check{\mathbf{y}}(\mathbf{x}, \mathbf{s}) := (\check{y}^1(x^1, s^1), \check{y}^2(x^2, s^2))$, and for $m \in \{1, 2\}$,

$$\check{y}^m(x^m, s^m) := \begin{cases} x^m, & \text{if } x^{m*}(\mathbf{s}^*) + d^m \leq x^m \\ x^{m*}(\mathbf{s}^*) + d^m, & \text{if } x^{m*}(\mathbf{s}^*) + d^m - \frac{P}{c_s^m} \leq x^m < x^{m*}(\mathbf{s}^*) + d^m \\ x^m + \frac{P}{c_s^m}, & \text{if } x^m < x^{m*}(\mathbf{s}^*) + d^m - \frac{P}{c_s^m} \end{cases} .$$

The stationary policy $\check{\pi}$ calls for the scheduler to allocate at most $\frac{P}{2}$ units of power for transmission to each user, and tries to bring receiver m 's buffer towards $x^{m*}(\mathbf{s}^*) + d^m$ (before transmission), regardless of the random channel conditions.¹⁵ For $m \in \{1, 2\}$, let $\tau^m(x^m, s^m)$ be the random number of time slots until receiver m 's buffer level at the beginning of a slot reaches $x^{m*}(\mathbf{s}^*)$ under policy $\check{\pi}$, starting from state (x^m, s^m) . Define also $\tau_{\max}(\mathbf{x}, \mathbf{s}) := \max\{\tau^1(x^1, s^1), \tau^2(x^2, s^2)\}$, and $\tau_{\min} := \min\{\tau^1(x^1, s^1), \tau^2(x^2, s^2)\}$. Note that if $x^m > x^{m*}(\mathbf{s}^*)$, then $\tau^m(x^m, s^m) = \left\lceil \frac{x - x^{m*}(\mathbf{s}^*)}{d^m} \right\rceil$, and the total discounted expected transmission and holding cost associated with receiver m for the first $\tau^m(x^m, s^m)$ slots is upper bounded by:

$$\begin{aligned} & \alpha^{\tau^m(x^m, s^m) - 1} \cdot c_{\max}^m \cdot d^m + \sum_{t=1}^{\tau^m(x^m, s^m)} \alpha^{t-1} \cdot h^m(x - t \cdot d^m) \\ & \leq c_{\max}^m \cdot d^m + \sum_{t=1}^{\left\lceil \frac{x - x^{m*}(\mathbf{s}^*)}{d^m} \right\rceil} h^m(x - t \cdot d^m) . \end{aligned} \quad (173)$$

¹⁵For the proof of Theorem 6, the policy $\check{\pi}$ calls for the scheduler to allocate the full P units of power for transmission to the single receiver when its buffer is below $x^*(s^*) + d$. The bounds are adjusted accordingly.

On the other hand, if $x^m \leq x^{m^*}(\mathbf{s}^*)$, $\mathbb{E}[\tau^m(x^m, s^m)]$ is finite.¹⁶ Therefore, by Wald's Lemma, the total discounted expected transmission and holding cost associated with receiver m for the first $\tau^m(x^m, s^m)$ slots is upper bounded by:

$$\sum_{t=1}^{\tau^m(x^m, s^m)} \alpha^{t-1} \cdot \left[\frac{P}{2} + h^m(x^{m^*}(\mathbf{s}^*)) \right] \leq \mathbb{E}[\tau^m(x^m, s^m)] \cdot \left[\frac{P}{2} + h^m(x^{m^*}(\mathbf{s}^*)) \right]. \quad (174)$$

So for $m \in \{1, 2\}$, we define:

$$\bar{\kappa}^m(x^m, s^m) := \begin{cases} c_{\max}^m \cdot d^m + \sum_{t=1}^{\left\lceil \frac{x - x^{m^*}(\mathbf{s}^*)}{d^m} \right\rceil} h^m(x - t \cdot d^m), & \text{if } x^{m^*}(\mathbf{s}^*) < x^m \\ \mathbb{E}[\tau^m(x^m, s^m)] \cdot \left[\frac{P}{2} + h^m(x^{m^*}(\mathbf{s}^*)) \right], & \text{if } x^m \leq x^{m^*}(\mathbf{s}^*) \end{cases}.$$

Next, let $\tau_{\text{switch}}(\mathbf{x}, \mathbf{s})$ be the random number of time slots until the state $(\mathbf{x}^*(\mathbf{s}^*), \mathbf{s}^*)$ is reached at the beginning of a slot under policy $\check{\pi}$, starting from state (\mathbf{x}, \mathbf{s}) . We define a new policy $\bar{\pi}$ that follows $\check{\pi}$ for $\tau_{\text{switch}}(\mathbf{x}, \mathbf{s})$ slots (a random stopping time), and then behaves optimally. Then we have:

$$V_{\infty, \alpha}(\mathbf{x}, \mathbf{s}) \leq V_{\infty, \alpha}^{\bar{\pi}}(\mathbf{x}, \mathbf{s}) \leq \bar{\kappa}(\mathbf{x}, \mathbf{s}) + V_{\infty, \alpha}(\mathbf{x}^*(\mathbf{s}^*), \mathbf{s}^*), \quad (175)$$

where

$$\begin{aligned} \bar{\kappa}(\mathbf{x}, \mathbf{s}) &:= \bar{\kappa}^1(x^1, s^1) + \bar{\kappa}^2(x^2, s^2) \\ &+ \mathbb{E}[\tau_{\text{switch}}(\mathbf{x}, \mathbf{s}) - \tau_{\min}(\mathbf{x}, \mathbf{s})] \cdot [\mathbf{c}_{\max}^T \mathbf{d} + h^1(x^{1^*}(\mathbf{s}^*)) + h^2(x^{2^*}(\mathbf{s}^*))]. \end{aligned} \quad (176)$$

The third term in (176) is an upper bound on the transmission and holding costs required to keep the vector of buffer levels at $\mathbf{x}^*(\mathbf{s}^*)$ while waiting for the vector of channel condition realizations to reach \mathbf{s}^* . Since the vector of channel conditions is a finite-state ergodic Markov process, this

¹⁶In order to guarantee $\mathbb{E}[\tau^m(x^m, s^m)]$ is finite, we actually need an additional assumption that $\Pr\left(\frac{P}{2} = d^m\right) < 1$. However, this assumption is harmless, for if it is not true, the channel condition does not vary over time, a scenario outside of our scope of interest.

quantity is finite. Equation (175) implies:

$$\begin{aligned}
 w_{\infty,\alpha}(\mathbf{x}, \mathbf{s}) &= V_{\infty,\alpha}(\mathbf{x}, \mathbf{s}) - m_{\infty,\alpha} \\
 &= V_{\infty,\alpha}(\mathbf{x}, \mathbf{s}) - V_{\infty,\alpha}(\mathbf{x}^*(\mathbf{s}^*), \mathbf{s}^*) \\
 &\leq \bar{\kappa}(\mathbf{x}, \mathbf{s}) < \infty .
 \end{aligned}$$

The important thing to note here is that the bounding function $\bar{\kappa}(\mathbf{x}, \mathbf{s})$ is independent of α , so Condition (B) holds. The function $\bar{\kappa}(\mathbf{x}, \mathbf{s})$ is also measurable and satisfies (167), so Condition (B2) also holds.

Finally, Condition (E) follows from the fact that for every $l \in \{1, 2, \dots\}$ and $\mathbf{s} \in \mathcal{S}$, $w_{\infty,\alpha(l)}(\cdot, \mathbf{s})$ is convex. Thus, by the finiteness of \mathcal{S} and essentially the same argument used by Fernández-Gaucherand, Marcus, and Arapostathis in [83, pp. 178-179], $\{w_{\infty,\alpha(l)}(\cdot, \cdot)\}_{l=1,2,\dots}$ is locally equi-Lipschitzian and equicontinuous.

REFERENCES

- [1] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, pp. 150–154, February 2001.
- [2] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.
- [3] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proceedings of the International Conference on Communications*, vol. 1, (Seattle, WA), pp. 331–335, June 1995.
- [4] F. Berggren and R. Jäntti, "Asymptotically fair transmission scheduling over fading channels," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 326–336, January 2004.
- [5] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, pp. 451–474, March 2003.
- [6] J. M. Holtzman, "CDMA forward link waterfilling power control," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3, (Tokyo, Japan), pp. 1663–1667, May 2000.
- [7] S. Borst and P. Whiting, "Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization," *IEEE Transactions on Vehicular Technology*, vol. 52, pp. 569–586, May 2003.
- [8] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queueing system with asynchronously varying service rates," *Probability in the Engineering and Informational Sciences*, vol. 18, pp. 191–217, 2004.

- [9] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, pp. 466–478, March 1993.
- [10] S. Shakkottai, R. Srikant, and A. Stolyar, "Pathwise optimality of the exponential scheduling rule for wireless channels," *Advances in Applied Probability*, vol. 36, no. 4, pp. 1021–1045, 2004.
- [11] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," in *Proceedings of IEEE INFOCOM*, vol. 1, (San Francisco, CA), pp. 745–755, April 2003.
- [12] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), 1999.
- [13] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, pp. 1135–1149, May 2002.
- [14] M. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," in *Proceedings of the IEEE INFOCOM*, (San Francisco, CA), pp. 311–320, March-April 2003.
- [15] H. Wang, *Opportunistic Transmission of Wireless Data Over Fading Channels Under Energy and Delay Constraints*. PhD thesis, Rutgers University, 2003.
- [16] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 125–144, January 2004.
- [17] R. A. Berry and E. M. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Processing Magazine*, vol. 21, pp. 59–69, September 2004.
- [18] D. V. Djonin and V. Krishnamurthy, "Structural results on the optimal transmission scheduling policies and costs for correlated sources and channels," in *Proceedings of the IEEE Conference on Decision and Control*, (Seville, Spain), pp. 3231–3236, December 2005.
- [19] A. Borkar, A. Karandikar, and V. S. Borkar, "Power optimal opportunistic scheduling," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, (San Francisco, CA), November 2006.
- [20] S. Kittipiyakul and T. Javidi, "Resource allocation in OFDMA with time-varying channel and bursty arrivals," *IEEE Communications Letters*, vol. 11, pp. 708–710, September 2007.
- [21] M. Agarwal, V. S. Borkar, and A. Karandikar, "Structural properties of optimal transmission policies over a randomly varying channel," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1476–1491, July 2008.
- [22] M. Goyal, A. Kumar, and V. Sharma, "Optimal cross-layer scheduling of transmissions over a fading multiaccess channel," *IEEE Transactions on Information Theory*, vol. 54, pp. 3518–3537, August 2008.
- [23] S. Kittipiyakul and T. Javidi, "Delay-optimal server allocation in multi-queue multi-server systems with time-varying connectivities," *IEEE Transactions on Information Theory*, to appear.
- [24] W. Chen, U. Mitra, and M. J. Neely, "Energy-efficient scheduling with individual delay constraints over a fading channel," in *Proceedings of the International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, (Limassol, Cyprus), April 2007.
- [25] W. Chen, U. Mitra, and M. J. Neely, "Energy-efficient scheduling with individual delay constraints over a fading channel," *Wireless Networks*, vol. 15, pp. 601–618, July 2009.

- [26] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proceedings of the IEEE Infocom*, vol. 2, (San Francisco, CA), pp. 1095–1105, April 2003.
- [27] A. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal transmission scheduling over a fading channel with energy and deadline constraints," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 630–641, March 2006.
- [28] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," in *Proceedings of the IEEE International Symposium on Information Theory*, (Toronto, Canada), July 2008.
- [29] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 1866–1875, April 2009.
- [30] J. Lee and N. Jindal, "Delay constrained scheduling over fading channels: Optimal policies for monomial energy-cost functions," in *Proceedings of the IEEE International Conference on Communications*, (Dresden, Germany), June 2009.
- [31] J. Lee and N. Jindal, "Asymptotically optimal policies for hard-deadline scheduling over fading channels," *submitted to IEEE Transactions on Information Theory*, 2009.
- [32] E. Uysal-Biyikoglu and A. E. Gamal, "On adaptive transmission for energy efficiency in wireless data networks," *IEEE Transactions on Information Theory*, vol. 50, pp. 3081–3094, December 2004.
- [33] A. Tarello, J. Sun, M. Zafer, and E. Modiano, "Minimum energy transmission scheduling subject to deadline constraints," *ACM Wireless Networks*, vol. 14, pp. 633–645, October 2008.
- [34] D. Shuman and M. Liu, "Opportunistic scheduling with deadline constraints in wireless networks," in *Performance Models and Risk Management in Communication Systems* (N. Gulpinar, P. Harrison, and B. Rustem, eds.), Springer, to appear.
- [35] X. Liu, E. K. P. Chong, and N. B. Shroff, "Optimal opportunistic scheduling in wireless networks," in *Proceedings of the Vehicular Technology Conference*, vol. 3, (Orlando, FL), pp. 1417–1421, October 2003.
- [36] X. Liu, N. B. Shroff, and E. K. P. Chong, "Opportunistic scheduling: An illustration of cross-layer design," *Telecommunications Review*, vol. 14, pp. 947–959, December 2004.
- [37] B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," in *Proceedings of the International Conference on Image Processing*, vol. 1, (Rochester, NY), pp. 9–12, September 2002.
- [38] Y. Li and N. Bambos, "Power-controlled wireless links for media streaming applications," in *Proceedings of the Wireless Telecommunications Symposium*, (Pomona, CA), pp. 102–111, May 2004.
- [39] Y. Li, A. Markopoulou, N. Bambos, and J. Apostolopoulos, "Joint power-playout control for media streaming over wireless links," *IEEE Transactions on Multimedia*, vol. 8, pp. 830–843, August 2006.
- [40] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 1710–1720, December 2003.
- [41] J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica*, vol. 65, pp. 487–516, May 1997.
- [42] C.-S. Chow and J. N. Tsitsiklis, "The complexity of dynamic programming," *Journal of Complexity*, vol. 5, pp. 466–488, December 1989.
- [43] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes*. Springer-Verlag, 1996.
- [44] D. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, 1996.

- [45] T. Fabian, J. L. Fisher, M. W. Sasieni, and A. Yardeni, "Purchasing raw material on a fluctuating market," *Operations Research*, vol. 7, pp. 107–122, January-February 1959.
- [46] B. Kalyon, "Stochastic prices in a single-item inventory purchasing model," *Operations Research*, vol. 19, pp. 1434–1458, October 1971.
- [47] B. G. Kingsman, "Commodity purchasing," *Operational Research Quarterly*, vol. 20, pp. 59–80, 1969.
- [48] B. G. Kingsman, *Commodity Purchasing in Uncertain Fluctuating Price Markets*. PhD thesis, University of Lancaster, 1969.
- [49] V. F. Magirou, "Stockpiling under price uncertainty and storage capacity constraints," *European Journal of Operational Research*, vol. 11, pp. 233–246, 1982.
- [50] V. F. Magirou, "Comments on 'On Optimal Inventory Policies when Ordering Prices are Random' by Kamal Golabi," *Operations Research*, vol. 35, pp. 930–931, November-December 1987.
- [51] K. Golabi, "A single-item inventory model with stochastic prices," in *Proceedings of the Second International Symposium on Inventories*, (Budapest, Hungary), pp. 687–697, 1982.
- [52] K. Golabi, "Optimal inventory policies when ordering prices are random," *Operations Research*, vol. 33, pp. 575–588, May-June 1985.
- [53] A. Federgruen and P. Zipkin, "An inventory model with limited production capacity and uncertain demands II. The discounted-cost criterion," *Mathematics of Operations Research*, vol. 11, pp. 208–215, May 1986.
- [54] S. Tayur, "Computing the optimal policy for capacitated inventory models," *Communications in Statistics. Stochastic Models*, vol. 9, no. 4, pp. 585–598, 1993.
- [55] M. J. Sobel, "Making short-run changes in production when the employment level is fixed," *Operations Research*, vol. 18, pp. 35–51, January-February 1970.
- [56] A. Bensoussan, M. Crouhy, and J.-M. Proth, *Mathematical Theory of Production Planning*. Elsevier Science, 1983.
- [57] F. C. Zahrn, *Studies of Inventory Control and Capacity Planning with Multiple Sources*. PhD thesis, Georgia Institute of Technology, August 2009.
- [58] E. L. Porteus, "Stochastic inventory theory," in *Stochastic Models* (D. P. Heyman and M. J. Sobel, eds.), pp. 605–652, Elsevier Science, 1990.
- [59] S. Karlin, "Optimal inventory policy for the Arrow-Harris-Marschak dynamic model," in *Studies in the Mathematical Theory of Inventory and Production* (K. J. Arrow, S. Karlin, and H. Scarf, eds.), pp. 135–154, Stanford University Press, 1958.
- [60] R. Evans, "Inventory control of a multiproduct system with a limited production resource," *Naval Research Logistics Quarterly*, vol. 14, no. 2, pp. 173–184, 1967.
- [61] G. A. DeCroix and A. Arreola-Risa, "Optimal production and inventory policy for multiple products under resource constraints," *Management Science*, vol. 44, pp. 950–961, July 1998.
- [62] S. X. Chen, "The optimality of hedging point policies for stochastic two-product flexible manufacturing systems," *Operations Research*, vol. 52, pp. 312–322, March-April 2004.
- [63] G. Janakiraman, M. Nagarajan, and S. Veeraraghavan, "Simple policies for capacitated multi-product inventory systems," *submitted to Management Science*, August 2008.

- [64] R. Bellman, I. Glicksberg, and O. Gross, “On the optimal inventory equation,” *Management Science*, vol. 2, pp. 83–104, October 1955.
- [65] J. T. Hawkins, *A Lagrangian Decomposition Approach to Weakly Coupled Dynamic Optimization Problems and its Applications*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [66] D. Adelman and A. J. Mersereau, “Relaxations of weakly coupled stochastic dynamic programs,” *Operations Research*, vol. 56, pp. 712–727, May-June 2008.
- [67] E. Antoniadou, *Lattice Programming and Economic Optimization*. PhD thesis, Stanford University, 1996.
- [68] E. Antoniadou, “Comparative statics for the consumer problem,” *Economic Theory*, vol. 31, no. 1, pp. 189–203, 2007.
- [69] D. Shuman, *From Sleeping to Stockpiling: Energy Conservation via Stochastic Scheduling in Wireless Networks*. PhD thesis, University of Michigan, Ann Arbor, 2010.
- [70] D. B. Brown, J. E. Smith, and P. Sun, “Information relaxations and duality in stochastic dynamic programs,” *Operations Research*, to appear.
- [71] P. J. Schweitzer and A. Seidmann, “Generalized polynomial approximations in Markovian decision processes,” *Journal of Mathematical Analysis and Applications*, vol. 110, pp. 568–582, September 1985.
- [72] D. P. de Farias and B. Van Roy, “The linear programming approach to approximate dynamic programming,” *Operations Research*, vol. 51, pp. 850–865, November-December 2003.
- [73] W. Karush, “A theorem in convex programming,” *Naval Research Logistics Quarterly*, vol. 6, pp. 245–260, September 1959.
- [74] E. L. Porteus, *Foundations of Stochastic Inventory Theory*. Stanford University Press, 2002.
- [75] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [76] D. M. Topkis, *Supermodularity and Complementarity*. Princeton University Press, 1998.
- [77] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [78] J. E. Smith and K. F. McCardle, “Structural properties of stochastic dynamic programs,” *Operations Research*, vol. 50, no. 5, pp. 796–809, 2002.
- [79] D. L. Iglehart, “Optimality of (s,S) policies in the infinite horizon dynamic inventory problem,” *Management Science*, vol. 9, pp. 259–267, January 1963.
- [80] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research*, vol. II. McGraw-Hill, 1984.
- [81] M. J. Sobel, “Production smoothing with stochastic demand II: Infinite horizon case,” *Management Science*, vol. 17, pp. 724–735, July 1971.
- [82] M. Schäl, “Average optimality in dynamic programming with general state space,” *Mathematics of Operations Research*, vol. 18, pp. 163–172, February 1993.
- [83] E. Fernández-Gaucherand, S. I. Marcus, and A. Arapostathis, “Structured solutions for stochastic control problems,” in *Stochastic Theory and Adaptive Control*, pp. 172–184, Springer, 1992.
- [84] R. Montes-de-Oca, “The average cost optimality equation for Markov control processes on Borel spaces,” *Systems and Control Letters*, vol. 22, pp. 351–357, 1994.