Quality-Driven Resource Allocation for Full-Duplex Delay-Constrained Wireless Video Transmissions

Chuang Ye, M. Cenk Gursoy, *Senior Member, IEEE*, and Senem Velipasalar, *Senior Member, IEEE*

***Abstract*— In this paper, wireless video transmission over full-duplex channels under total bandwidth and minimum required quality constraints is studied. In order to provide the desired performance levels to the end-users in real-time video transmissions, quality of service requirements such as statistical delay constraints are also considered. Effective capacity is used as the throughput metric in the presence of such statistical delay constraints since deterministic delay bounds are difficult to guarantee due to the time-varying nature of wireless fading channels. A communication scenario with multiple pairs of users in which different users have different delay requirements is addressed. Following characterizations from the rate-distortion theory, a logarithmic model of the quality-rate relation is used for predicting the quality of the reconstructed video in terms of the peak signal-to-noise ratio at the receiver side. Since the optimization problem is not concave or convex, the optimal bandwidth and power allocation policies that maximize the weighted sum video quality subject to total bandwidth, maximum transmission power level and minimum required quality constraints are derived by using monotonic optimization theory.**

***Index Terms*— Delay constraints, effective capacity, full-duplex operation, monotonic optimization, quality of service, rate distortion, resource allocation.**

I. Introduction

RECENTLY, with rapid developments in communication technology, multimedia applications such as video telephony, teleconferencing, and video streaming which are delay sensitive and bandwidth intensive, have started becoming predominant in data transmission over wireless networks. For instance, as revealed in [1], mobile video traffic accounted for 60% of the total mobile data traffic in 2016, and more than three-fourths of the global mobile data traffic is expected to be video traffic by 2021. Indeed, mobile video has the highest growth rate of any application category measured among the mobile data traffic types. Such dramatic increase in wireless video traffic, coupled with the limited spectrum resources, brings a great challenge to today's wireless networks. Therefore, it is important to improve the wireless network capacity by allocating the limited resource efficiently. In such multimedia applications, certain quality of service (QoS) guarantees also need to be provided in order to satisfy the performance requirements of the end-users. For instance, in order to ensure a satisfactory user experience, bounds on time delay are imposed in real-time video transmissions. The strictness of the delay constraints varies based on the specific wireless multimedia application. For instance, live video streaming may tolerate some delay whereas bidirectional video conferencing requires much more stringent time delay bounds on the order of few milliseconds in order to guarantee satisfactory user experience. Supporting such QoS requirements with stringent delay limitations requires larger transmission rates that can be achieved by using more resources such as bandwidth and power, and facing less interference. Therefore, it is critical to allocate the limited resources efficiently taking into account the QoS requirements of different users in the wireless network.

Khalek *et al.* [2] proposed a strategy to maximize the sum quality of the received reconstructed videos subject to different delay constraints on different users and a total bandwidth constraint in a multiuser setup by allocating the optimal amount of bandwidth to each user in a downlink wireless network. They also derived user admission and scheduling policies that enable selecting a maximal user subset such that all selected users can meet their statistical delay requirements. A content-aware framework for spectrum- and energy-efficient mobile association and resource allocation in wireless heterogeneous networks was proposed in [3]. Two content-aware performance metrics, namely quality-of-experience-aware spectral efficiency (QSE) and quality-of-experience-aware energy efficiency (QEE), were used to capture spectrum usage and energy consumption from the perspective of video quality. The goal was to obtain the optimal system level QSE and QEE by determining the mobile association and allocating the resources optimally via nonlinear fractional programming approach and dual decomposition method. In this work, delay QoS constraints were not considered. On the other hand, reference [4] addressed the maximization of the system throughput subject to delay QoS and average power constraints for time-division multiple access (TDMA) communication links. [5] proposed a QoS-driven power and rate adaptation scheme that aims at maximizing the throughput of multichannel systems subject to a given delay QoS constraint over wireless links. Multichannel communication can achieve high throughput and satisfy stringent QoS requirements simultaneously. Wang *et al.* [6] developed an optimal power allocation

Manuscript received April 23, 2017; revised October 8, 2017 and February 10, 2018; accepted April 2, 2018. Date of publication April 16, 2018; date of current version September 14, 2018. This work was supported by the National Science Foundation under Grants CNS-1443966, ECCS-1443994, and CCF-1618615. The associate editor coordinating the review of this paper and approving it for publication was A. Tajer. *(Corresponding author: M. Cenk Gursoy.)*

The authors are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: chye@syr.edu; mcgursoy@syr.edu; svelipas@syr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2827063

scheme for the cognitive network with the goal of maximizing the effective capacity of the secondary user link under constraints on the primary user's outage probability and secondary user's average and peak transmission power. The scheme also satisfied the QoS requirements of both secondary users and primary users simultaneously. Statistical QoS provisioning in next generation heterogeneous mobile cellular networks was investigated in [7]. Under certain assumptions, a lower bound for the system performance was introduced in order to facilitate the analysis of the effective capacity. Based on the proposed lower bound, performance of dense next generation heterogeneous cellular networks under statistical QoS requirements was analyzed by building a scalable mathematical framework.

Cheng *et al.* [8] proposed a QoS-driven power allocation scheme for full-duplex wireless links with the goal of maximizing the overall effective capacity under a given delay QoS constraint. Two models namely local transmit power related self-interference (LTPRS) model and local transmit power unrelated self-interference (LTPUS) were built to analyze the full-duplex transmission, respectively. However, an approximation of the sum Shannon capacity was used in the formulation of the effective capacity under the assumption that the signal-to-interference-plus-noise ratio is much larger than 1. Reference [9] considered the problem of distributed power allocation in a full-duplex wireless network consisting of multiple pairs of nodes with the goal of maximizing the network-wide capacity. Shannon capacity was used as the performance metric and the optimal transmission powers for the full-duplex transmitters were derived based on the high SINR approximation and a more general approximation method for the logarithm function.

The problem of joint subchannel allocation and power control was discussed in many studies. For instance, resource allocation in multicell uplink orthogonal frequency division multiple access (OFDMA) systems was considered in [10], and the problem was solved via noncooperative games for subcarrier allocation and transmit power control. Reference [11] proposed a joint power control and subchannel allocation for OFDMA femtocell networking using distributed auction game in order to minimize the total power radiated by the femtocell base station and guaranteeing the throughput. Reference [12] considered the problem of joint subcarrier and power allocation for the downlink of a multiuser OFDM cellular network in order to minimize the power consumption subject to meeting the target rates of all users in the network. Wang *et al.* [13] considered the adaptive subcarrier assignment and fair power control strategy that minimize a cost function of average relay powers in multiuser wireless OFDM networks.

However, the aforementioned works have not considered statistical QoS requirements, bandwidth limitations, power limitations and interference jointly in full-duplex wireless networks. In this paper, we address the problem of maximizing the weighted sum quality of reconstructed videos at the receivers subject to total bandwidth, minimum video quality, maximum transmission power and delay QoS constraints by allocating the bandwidth and determining the optimal power level for each user when statistical channel side information (CSI) is available in the full-duplex wireless network. Since the optimization problem is neither a concave nor convex problem due to the existence of the interference, we employ the monotonic optimization (MO) framework.[1] Our more specific contributions include the following:

1) We reformulate the optimization problem as a monotonic optimization problem, and propose a framework to study full-duplex communication via monotonic optimization.
2) We derive several key properties of the optimal solution space.
3) We develop algorithms to efficiently determine the optimal resource allocation policies. In particular, we develop algorithms for enclosing polyblock initialization, projection onto the upper boundary, and iterative derivation of new enclosing polyblocks.
4) We analyze the impact of important system parameters (e.g., video quality parameters, QoS constraints, and weights) on the optimal resource allocation strategies and received video quality in terms of peak signal-to-noise ratio.

The remainder of this paper is organized as follows: The system model is presented in Section II. Statistical QoS guarantees, effective capacity as a throughput metric, and quality-rate model are described as preliminary concepts in Section III. The optimization problems are formulated and the optimal policies are derived in Section IV. Simulation results are presented and discussed in Section V. Finally, we conclude the paper in Section VI. Proofs are relegated to the Appendix.

II. System Model

Fig. 1 depicts the considered system model. We consider K pairs of users, denoted as $(U_{1,1}, U_{2,1})$, $(U_{1,2}, U_{2,2})$, ..., $(U_{1,K}, U_{2,K})$,[2] orthogonally sharing a total bandwidth of B Hz in full-duplex mode. Specifically, the kth full-duplex link between $U_{1,k}$ and $U_{2,k}$ is allocated a bandwidth of B_k Hz for the transmission of the video data under the constraint that the total bandwidth is $\sum_{k=1}^{K} B_k = B$. It is assumed that flat fading is experienced in each subchannel. The channel coherence time is denoted by T_c, and the timescale of video rate adaptation is much larger than T_c in practice for video transmission since video source rate is adapted at the group of pictures (GOP) time scale which is measured in seconds. The case in which the channel state changes faster than the source rate is considered in our system since if the fading channel state varies at the same timescale as the source rate, statistical delay guarantees become less interesting [2].

The practical application of this model includes, for instance, scenarios in which device-to-device (D2D) users exchange multimedia data (e.g., via social media sites) or conduct teleconferencing (i.e., engage in interactive video) in full-duplex mode. Assuming the availability of only statistical channel side information (CSI), base

[1] We note that we have recently addressed in the conference paper [14] a simpler scenario with only a single pair of users communicating in full-duplex mode and studied only optimal power allocation via monotonic optimization.

[2] Throughout the paper, the subscripts $(1, k)$ and $(2, k)$ are used for parameters and notations related to users 1 and 2 of the k^{th} pair, respectively.

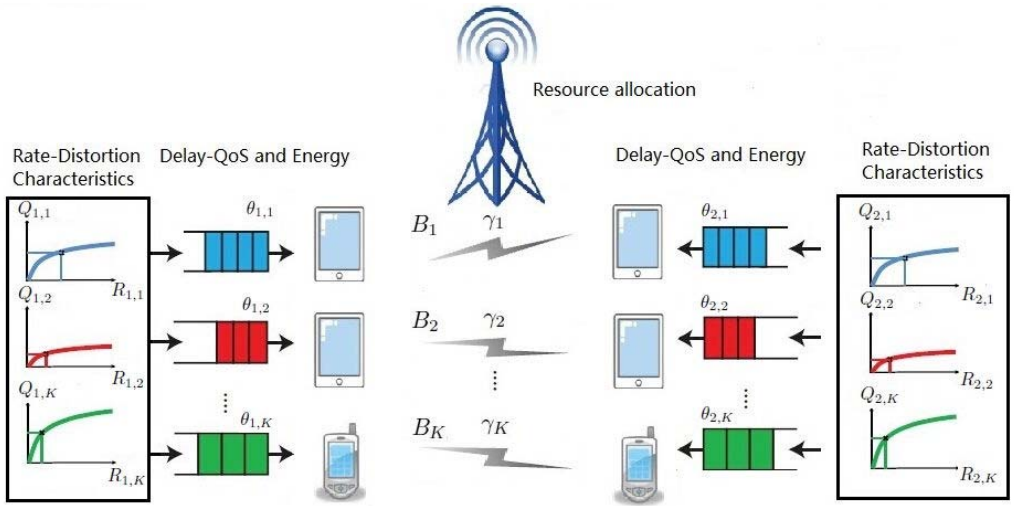


Fig. 1. Wireless system model in which each pair of users communicates in full-duplex mode under quality and delay constraints.

station acts as a coordinating agent and performs quality-driven resource allocation. Or in a different scenario, we can have one base station performing full-duplex multimedia communication with multiple users over different subchannels (e.g., via othogonal frequency division multiple access (OFDMA)). In this case, all the users on the left-hand side of Fig. 1 essentially represent (or collapse to) a single base station in which there are multiple buffers and multiple flows of multimedia data to be sent to different users on the right-hand side. Base station again performs quality-driven resource allocation.

III. Preliminaries

A. Notations

Throughout this paper, vectors are denoted by boldface letters, the j-th entry of a vector $\mathbf{x}$ is denoted by x_j. $\mathcal{R}$ and $\mathcal{R}_+$ denote the set of real numbers and nonnegative real numbers, respectively. $\mathcal{R}^n$ and $\mathcal{R}_+^n$ denote the space of n-dimensional real-valued vectors and nonnegative real-valued vectors, respectively. For any two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\mathbf{x} \geq \mathbf{y}$ if $x_j \geq y_j$ for all $j = 1, 2, \ldots, n$. $\cup$, $\cap$ and $\setminus$ represent set union, set intersection and set difference operators, respectively. $\mathbf{e}_j \in \mathcal{R}^n$ denotes the j-th unit vector of $\mathcal{R}^n$, i.e., the vector such that $e_j = 1$ and $e_i = 0$ for all $i \neq j$.

B. Delay QoS Constraints and Effective Capacity

In wireless video transmissions, queue length in the buffer is subject to limitations to control the queueing delay. The behavior of the queue length in wireless communication is addressed in [15]. In particular, we assume that the overflow probabilities in the buffer storing the data to be transmitted at each pair of users decay exponentially for large buffer threshold, i.e.,

$$\Pr\{l_{i,k} > l_{i,k}^{th}\} \approx e^{-\theta_{i,k} l_{i,k}^{th}}, i \in \mathcal{I}, k \in \mathcal{K} \tag{1}$$

where $l_{i,k}$ and $l_{i,k}^{th}$ are the queue length and threshold at $U_{i,k}$, respectively, and $\mathcal{I} = \{1, 2\}$ and $\mathcal{K} = \{1, 2, \ldots, K\}$. $\theta_{i,k}$, referred to as the QoS exponent, determines the decay rate of the buffer overflow probability, and characterizes how strict the queueing/delay constraints are[3]. Larger $\theta_{i,k}$ leads to more stringent QoS requirements while smaller $\theta_{i,k}$ represents looser QoS requirements. In the presence of such QoS requirements, two key performance metrics are effective capacity and effective bandwidth. Effective capacity (EC), $C(\theta_{i,k})$, characterizes the maximum constant arrival rate which can be supported by the service process (i.e., wireless transmissions) in the presence of statistical buffer overflow constraints specified by the QoS exponent $\theta_{i,k}$. Effective bandwidth, $A(\theta_{i,k})$, provides the minimum constant service rate needed to guarantee that the overflow probability decades with rate specified by $\theta_{i,k}$ for the given arrival process $\{A\}$.

Now, we express the EC formulations for the pair of users operating in full-duplex mode. Considering independent and identically distributed fading in each coherence block of duration T_c, we can write the EC expressions for the k^{th} pair of users as [5]

$$\begin{aligned} C_{1,k}(\theta_{1,k}) &= -\frac{1}{\theta_{1,k} T_c} \ln\left(\mathbb{E}_{\gamma_k}\{e^{-\theta_{1,k} r_{1,k}}\}\right) \\ &= -\frac{1}{\theta_{1,k} T_c} \ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{1,k} B_k T_c \log\left(1+\frac{P_{1,k}\gamma_k}{N_0 B_k + I_{2,k}}\right)}\right\}\right) \end{aligned} \tag{2}$$

$$\begin{aligned} C_{2,k}(\theta_{2,k}) &= -\frac{1}{\theta_{2,k} T_c} \ln\left(\mathbb{E}_{\gamma_k}\{e^{-\theta_{2,k} r_{2,k}}\}\right) \\ &= -\frac{1}{\theta_{2,k} T_c} \ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{2,k} B_k T_c \log\left(1+\frac{P_{2,k}\gamma_k}{N_0 B_k + I_{1,k}}\right)}\right\}\right) \end{aligned} \tag{3}$$

where B_k is the allocated bandwidth for the full-duplex communication of these users, $P_{i,k}$ is the power of user $U_{i,k}$, and $\theta_{i,k}$ is the QoS exponent of $U_{i,k}$. Moreover, N_0 is the power spectral density of the background Gaussian noise, and $I_{1,k}$ and $I_{2,k}$ are the self-interference terms at $U_{1,k}$ and $U_{2,k}$, respectively.

[3]Given the threshold $l_{i,k}^{th}$ and the overflow probability limit $\Pr\{l_{i,k} > l_{i,k}^{th}\} = \varepsilon$, the corresponding QoS exponent value can be obtained from $\theta_{i,k} = \frac{1}{l_{i,k}^{th}} \log \frac{1}{\varepsilon}$. In a multimedia application, threshold and the overflow probability limit are influenced by the buffer size, video frame sizes, and the given QoS and video quality requirements.

The EC should be equal to the effective bandwidth of the arrival process for the given QoS exponent θ [16] in order to support the highest arrival rates. For constant arrival rate R, the effective bandwidth of the arrival process is $A(\theta_{i,k}) = R$. Therefore, the maximum constant arrival rates at users $U_{1,k}$ and $U_{2,k}$ can be expressed, respectively, as

$$R_{1,k} = A_{1,k}(\theta_{1,k}) = C_{1,k}(\theta_{1,k}) = -\frac{1}{\theta_{1,k}T_c}\ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{1,k}B_kT_c\log\left(1+\frac{P_{1,k}\gamma_k}{N_0\ B_k+I_{2,k}}\right)}\right\}\right), \tag{4}$$

$$R_{2,k} = A_{2,k}(\theta_{2,k}) = C_{2,k}(\theta_{2,k}) - \frac{1}{\theta_{2,k}T_c}\ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{2,k}B_kT_c\log\left(1+\frac{P_{2,k}\gamma_k}{N_0B_k+I_{1,k}}\right)}\right\}\right). \tag{5}$$

C. Video Quality-Rate Model

Lossy data compression, which focuses on the tradeoff between the distortion and bit rate, is used in video coding algorithms, where an increased distortion leads to a decreased rate and vice-versa. Rate-distortion (R-D) theory addresses the problem of determining the minimal source bit rate so that the distortion of the reconstructed data at the receiver does not exceed a given distortion value. Thus, the R-D function can estimate the bit rate at given distortion, or estimate the distortion at a given bit rate. Moreover, operational R-D (ORD) theory is applied to lossy data compression with finite number of possible R-D pairs, and the ORD function shows that the bit rate is a convex function of distortion. In [17], the quality of video is measured in terms of the reversed difference mean opinion score (RDMOS), and the following rate-quality model to predict $q_u(t)$ using the video data rate $r_u(t)$ is employed:

$$q_u(t) = \alpha_u(t)\log(r_u(t)) + \beta_u(t) \tag{6}$$

where model parameters $\alpha_u(t)$ and $\beta_u(t)$ can be determined by minimizing the prediction error. Also several R-D models are proposed in [18], in which the quality is measured in terms of peak signal-to-noise ratio (PSNR). The exponential model for the rate-PSNR curve is used in our paper. Thus, PSNR-rate curve is described by a logarithmic model and can be expressed as follows:

$$Q_{i,k} = a_{i,k}\ln(R_{i,k}) + b_{i,k} \tag{7}$$

where $R_{i,k}$ and $Q_{i,k}$ are the arrival rate and PSNR of the transmitted video at $U_{i,k}$, respectively, and $a_{i,k}$ and $b_{i,k}$ are the parameters that can be determined by minimizing the prediction error. As discussed in the previous subsection, the source rate of the channel is given by the effective capacity (which quantifies the maximum constant arrival rate), i.e., $R_{i,k} = C_{i,k}$ in order to achieve the maximum video quality.

IV. Weighted Sum Quality-Maximizing Policies

In this section, optimization problems are formulated to maximize the weighted sum video quality subject to maximum transmission power and minimum video quality constraints at each user and a total bandwidth constraint. More specifically, we address the optimal allocation of bandwidth and the determination of transmission power levels assuming the availability of statistical CSI. It is assumed that each user just has one antenna for transmitting and receiving the data. Thus, the self-interference just depends on the self-transmission power, and the maximum constant arrival rate in (4) and (5) can be rewritten as

$$R_{1,k} = C_{1,k}(\theta_{1,k}) = -\frac{1}{\theta_{1,k}T_c}\ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{1,k}B_kT_c\log\left(1+\frac{P_{1,k}\gamma_k}{N_0\ B_k+\mu_{2,k}P_{2,k}}\right)}\right\}\right) \tag{8}$$

$$R_{2,k} = C_{2,k}(\theta_{2,k}) = -\frac{1}{\theta_{2,k}T_c}\ln\left(\mathbb{E}_{\gamma_k}\left\{e^{-\theta_{2,k}B_kT_c\log\left(1+\frac{P_{2,k}\gamma_k}{N_0B_k+\mu_{1,k}P_{1,k}}\right)}\right\}\right) \tag{9}$$

where $\mu_{i,k} \in (0,1]$ is the self-interference suppression factor at $U_{i,k}$. We can now express the weighted sum video quality at users $U_{1,k}$ and $U_{2,k}$ as

$$Q_k = \omega_{1,k}Q_{1,k} + \omega_{2,k}Q_{2,k} = \sum_{i=1}^{2}\omega_{i,k}\big(a_{i,k}\ln(R_{i,k}) + b_{i,k}\big), \tag{10}$$

where $\omega_{i,k} \in [0,1]$ denotes the weight for the quality of the video transmitted by user $U_{i,k}$ such that $\sum_{k=1}^{K}\sum_{i=1}^{2}\omega_{i,k} = 1$.

Now, the problem of maximizing the overall sum video quality of all users over bandwidth and power allocation strategies can be expressed as follows:

$$\max_{\mathbf{B},\mathbf{P_1},\mathbf{P_2}} \sum_{k=1}^{K}\sum_{i=1}^{2}\Big(\omega_{i,k}Q_{i,k}(R_{i,k})\Big) \tag{11a}$$

$$\text{s.t.} \sum_{k=1}^{K} B_k \le B; \quad B_k \ge 0, \quad \forall k \in \mathcal{K} \tag{11b}$$

$$P_{i,k} \le P_{i,k}^{max}; \quad P_{i,k} \ge 0, \quad \forall i \in \mathcal{I}, k \in \mathcal{K} \tag{11c}$$

$$Q_{i,k}(R_{i,k}) \ge Q_{i,k}^{min}, \quad \forall i \in \mathcal{I}, k \in \mathcal{K} \tag{11d}$$

Above, (11b) is the total bandwidth constraint, (11c) is the maximum transmission power constraint at each user and (11d) is the minimum required video quality constraint. Specifically, $P_{i,k}^{max}$ and $Q_{i,k}^{min}$ are the maximum available transmission power and minimum transmitted video quality at $U_{i,k}$, respectively. $\mathbf{B}$, $\mathbf{P}_1$ and $\mathbf{P}_2$ are $K \times 1$ vectors of bandwidth allocated to each link, power allocated to $U_{1,k}$ and $U_{2,k}$, respectively. The feasible set of $\mathbf{B}$ is denoted by $\mathcal{B} = \{\mathbf{B}|\sum_{k=1}^{K}B_k \le B\}$, and the feasible sets of $\mathbf{P}_1$ and $\mathbf{P}_2$ are denoted by $\mathcal{P}_1 = \{\mathbf{P}_1|P_{1,k} \le P_{1,k}^{max}, \forall k \in \mathcal{K}\}$ and $\mathcal{P}_2 = \{\mathbf{P}_2|P_{2,k} \le P_{2,k}^{max}, \forall k \in \mathcal{K}\}$, respectively.

A. Problem Reformulation as Monotonic Optimization

First, we introduce some definitions used in monotonic optimization (MO) from [19], and then show that problem (11) can be reformulated as an MO problem.

Definition 1 (Box): For two vectors $\mathbf{a} \in \mathcal{R}^n$, $\mathbf{b} \in \mathcal{R}^n$ with $\mathbf{a} \le \mathbf{b}$, the box $[\mathbf{a}, \mathbf{b}]$ is the set of all vectors $\mathbf{x} \in \mathcal{R}^n$ satisfying

$\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$. In other words, a hyperrectangle $[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} | a_j \leq x_j \leq b_j, j = 1, 2, ..., n\}$ is referred as a box.

Definition 2 (Normal Set): A set $\mathcal{G} \subset \mathcal{R}_+^n$ (the n-dimensional nonnegative real domain) is normal if for any element $\mathbf{x} \in \mathcal{G}$, all other elements $\mathbf{x}'$ such that $\mathbf{0} \leq \mathbf{x}' \leq \mathbf{x}$ are in the same set $\mathcal{G}$. In other words, $\mathcal{G} \subset \mathcal{R}_+^n$ is normal if for any $\mathbf{x} \in \mathcal{G}$, the set $[\mathbf{0}, \mathbf{x}] \subset \mathcal{G}$.

Definition 3 (Conormal Set): A set $\mathcal{H} \subset \mathcal{R}_+^n$ is conormal if for any element $\mathbf{x} \in \mathcal{H}$, all other elements $\mathbf{x}'$ such that $\mathbf{x}' \geq \mathbf{x}$ are in the same set $\mathcal{H}$. In other words, a set $\mathcal{H}$ is conormal in $[\mathbf{0}, \mathbf{b}]$ if for any $\mathbf{x} \in \mathcal{H}$, $[\mathbf{x}, \mathbf{b}] \subset \mathcal{H}$.

Definition 4 (Upper Boundary): An element $\bar{\mathbf{x}}$ of a normal closed set $\mathcal{G}$ is an upper boundary point of $\mathcal{G}$ if $\mathcal{G} \cap \{\mathbf{x} \in \mathcal{R}_+^n | \mathbf{x} > \bar{\mathbf{x}}\} = \emptyset$. The set of all upper boundary points of the set $\mathcal{G}$ is called its upper boundary and denoted by $\partial^+ \mathcal{G}$.

Definition 5 (Polyblocks): A set $\mathcal{S} \subset \mathcal{R}_+^n$ is a polyblock if it is a union of a finite number of boxes $[\mathbf{0}, \mathbf{z}]$, where $\mathbf{z} \in \mathcal{T}$ and $|\mathcal{T}| < +\infty$. The set $\mathcal{T}$ is the vertex set of the polyblock.

Definition 6 (Proper): An element $\mathbf{x} \in \mathcal{T}$ is said to be proper if there is no $\mathbf{x}' \in \mathcal{T}$ such that $\mathbf{x}' \neq \mathbf{x}$ and $\mathbf{x}' \geq \mathbf{x}$. If every element $\mathbf{x}' \in \mathcal{T}$ is proper, then the set $\mathcal{T}$ is a proper set.

From [19], an optimization problem belongs to the class of MO if it can be represented in the following form:

$$\max f(\mathbf{x}) \tag{12}$$

$$\text{s.t. } \mathbf{x} \in \mathcal{G} \cap \mathcal{H} \tag{13}$$

where $f(\mathbf{x}) : \mathcal{R}_+^n \to \mathcal{R}$ is an increasing function, $\mathcal{G} \subset [\mathbf{0}, \mathbf{b}] \subset \mathcal{R}_+^n$ is a compact normal set, and $\mathcal{H}$ is a closed conormal set on $[\mathbf{0}, \mathbf{b}]$. A simpler case is the one in which $\mathcal{H}$ is not present in the formulation (which occurs e.g., if the conormal set $\mathcal{H}$ is box $[\mathbf{0}, \mathbf{b}]$). In general, if $\mathcal{G} \cap \mathcal{H} \neq \emptyset$, the problem is considered feasible.

We note that it is not possible to obtain the optimal solution of (11) based on the theory of convex optimization [20] because of the non-convexity of the optimization problem in (11) in terms of $P_{i,k}$ and B_k jointly. This non-convexity is primarily due to the presence of the self-interference terms. In operations research, monotonicity is regarded as another important property for effectively solving an optimization problem. Therefore, we follow the approach to solve the non-convex problem (11) by transforming it into an MO problem, and then solving the corresponding MO problem based on recent advances in monotonic optimization [21].

We first rewrite the objective function in (11a) in terms of auxiliary variables[4]. Let $\mathbf{Y}$denote the vector $(Y_1, Y_2, \ldots, Y_{2K})$ with Y_j being the j-th component of $\mathbf{Y}$. We define the function

$$\Phi(\mathbf{Y}) = \sum_{k=1}^{K} \sum_{i=1}^{2} \omega_{i,k} \left[a_{i,k} \ln \left(\frac{1}{\theta_{i,k} T_c} \ln(Y_{(i-1)K+k}) \right) + b_{i,k} \right].$$

It is easy to see that $\Phi(\mathbf{Y})$ is an increasing function of $\mathbf{Y}$ on $\mathcal{R}_+^{2\times K}$. In other words, for any two vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$, $\Phi(\mathbf{Y}_1) \geq \Phi(\mathbf{Y}_2)$ if $\mathbf{Y}_1 \geq \mathbf{Y}_2$. Now, problem (11) can be rewritten in the MO formulation as

$$\max \Phi(\mathbf{Y}) = \sum_{k=1}^{K} \sum_{i=1}^{2} \omega_{i,k} \left[a_{i,k} \ln \left(\frac{\ln(Y_{(i-1)K+k})}{\theta_{i,k} T_c} \right) + b_{i,k} \right] \tag{14a}$$

$$\text{s.t. } \mathbf{Y} \in \mathcal{G} \cap \mathcal{H}. \tag{14b}$$

Above, the normal set is

$$\mathcal{G} = \{\mathbf{Y} | 0 \leq Y_{(i-1)K+k} \leq V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k), \forall i \in \mathcal{I}, \forall k \in \mathcal{K}, \mathbf{P}_1 \in \mathcal{P}_1, \mathbf{P}_2 \in \mathcal{P}_2, \mathbf{B} \in \mathcal{B}\} \tag{15}$$

where

$$V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k) = \left(\mathbb{E}_{\gamma_k} \left\{ e^{-\theta_{i,k} B_k T_c \log\left(1 + \frac{P_{i,k}\gamma_k}{N_0 B_k + \mu_{3-i,k} P_{3-i,k}}\right)} \right\} \right)^{-1}. \tag{16}$$

Note that when $Y_{(i-1)K+k}$ in the objective function in (14) is replaced with the upper bound $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k)$, the objective function becomes the same as that in (11). In (14b), the conormal set is

$$\mathcal{H} = \{\mathbf{Y} | Y_{(i-1)K+k} \geq V_{(i-1)K+k}^{min}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}\} \tag{17}$$

where

$$V_{(i-1)K+k}^{min} = e^{\theta_{i,k} T_c e^{\frac{Q_{i,k}^{min} - b_{i,k}}{a_{i,k}}}}.$$

Note that the normal set $\mathcal{G}$ describes the combination of total bandwidth constraint (11b) and maximum transmission power constraint (11c), and the cornormal set $\mathcal{H}$ corresponds to the minimum quality constraint (11d).

Since $\Phi(\mathbf{Y})$ is an increasing function of $\mathbf{Y}$, the optimal solution of Problem (14), denoted by $\mathbf{Y}^*$, must be located at the upper boundary of $\mathcal{G}$, denoted by $\partial^+ \mathcal{G}$. This means that we can find a bandwidth allocation $\mathbf{B}^*$ and power allocations $\mathbf{P}_1^*$ and $\mathbf{P}_2^*$ corresponding to the optimal solution $\mathbf{Y}^*$ such that

$$Y_{(i-1)K+k}^* = \left(\mathbb{E}_{\gamma_k} \left\{ e^{-\theta_{i,k} B_k^* T_c \log\left(1 + \frac{P_{i,k}^* \gamma_k}{N_0 B_k^* + \mu_{3-i,k} P_{3-i,k}^*}\right)} \right\} \right)^{-1} \tag{18}$$

for all $i \in \mathcal{I}$ and $k \in \mathcal{K}$. Therefore, such $\mathbf{B}^*$, $\mathbf{P}_1^*$ and $\mathbf{P}_2^*$ are clearly the optimal solutions to Problem (11). Hence Problem (11) and (14) are equivalent. We must also note that $Y_{(i-1)K+k}$ is lower bounded by 1, i.e., $Y_{(i-1)K+k} \geq 1$ for all i and k. Consequently, the optimal solution $\mathbf{Y}^*$ to Problem (14), which is located only at the upper boundary of set $\mathcal{G}$, is also

[4]Note that the objective function in (11a) can be expressed as $\sum_{k=1}^{K} \sum_{i=1}^{2} \left(\omega_{i,k} Q_{i,k}(R_{i,k}) \right) = \sum_{k=1}^{K} \sum_{i=1}^{2} \omega_{i,k} \times \left[a_{i,k} \ln \left(\frac{1}{\theta_{i,k} T_c} \ln \left(\mathbb{E}_{\gamma_k} \left\{ e^{-\theta_{i,k} B_k T_c \log \left(1 + \frac{P_{i,k}\gamma_k}{N_0 B_k + \mu_{3-i,k} P_{3-i,k}} \right)} \right\} \right)^{-1} \right) + b_{i,k} \right]$ by incorporating (8), (9), and (10). In the ensuing discussion above, while reformulating the problem within the framework of monotonic optimization, we essentially replace $\left(\mathbb{E}_{\gamma_k} \left\{ e^{-\theta_{i,k} B_k T_c \log\left(1 + \frac{P_{i,k}\gamma_k}{N_0 B_k + \mu_{3-i,k} P_{3-i,k}}\right)} \right\} \right)^{-1}$ with auxiliary variables $Y_{(i-1)K+k}$ for $k = 1, \ldots, K$ and $i = 1, 2$.

lower bounded by $\mathbf{1}$. That means that the optimal solution $\mathbf{Y}^* \in \mathcal{G} \cap \mathcal{H} \cap \mathcal{L}$, where

$$\mathcal{L} = \{\mathbf{Y} | Y_{(i-1)K+k} \geq 1, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}\}.$$

B. Initialization of the Enclosing Polyblock

In order to better approximate the upper boundary of the feasible set, we need to initialize the polyblock that contains the feasible set properly. In other words, we need to find the smallest box $[\mathbf{0}, \mathbf{v}']$ that contains $\mathcal{G} \cap \mathcal{H} \cap \mathcal{L}$. Since both sets $\mathcal{H}$ and $\mathcal{L}$ are cornormal, the set

$$\begin{aligned}\mathcal{J} &= \mathcal{H} \cap \mathcal{L} \\ &= \{\mathbf{Y} | Y_{(i-1)K+k} \geq \max\{V^{min}_{(i-1)K+k}, 1\}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}\}\end{aligned}$$

is also cornormal. The smallest $\mathbf{v}'$ such that $[\mathbf{0}, \mathbf{v}']$ contains $\mathcal{G} \cap \mathcal{J}$ is given by the following:

$$v'_j = \max\{Y_j | \mathbf{Y} \in \mathcal{G} \cap \mathcal{J}\} \quad \forall j = 1, \ldots, 2K. \tag{19}$$

Before describing the enclosing polyblock initialization algorithm, we provide the following characterization for the functional properties of $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k)$.

Theorem 1: Consider the functions

$$V_1(P_1, P_2, B) = \left(\mathbb{E}_\gamma \left\{ e^{-\theta B T_c \log\left(1+\frac{P_1 \gamma}{N_0 \ B+\mu P_2}\right)} \right\}\right)^{-1} \text{and} \tag{20}$$

$$V_2(P_1, P_2, B) = \left(\mathbb{E}_\gamma \left\{ e^{-\theta B T_c \log\left(1+\frac{P_2 \gamma}{N_0 B+\mu P_1}\right)} \right\}\right)^{-1} \tag{21}$$

and assume that $P_1 \leq P^{\max}$ and $P_2 \leq P^{\max}$. Then, we have the following properties:

1) For given bandwidth B, V_1 is maximized when either $P_1 = P^{\max}$ or $P_2 = P^{\max}$. Hence, at least one power value should be at its maximum level.
2) For given P_1 and P_2, V_1 is an increasing function of B.
3) The above properties hold for V_2 as well due to the similarity in their definitions (with only roles of P_1 and P_2 switched).
4) The bandwidth required to achieve two target values $V_1(P_1, P_2, B) = V_1^*$ and $V_2(P_1, P_2, B) = V_2^*$ is minimized if either $P_1 = P^{\max}$ or $P_2 = P^{\max}$.

Proof: See Appendix VI-A.

The detailed algorithm for initializing the enclosing polyblock is provided below in Algorithm 1. We note that Step 3 of Algorithm 1 makes use of Theorem 1, i.e., the fact that the minimum bandwidth always occurs at $P_{1,k} = P^{max}_{1,k}$ or $P_{2,k} = P^{max}_{2,k}$. In particular, in this step, we identify the minimum bandwidth needed by the users while power and minimum quality constraints are satisfied. Subsequently, starting with Step 8, we allocate the remaining bandwidth to user k (after providing the minimum required bandwidth to all users) and determine $V^{max}_{(i-1)K+k}$, which is essentially the solution of the maximization problem in (19) for $j = (i-1)K + k$ within the feasible set $\mathcal{G} \cap \mathcal{J}$.

We now provide an illustration for the enclosing polyblock initialization. For instance, assume that $\mathcal{G}$ and $\mathcal{J}$ are two-dimensional sets by assuming $K = 1$. As shown in Fig. 2, the box $[\mathbf{0}, \mathbf{v}']$ constrained by the red lines is the smallest box that contains $\mathcal{G} \cap \mathcal{J}$, where $\mathbf{v}' = (v'_1, v'_2)$. And $\mathbf{v}'$ can be obtained by the algorithm provided above.

Algorithm 1 The Enclosing Polyblock Initialization Algorithm

Input: $\mathcal{G}$, $\mathcal{H}$ and $\mathcal{L}$

Output: Polyblock $\mathcal{S}_1$

1: Initialize $s = 1$.
2: **for** $k = 1 : K$ **do**
3: Set $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k) = \max\{V^{min}_{(i-1)K+k}, 1\}$ for $i = 1, 2$. Let $P_{1,k} = P^{max}_{1,k}$, find the bandwidth $B_k = B_{k1}$ and power $P_{2,k}$ by solving (16). Similarly, let $P_{2,k} = P^{max}_{2,k}$, find the bandwidth $B_k = B_{k2}$ and $P_{1,k}$ by solving (16). $B^{min}_k = \min\{B_{k1}, B_{k2}\}$ if both $P_{1,k} \in \mathcal{P}_1$ and $P_{2,k} \in \mathcal{P}_2$, and $B^{min}_k = B_{ki}$ if just one $P_{i,k} \in \mathcal{P}_i$ for $i = 1$ or $i = 2$. Otherwise, Problem (14) does not have solution and set $s = 0$.
4: **end for**
5: If $s = 1$, and $\sum_{l=1}^{K} B^{min}_l > B$, the Problem (14) does not have solution and set $s = 0$.
6: **if** $s = 1$ **then**
7: **for** $k = 1 : K$ **do**
8: $B_k = B - \sum_{l \neq k} B^{min}_l$.
9: **for** i = 1:2 **do**
10: Let $V_{(2-i)K+k} = \max\{V^{min}_{(2-i)K+k}, 1\}$ and $P_{3-i,k} = P^{max}_{3-i,k}$, and find the power $P_{i,k}$ from (16).
11: Calculate $V^{max}_{(i-1)K+k}$ from (16) by substituting $P_{j,k}$ and $P_{i,k}$ obtained above.
12: **end for**
13: **end for**
14: **end if**
15: Therefore, the vector $\mathbf{v}' = (V^{max}_1, \ldots, V^{max}_{2K})$ is the vertex of the initial polyblock $\mathcal{S}_1$.

Before we solve the optimization problem by using MO theory, we provide the following proposition from [19].

Proposition 1 (Projection on the Upper Boundary [17]): Let $\mathcal{G} \subset \mathcal{R}^n_+$ be a compact normal set with nonempty interior. Then, for any point $\mathbf{x} \in \mathcal{R}^n_+ \setminus \mathcal{G}$, the line connecting $\mathbf{0}$ and $\mathbf{x}$ intersects the upper boundary $\partial^+\mathcal{G}$ of $\mathcal{G}$ at a unique point $\pi_{\mathcal{G}}(\mathbf{x})$, which is defined as

$$\pi_{\mathcal{G}}(\mathbf{x}) = \lambda \mathbf{x}, \ \text{ where } \lambda = \arg\max\{\alpha > 0 \mid \alpha \mathbf{x} \in \mathcal{G}\}. \tag{22}$$

$\pi_{\mathcal{G}}(\mathbf{x})$ is the projection of $\mathbf{x}$ on the upper boundary $\partial^+\mathcal{G}$.

Due to the presence of $\mathcal{J}$, $\pi_{\mathcal{G}(\mathbf{x})}$ may be located outside the feasible set $\mathcal{G} \cap \mathcal{J}$ if one end point of the line is $\mathbf{0}$. In order to avoid this situation, we modify the projection by changing the line connecting $\mathbf{0}$ and $\mathbf{x}$ to the line connecting $\mathbf{u}$ and $\mathbf{x}$, and we denote by $\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{x})$ the projection of $\mathbf{x}$ on the upper boundary $\partial^+\mathcal{G}$ with $\mathbf{u}$ acting as the origin. Therefore,

$$\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{x}) = \lambda(\mathbf{x} - \mathbf{u}) + \mathbf{u}, \tag{23}$$

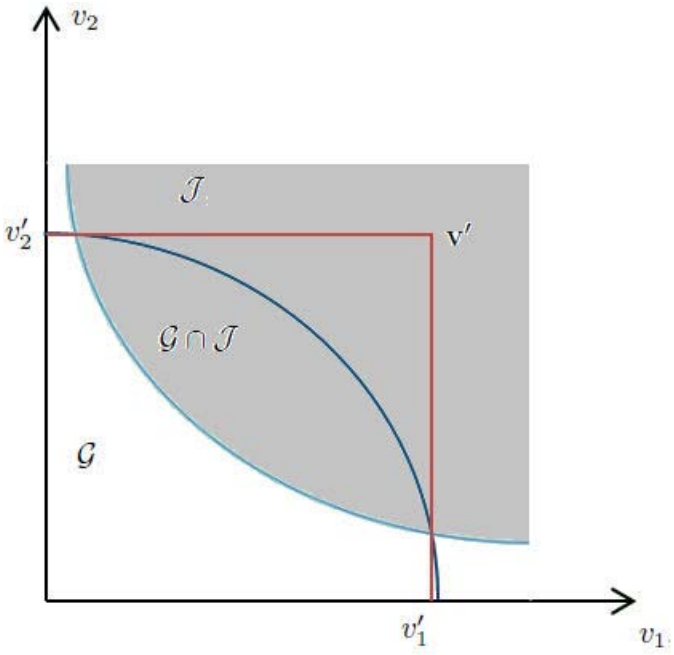


Fig. 2. Example of initialized enclosing polyblock

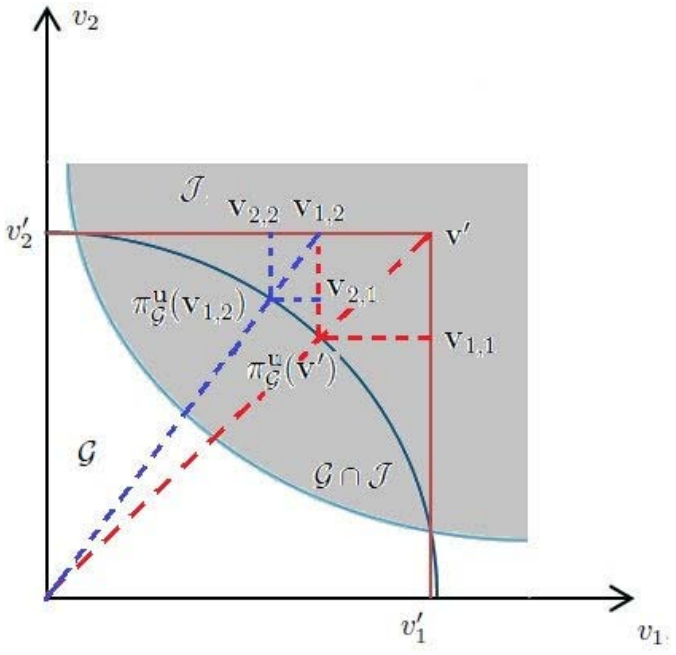


Fig. 3. Iterative procedure of projection onto the boundary and formation of new set of vertices.

where $\lambda = \arg\max\{\alpha > 0 \mid \alpha\mathbf{x} \in \mathcal{G}\}$ and $\mathbf{u} = (\max\{V_1^{min}, 1\}, \ldots, \max\{V_{2K}^{min}, 1\})$.

C. Algorithms and Optimal Solution via Monotonic Optimization

After obtaining the proper initial polyblock, we next develop algorithms and determine the optimal solution to Problem (14) via MO approach. The key idea of MO is to iteratively derive a new enclosing polyblock $\mathcal{S}_{j+1}$ from the previous polyblock $\mathcal{S}_j$ by cutting off the points that is in the infeasible set until reaching the ϵ-error-tolerance solution. Following [19, Proposition 3.8], we let $\mathcal{S} \subset \mathcal{R}_+^n$ be a polyblock with a proper vertex set $\mathcal{T} \subset \mathcal{R}_+^n$ and let $\mathbf{x} \in \mathcal{S}$. Then, the new polyblock $\mathcal{S}_*$ has a vertex set

$$\mathcal{T}' = (\mathcal{T} \setminus \mathcal{T}_*) \cup \{\mathbf{v} = \mathbf{v} + (x_j - v_j)e_j | \mathbf{v} \in \mathcal{T}_*, j \in \{1, \ldots, n\}\} \tag{24}$$

where $\mathcal{T}_*$ is the subset of $\mathcal{T}$, consisting of the vertices at which $\Phi(\mathbf{Y})$ is maximized. It is easy to see that if $\mathcal{S}$ is the proper polyblock such that $\mathcal{G} \cap \mathcal{J} \subset \mathcal{S}$ and $\mathbf{x} \in \partial^+\mathcal{G}$, then we have $\mathcal{G} \cap \mathcal{J} \subset \mathcal{S}_* \subset \mathcal{S}$.

We first construct a proper polyblock $\mathcal{S}_1$ that contains the feasible set, $\mathcal{G} \cap \mathcal{J}$ of Problem (14) by using Algorithm 1, and let $\mathcal{T}_1$ denote the initial proper vertex set of $\mathcal{S}_1$. There is just one vertex, $\mathbf{v}'$, in $\mathcal{T}_1$. Since the objective function of Problem (14), $\Phi(\mathbf{Y})$ is monotonically increasing over set $\mathcal{S}_1$, the maximum of $\Phi(\mathbf{Y})$ occurs at some proper vertex $\mathbf{Y}_1$ of $\mathcal{S}_1$, i.e., $\mathbf{Y}_1 \in \mathcal{T}_1$. If $\mathbf{Y}_1$ is also in the feasible set $\mathcal{G} \cap \mathcal{J}$, then the optimization problem is solved and $\mathbf{Y}^* = \mathbf{Y}_1$. Otherwise, a smaller polyblock $\mathcal{S}_2 \subset \mathcal{S}_1$ is constructed such that $\mathcal{G} \cap \mathcal{J} \subset \mathcal{S}_2$ but excludes $\mathbf{Y}_1$ by using [19, Proposition 3.8]. Therefore, a new vertex set $\mathcal{T}_2$ is constructed by replacing $\mathbf{Y}_1$ in $\mathcal{T}_1$ with $2 \times K$ new vertices and removing the improper vertices. This procedure is repeated until an ϵ-error-tolerance solution is found. If $\mathbf{Y}_j$ denotes the optimal vertex that maximizes $\Phi(\mathbf{Y})$ over set $\mathcal{S}_j$ at the j-th iteration, we have $\mathcal{S}_1 \supset \mathcal{S}_2 \supset \cdots \supset \mathcal{G}$ and $\Phi(\mathbf{Y}_1) \geq \Phi(\mathbf{Y}_2) \geq \cdots \geq \Phi(\mathbf{Y}^*)$. $\mathbf{Y}_j' = \arg\max\{\Phi(\mathbf{Y}) | \mathbf{Y} \in \{\pi_{\mathcal{G}}^{\mathbf{u}}(\mathbf{Y}_j), \mathbf{Y}_{j-1}'\}\}$ denotes the current best solution (CBS), and the current best value (CBV) is $\Phi(\mathbf{Y}_j')$ in the j-th iteration. Consequently, we have $\Phi(\mathbf{Y}_1') \leq \Phi(\mathbf{Y}_2') \leq \cdots \leq \Phi(\mathbf{Y}^*)$. The algorithm terminates at the j-th iteration if $\mathbf{Y}_j \in \mathcal{S}_j$, and $(1+\epsilon)\Phi(\mathbf{Y}_j') \geq \Phi(\mathbf{Y}_j)$ or $|\Phi(\mathbf{Y}_j') - \Phi(\mathbf{Y}_j)| \leq \epsilon$ based on the chosen strategy, where $\epsilon > 0$ is a small positive number representing the error tolerance. $\mathbf{Y}_j'$ is the optimal ϵ-error-tolerance solution.

In order to illustrate the iterative process described above, we again consider the simple setting of Fig. 2 with $K = 1$ and plot Fig. 3 above. As seen in the figure, in the first iteration, the vertex $\mathbf{v}'$ is projected onto the boundary at point $\pi_{\mathcal{G}}^{\mathbf{u}}(\mathbf{v}')$ by using the projection algorithm. Following this, we generate the polyblock with vertices $\mathbf{v}_{1,1}$ and $\mathbf{v}_{1,2}$ and form the set of vertices $\mathcal{T}$. Any point inside this polyblock is possible to be the optimal one. Let us assume that $\mathbf{v}_{1,2}$ leads to a higher value of the objective function than $\mathbf{v}_{1,1}$. Then, in the second iteration, we project $\mathbf{v}_{1,1}$ onto the boundary and get the projection point $\pi_{\mathcal{G}}^{\mathbf{u}}(\mathbf{v}_{1,2})$. Following this, two new vertices $\mathbf{v}_{2,1}$ and $\mathbf{v}_{2,2}$ are added to $\mathcal{T}$ while the previous vertex point $\mathbf{v}_{1,2}$ is removed from $\mathcal{T}$. Therefore, the updated set of vertices now contains $\mathbf{v}_{1,1}$, $\mathbf{v}_{2,1}$ and $\mathbf{v}_{2,2}$, and these three vertices form the new polyblock. And, we continue this procedure iteratively until an $\epsilon-$error-tolerance solution is obtained.

As discussed in the previous subsection with Proposition 1, iterations in finding $\{\mathbf{Y}_j\}$ involve projection on the upper boundary. We provide our projection algorithm for finding $\pi_{\mathcal{G}}^{\mathbf{u}}(\mathbf{Y}_j)$ as Algorithm 2 above. In steps 6 and 8 of this algorithm, the reason for considering $P_{1,k}$ or $P_{2,k}$ to be at

Algorithm 2 Projection Algorithm (for Finding $\pi_{\mathcal{G}}(\mathbf{Y}_j)$)

Input: $\mathbf{Y}_j$, $\mathcal{G}$

Output: λ_j such that $\lambda_j = \arg\max\{\lambda_j > 0 | \lambda_j \mathbf{Y}_j \in \mathcal{G}\}$

Initialize $\lambda_j = 0$
for $d = 0 : 2^K - 1$ **do**
  Let c be a K-digit binary integer corresponding to d, and c_l denote the l-th binary digit of c.
  for $k = 1 : K$ **do**
    if $c_k = 0$ **then**
      $P_{1,k} = P_{1,k}^{max}$
    else
      $P_{2,k} = P_{2,k}^{max}$
    end if
    From (16), we set $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k) = \lambda_{j,d+1}(Y^j_{(i-1)K+k} - u_{(i-1)K+k}) + u_{(i-1)K+k}$.
  end for
  Set $\sum_{k=1}^{K} B_k = B$.
  Therefore, we get $2K+1$ equations, K unknown power variables $P_{1,k}$ or $P_{2,k}$, K unknown bandwidth variables B_k for all $k = 1, \ldots, K$, and unknown variable $\lambda_{j,d+1}$. We can get the value of $\lambda_{j,d+1}$ by solving this $2K+1$ equations. If $P_{i,k} \leq P_{i,k}^{max}$ for all $i = 1, 2$ and $k = 1, \ldots, K$, $\lambda_j = \max\{\lambda_j, \lambda\}$.
end for
$\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j) = \lambda_j(\mathbf{Y}_j - \mathbf{u}) + \mathbf{u}$.

the maximum level for all $k = 1, \ldots, K$ and $\sum_{k=1}^{K} B_k = B$ is that $\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j)$ is attained at the upper boundary of $\mathcal{G}$, and the upper boundary $\partial^+\mathcal{G}$ is reached only if one of the users transmits at the peak power level. The proof for this characterization is provided in Appendix VI-B, which primarily follows from the results of Theorem 1.

After having obtained the initial enclosing polyblock $\mathcal{S}_1$ and identified the algorithm for projection on the boundary, we can now iteratively derive a new enclosing polyblock $\mathcal{S}_{j+1}$ from the previous polyblock $\mathcal{S}_j$ by using Algorithm 3 below. Eventually, we obtain the ϵ-error-tolerance solution after terminating the iteration under a certain condition.

Via Algorithms 1–3, we determine the optimal bandwidth allocation and power allocation (BAPA) maximizing weighted sum quality of the videos of the users under total bandwidth, individual power, and individual video quality constraints (i.e., we solve the optimization problem in (11)).

In the numerical results presented in the next section, we demonstrate the optimal performance and identify the key tradeoffs. Additionally, we analyze the equal-bandwidth (EB) scenario in which bandwidth is equally allocated to the users, i.e., $B_k = \frac{B}{K}$, and power allocation is performed separately for each pair of full-duplex users, and provide comparisons.

V. Numerical and Simulation Results

Five CIF video sequences namely *Akiyo, Bus, Coastguard, Foreman* and *News* are used for the simulation results [22]. Size of each frame is 352×288 pixels. FFMPEG is used for encoding the video sequences and GOP is set as 10. Frame rate

Algorithm 3 The Optimal Resource Allocation Algorithm

Input: Function $\Phi(\mathbf{Y}) : \mathcal{R}_+^{2\times K} \to \mathcal{R}$, compact normal set $\mathcal{G} \subset \mathcal{R}_+^{2\times K}$, and a closed conormal set $\mathcal{J} \subset \mathcal{R}_+^{2\times K}$ such that $\mathcal{G} \cap \mathcal{J} \neq \emptyset$

Output: An ϵ error tolerance solution $\mathbf{Y}^*$ and the corresponding $\mathbf{P}_1^*$, $\mathbf{P}_2^*$ and $\mathbf{B}^*$.

Initialization: Let the initial polyblock $\mathcal{S}_1$ be the box $[\mathbf{0}, \mathbf{b}]$ that encloses $\mathcal{G} \cap \mathcal{J}$ (This can be obtained by using Algorithm 1). The vertex set $\mathcal{T}_1 = \mathbf{b}$. $\epsilon > 0$ is a small positive number. CBV $\Omega_0 = 0$ and $j = 0$.
repeat
  $j = j + 1$.
  Select $\mathbf{Y}_j \in \arg\max\{\Phi(\mathbf{Y}) | \mathbf{Y} \in \mathcal{T}_j\}$.
  Compute $\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j)$ by projecting $\mathbf{Y}_j$ on the upper boundary of $\mathcal{G}$ (Algorithm 2).
  if $\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j) = \mathbf{Y}_j$, i.e., $\mathbf{Y}_j \in \partial^+\mathcal{G}$ **then**
    CBS $\mathbf{Y}' = \mathbf{Y}_j$ and CBV $\Omega_j = \Phi(\mathbf{Y}_j)$.
  else
    if $\Phi(\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j)) \geq \Omega_{j-1}$ **then**
      $\mathbf{Y}'_j = \pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j)$ and $\Omega_j = \Phi(\pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j))$.
    else
      $\mathbf{Y}'_j = \mathbf{Y}'_{j-1}$ and $\Omega_j = \Omega_{j-1}$.
    end if
    Let $\mathbf{x} = \pi^{\mathbf{u}}_{\mathcal{G}}(\mathbf{Y}_j)$ and $\mathcal{T}_{j+1} = (\mathcal{T}_j \setminus \mathcal{T}_*) \cup \{\mathbf{v} = \mathbf{v} + (x_t - v_t)e_t | \mathbf{v} \in \mathcal{T}_*, t \in \{1, \ldots, 2K\}\}$, where $\mathcal{T}_* = \{\mathbf{v} \in \mathcal{T}_j | \mathbf{v} > \mathbf{x}\}$.
    Remove the improper vertices from $\mathcal{T}_{j+1}$.
  end if
until $|\Phi(\mathbf{Y}_j) - \Omega_j| \leq \epsilon$.
$\mathbf{Y}^* = \mathbf{Y}'_j$ is the optimal solution and corresponding $\mathbf{P}_1^*$, $\mathbf{P}_2^*$ and $\mathbf{B}^*$ is the optimal resource allocation.

TABLE I
Parameter Values of the Quality Rate Model for Different Video Sequences

	Akiyo	Bus	Coastguard	Foreman	News
a_k	5.0545	4.7205	3.5261	4.5006	5.6218
b_k	17.1145	5.4764	13.8425	13.0780	10.0016

is set as 15 frames per second. Table I shows the parameters a_k and b_k that make the rate-distortion function of the five video sequences fit the quality rate model in (7), where the unit of R_k is kbit/s. Unless mentioned explicitly, we assume that the subchannel power gain for each link is exponentially distributed with mean $Z_k = \mathbb{E}\{\gamma_k\}$. The power spectrum density of the AWGN is set to $N_0 = 10^{-6}$ W/Hz, and the channel coherence time is assumed to be 0.001 seconds. The self-interference factor at each user is set to 0.1.

Fig. 4 shows the actual PSNR values as a function of the source bit rate for different video sequences, where we see that the increasing concave quality rate model fits the actual values very well. Throughout the numerical results, we assume the minimum required video quality is $Q_{i,k}^{min} = 20$dB and maximum transmission power is $P_{i,k}^{max} = 5$ for all users.

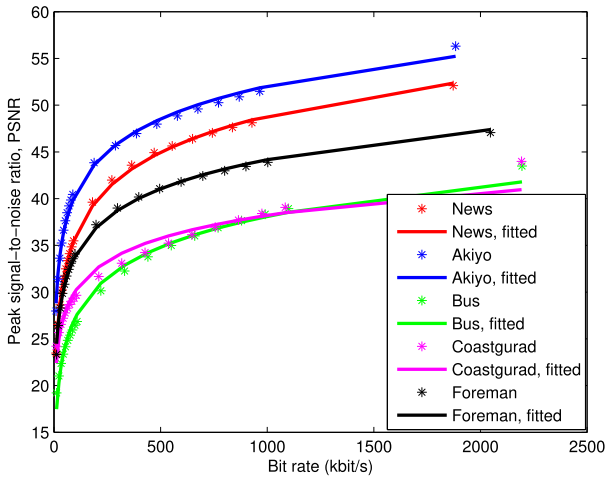
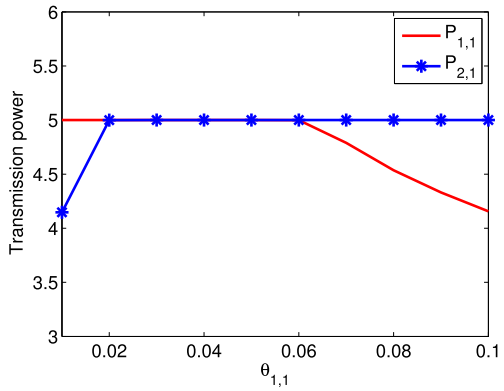
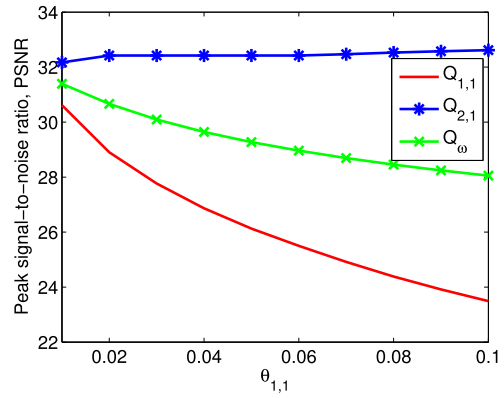


Fig. 4. Actual PSNR values vs. rate and fitted quality rate curves.



(a)



(b)

Fig. 5. (a) Optimal power allocation and (b) the corresponding quality Q (or equivalently PSNR) of video sequences as a function of $\theta_{1,1}$.

A. One Pair of Full-Duplex Users

In this section, we consider the power allocation between a single pair of full-duplex users. The bandwidth B is set to 0.1 MHz, average channel power gain is $Z_1 = 1$. $U_{1,1}$ transmits video sequence *Bus* to $U_{2,1}$, while $U_{2,1}$ transmits video sequence *Coastguard* to $U_{1,1}$, with the corresponding parameters ($a_{1,1} = 4.7205$, $b_{1,1} = 5.4764$), and ($a_{2,1} = 3.5261$, $b_{2,1} = 13.8425$) from Table I.

1) The Impact of the QoS Exponent on Multimedia Quality: In Fig. 5, we set $\omega_{1,1} = \omega_{2,1} = 0.5$ (meaning that two videos are equally weighted), and increase the value of $\theta_{1,1}$ (the QoS exponent of user $U_{1,1}$) from 0.01 to 0.1 while keeping $\theta_{2,1} = 0.01$. Note that increased $\theta_{1,1}$ implies that more stringent delay constraints are imposed on the video transmission of $U_{1,1}$. Fig. 5a plots the power allocated to the users as $\theta_{1,1}$ increases. Since quality parameter $a_{1,1}$ of *Bus* video is greater than $a_{2,1}$ of *Coastguard* video, quality $Q_{1,1}$ of the *Bus* video increases faster than $Q_{2,1}$ of the *Coastguard* video as the transmission power and correspondingly the arrival rate R grow, according to the logarithmic model in (7). Therefore, initially when $\theta_{1,1} = \theta_{2,1} = 0.01$ and $U_{1,1}$ and $U_{2,1}$ are subject to the same delay constraint, $U_{1,1}$ transmits at the peak power level in a greedy fashion to maximize the sum video quality, while $U_{2,1}$ uses less power.

As $\theta_{1,1}$ increases, more stringent delay constraints are imposed on user $U_{1,1}$ and the arrival rate $R_{1,1}$ of the *Bus* video is reduced to avoid delay violations. Consequently, the video quality $Q_{1,1}$ (or equivalently the PSNR of the video) starts diminishing as seen in Fig. 5b. Eventually, when $\theta_{1,1}$ exceeds 0.06, the lower arrival rates can be supported by smaller transmission power and $P_{1,1}$ is reduced as observed in Fig. 5a. In the meantime, we notice that quality $Q_{2,1}$ of *Costguard* video slightly increases due to increased transmission power $P_{2,1}$ at $U_{2,1}$ and smaller self-interference at $U_{1,1}$ (because of smaller transmission power $P_{1,1}$). However, since the drop in $Q_{1,1}$ is more significant, the weighted sum quality Q_{ω} is seen to decrease in Fig. 5b. Finally, it is interesting to note that, as predicted by Theorem 1 and discussed subsequently, at least power value is at the maximum level of 5, i.e., $P_{1,1} = 5$ or $P_{2,1} = 5$, for any given value of $\theta_{1,1}$ in Fig. 5a.

In Fig. 6, both $\theta_{1,1}$ and $\theta_{2,1}$ increase from 0.01 to 0.1 together. Since $U_{1,1}$ and $U_{2,1}$ now all the time operate under the same QoS constraints while transmitting different video sequences, Fig. 6a demonstrates that $P_{1,1}$ is always greater than $P_{2,1}$ due to, as discussed above, the impact of video quality parameters, or more specifically due to having $a_{1,1} > a_{2,1}$. Fig. 6b shows that both $Q_{1,1}$ and $Q_{2,1}$ decrease as both $\theta_{1,1}$ and $\theta_{2,1}$ increase. That is because larger $\theta_{1,1}$ and $\theta_{2,1}$ lead to smaller source rates $R_{1,1}$ and $R_{2,1}$, which in turn reduce the video quality.

2) The Impact of Weights on Multimedia Quality: Now, we set $\theta_{1,1} = \theta_{2,1} = 0.01$, and increase the weight $\omega_{1,1}$ from 0 to 1 while keeping $\omega_{1,1} + \omega_{2,1} = 1$. Hence, the weight of user $U_{1,1}$ gradually increases in the weighted sum quality maximization in (11). Fig. 7a shows that, as expected, $P_{1,1}$ grows and reaches the peak value as $\omega_{1,1}$ increases due to higher emphasis on the quality $Q_{1,1}$. At the same time, $P_{2,1}$ starts diminishing when $\omega_{1,1}$ increases beyond 0.4 and hence $\omega_{2,1}$ drops below 0.6. Fig. 7b plots the corresponding qualities of the video sequences. Following similar trends as in the power curves, $Q_{1,1}$ improves whereas $Q_{2,1}$ is reduced. Finally, we note that we have $Q_{1,1} = 20$dB when $\omega_{1,1} = 0$, and $Q_{2,1} = 20$dB when $\omega_{1,1} = 1$ due to the fact that a minimum quality of 20dB is imposed on both video transmissions.

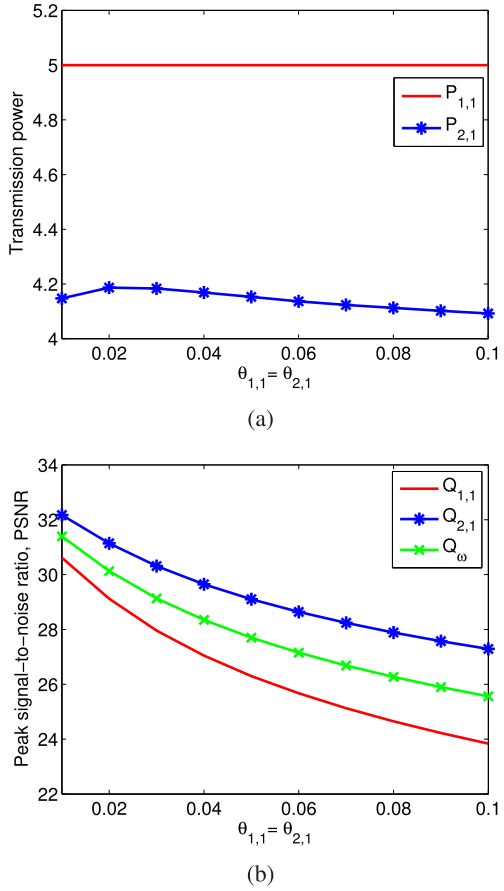


Fig. 6. (a) Optimal power allocation and (b) the corresponding quality Q (or equivalently PSNR) of video sequences as a function of $\theta_{1,1} = \theta_{2,1}$.

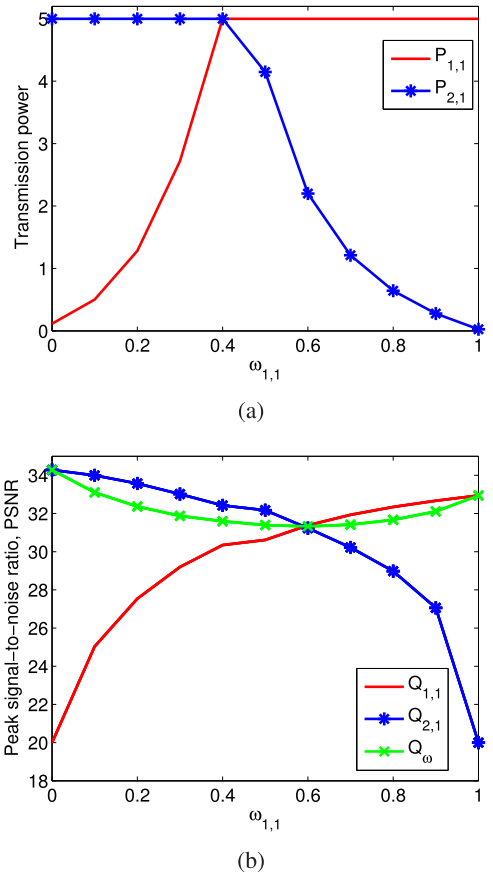


Fig. 7. (a) Optimal power allocation and (b) the corresponding quality Q (or equivalently PSNR) of video sequences as a function of $\omega_{1,1}$.

B. Two Pairs of Full-Duplex Users

In this section, we consider bandwidth and power allocation for two pairs of full-duplex users. The total bandwidth B is set to 0.2 MHz, and the average channel power gains are $Z_1 = 1$ between first pair of users and $Z_2 = 3$ between the second pair of users. $U_{1,1}$ and $U_{1,2}$ transmit the same video sequence *Bus* to $U_{2,1}$ and $U_{2,2}$, respectively. And video sequence *Coastguard* is transmitted to $U_{1,1}$ and $U_{1,2}$ by $U_{2,1}$ and $U_{2,2}$ respectively. For these video sequences, we have $a_{1,1} = a_{1,2} = 4.7205$ and $b_{1,1} = b_{1,2} = 5.4764$, $a_{2,1} = a_{2,2} = 3.5261$ and $b_{2,1} = b_{2,2} = 13.8425$.

1) The Impact of the QoS Exponent on Multimedia Quality: In this subsection, we initially set $\omega_{1,1} = \omega_{2,1} = \omega_{1,2} = \omega_{2,2} = 0.25$, and increase the values of the QoS exponents of the first pair of users $\theta_{1,1}$ and $\theta_{2,1}$ from 0.01 to 0.1 together (i.e., $\theta_{1,1} = \theta_{2,1}$) while keeping the QoS exponents of the second pair of users at $\theta_{1,2} = \theta_{2,2} = 0.01$. Fig. 8a and Fig. 8b show the results of the optimal power and bandwidth allocation as a function of $\theta_{1,1} = \theta_{2,1}$. Note that as QoS exponents $\theta_{1,1} = \theta_{2,1}$ increase (hence more stringent QoS constraints are imposed), lower arrival rates are supported and the quality of the video sequences of the first pair of users degrades. With this, bandwidth allocated to the first pair of users is reduced as noticed in Fig. 8b. Due to similar reasons (regarding the video quality parameters) as discussed in the case of one pair of full-duplex users (i.e., $a_{1,k} > a_{2,k}$ for $k = 1, 2$), $P_{1,1}$ and $P_{1,2}$ are always at their maximum levels. We also observe that as $\theta_{1,1} = \theta_{2,1}$ increase, $P_{2,1}$ diminishes whereas $P_{2,2}$ grows. These are due to the facts that the bandwidth allocated to the link between $U_{1,1}$ and $U_{2,1}$ decreases while the bandwidth allocated to the link between $U_{1,2}$ and $U_{2,2}$ increases. Hence, an opportunistic strategy is employed and more power is allocated to the link with more bandwidth. Fig. 8c demonstrates that the average PSNR value of first pair of video sequences degrades due to increasing QoS exponents and smaller bandwidth.

Fig. 9 plots the weighted sum quality of video sequences assuming optimal and also equal bandwidth allocation. In both cases, power is optimally allocated. We note that the equal bandwidth optimal power (EBOP) allocation scheme provides a performance close to that of the optimal bandwidth and power allocation scheme, but the gap widens as $\theta_{1,1} = \theta_{2,1}$ increase.

Hence, the performance improvements are highly dependent on the parameter values. For instance, in Fig. 10, we consider smaller values of the QoS exponent, which imply looser QoS requirements. Specifically, we vary $\theta_{1,1}$ and $\theta_{2,1}$ from 0.001 to 0.01 together while keeping the QoS exponents of the second pair of users at $\theta_{1,2} = \theta_{2,2} = 0.001$. We also set $\omega_{1,1} = \omega_{2,1} = 0.45$ and $\omega_{1,2} = \omega_{2,2} = 0.05$. In Fig. 10a, we plot the PSNR for both optimal bandwidth and power allocation, EBOP

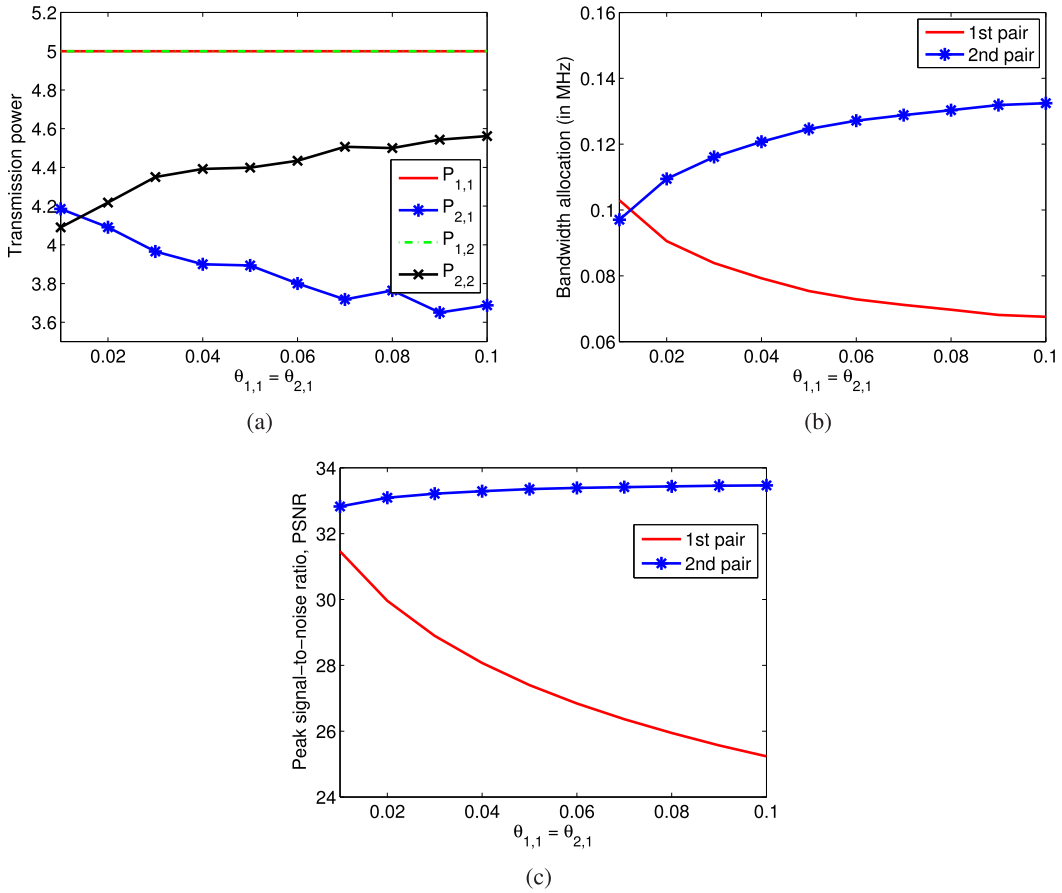


Fig. 8. (a) Optimal power allocation, (b) optimal bandwidth allocation, and (c) the corresponding quality of video sequences as $\theta_{1,1} = \theta_{2,1}$ increase.

allocation, and equal bandwidth and maximum power (EBMP) allocation. We observe the performance gains of optimal allocation while EBOP and EBMP curves are almost overlapping. More interestingly, as seen in Fig. 10b where average power consumption is plotted, the performance improvements with optimal allocation is attained while consuming less average power. Hence optimal allocation improves power efficiency as well. Expectedly, EBMP is the worst strategy, consuming the highest levels of power without any improvements in PSNR.

Next, we address in more detail the impact of having different weights in the weighted sum video quality maximization.

2) The Impact of Weights on Multimedia Quality: Fig. 11 shows the optimal bandwidth and power allocation and the corresponding quality of video sequences as the weights $\omega_{1,1} = \omega_{2,1}$ vary from 0.05 to 0.45. We also assume that $\omega_{1,2} = \omega_{2,2}$ while keeping the sum of all weights equal to 1. Fig. 11b indicates that bandwidth B_1 allocated to the first pair of users increases with increasing $\omega_{1,1} = \omega_{2,1}$ since growing emphasis is given to the quality of the video sequences transmitted between first pair of users. Consequently, the bandwidth allocated to the link between second pair of users $U_{1,2}$ and $U_{2,2}$ decreases. Since $\omega_{1,1} = \omega_{2,1}$ and $a_{1,1} > a_{2,1}$, $U_{1,1}$ always transmits the video sequence at the maximum transmission power level. Due to the same reason, $P_{1,2}$ always attains the maximum level. Again, due to the optimality of the opportunistic approach, $P_{2,1}$ increases as B_1 gets larger, whereas $P_{2,2}$ diminishes as B_2 becomes smaller. Correspondingly, Fig. 8c demonstrates that the average PSNR values $Q_{1,1}$ and $Q_{2,1}$ improve as higher weights $\omega_{1,1} = \omega_{2,1}$ are given to the video communication between the first pair of users, while the average PSNR values $Q_{1,2}$ and $Q_{2,2}$ are lowered.

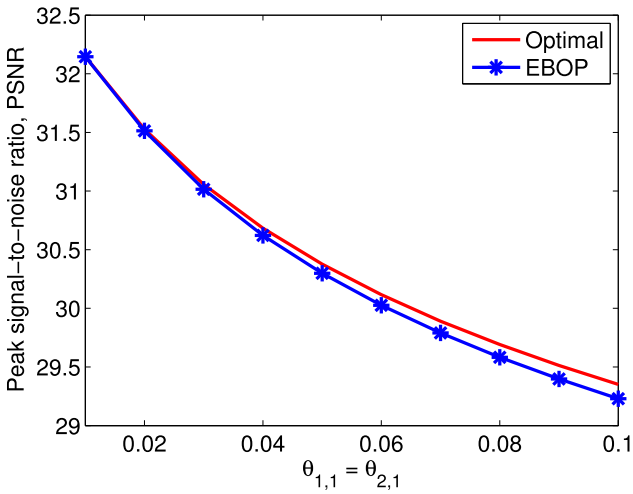


Fig. 9. Quality of video sequences as a function of $\theta_{1,1} = \theta_{2,1}$. Both optimal and equal bandwidth allocation are considered.

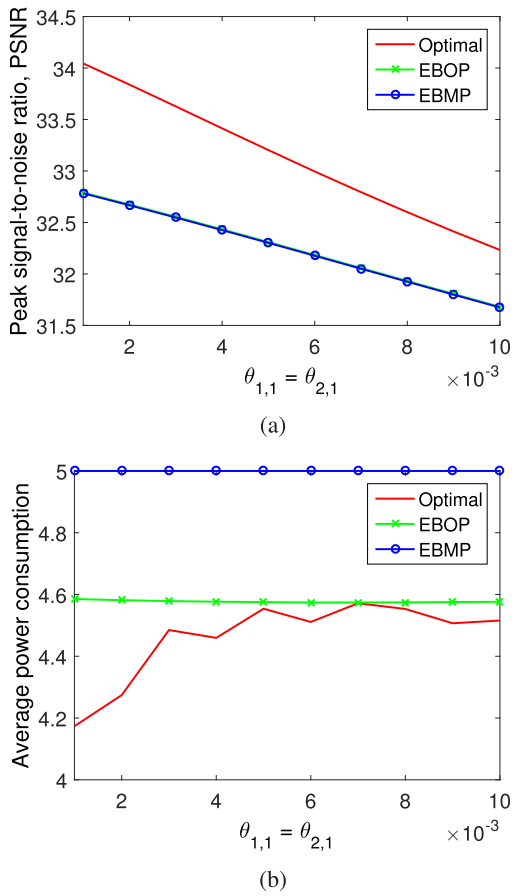


Fig. 10. (a) Quality of video sequences as a function of $\theta_{1,1} = \theta_{2,1}$ and (b) average power consumption as a function of $\theta_{1,1} = \theta_{2,1}$.

Fig. 12 shows the weighted sum quality of video sequences again considering optimal and equal bandwidth allocation schemes. As expected, the optimal bandwidth and power allocation scheme outperforms the case in which bandwidth is equally allocated among the pairs of users and power is allocated optimally (i.e., EBOP scheme). The performance gap is smallest when the weights are all equal (i.e., $\omega_{1,1} = \omega_{2,1} = \omega_{1,2} = \omega_{2,2} = 0.25$), and the gap grows as the difference in the weights increases.

3) Performance Comparison of Full-Duplex and Half-Duplex Operations: In Fig. 13, we compare the performances of half-duplex and full-duplex operations. In particular, we plot the PSNR (or equivalently the quality) of the video sequences as a function of the QoS exponents $\theta_{1,1} = \theta_{2,1}$ for four different cases (i.e., three full-duplex scenarios and one half-duplex). Since full-duplex performance is interference dependent, self interference suppression factor, μ, plays an important role in these cases. In the figure, for the three curves corresponding to full-duplex operation, we set $\mu = 0.1$, $\mu = 0.05$, and $\mu = 0.01$ respectively. $\mu = 0.01$ reflects the highest level of self-interference cancelation. We note that in the half-duplex case, the communicating pair of users employ time-division multiplexing, i.e., each user in the pair transmits half of the time and receives in the remaining half with no interference. We further note that while the users always transmit at the peak power level in the half-duplex case, bandwidth is still optimally allocated among different pairs of users by solving a convex optimization problem.

In the figure, we notice that half-duplex operation outperforms full-duplex case when $\mu = 0.1$. On the other hand, when self-interference is suppressed further i.e., when we have $\mu = 0.05$, full-duplex operation performs better as long as θ values are less than approximately 0.05. As the QoS exponent grows further, half-duplex starts leading to slightly higher PSNR levels. Hence, we interestingly observe that under stringent buffer/delay constraints, we need more self-interference suppression in order to surpass the performance levels achieved with half-duplex communication. Indeed, when we have $\mu = 0.01$, full-duplex operation outperforms half-duplex scheme over all values of the QoS exponents shown in the figure.

TABLE II
PERFORMANCE WITH 3 PAIRS OF FULL-DUPLEX USERS

k	$P_{1,k}$	$P_{2,k}$	B_k	$Q_{1,k}$	$Q_{2,k}$
1	5	3.8971	51.626	23.2390	26.7099
2	5	4.0473	150.691	34.5854	38.8709
3	5	4.3400	97.683	28.1572	34.4601

TABLE III
PERFORMANCE WITH 4 PAIRS OF FULL-DUPLEX USERS

k	$P_{1,k}$	$P_{2,k}$	B_k	$Q_{1,k}$	$Q_{2,k}$
1	5	3.7464	26.720	22.4085	25.9285
2	5	4.9929	146.759	34.4014	39.0498
3	5	4.3079	36.258	26.7723	33.0397
4	5	4.9990	190.263	43.6321	40.0009

C. More than Two Pairs of Full-Duplex Users

In this subsection, we apply our optimal resource allocation algorithms to cases in which there are more than two pairs of full-duplex users Table II provides results on the optimal bandwidth and power allocation and the resulting video qualities when there 3 pairs of users. In these results, it is assumed that $\omega_{1,1} = \omega_{2,1} = 0.05$, $\omega_{1,2} = \omega_{2,2} = 0.3$ and $\omega_{1,3} = \omega_{2,3} = 0.15$. Moreover, we set $\theta_{1,1} = \theta_{2,1} = 0.1$, $\theta_{1,2} = \theta_{2,2} = 0.07$ and $\theta_{1,3} = \theta_{2,3} = 0.04$. Overall, optimal bandwidth and power allocation leads to a weighted sum quality of 33.9269dB. We notice that since the weights $\omega_{1,2}$ and $\omega_{2,2}$ of the second pair of users are the largest, most bandwidth (out of a total bandwidth of $B = 0.3$MHz $= 300$kHz) is allocated to these users. Also, it is interesting to note that due to the need to control the self-interference, several power levels are less than the maximum allowed peak power level of 5 (while at least one power value is at the peak level), as also noted in the previous cases.

Table III shows the performances of video transmissions between 4 pairs of full-duplex users again considering optimal bandwidth and power allocation with $\omega_{1,1} = \omega_{2,1} = 0.05$, $\omega_{1,2} = \omega_{2,2} = 0.2$, $\omega_{1,3} = \omega_{2,3} = 0.05$

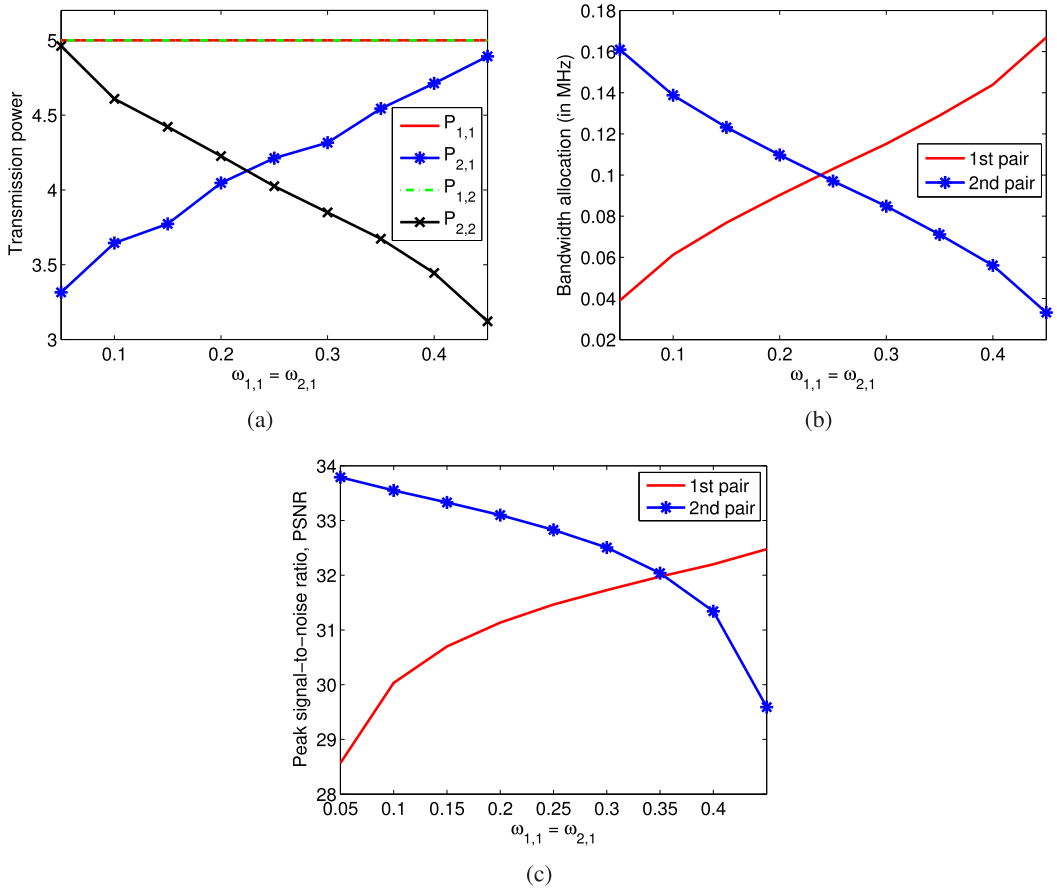


Fig. 11. (a) Optimal power allocation, (b) Optimal bandwidth allocation, and (c) the corresponding quality of video sequences as a function of $\omega_{1,1} = \omega_{2,1}$.

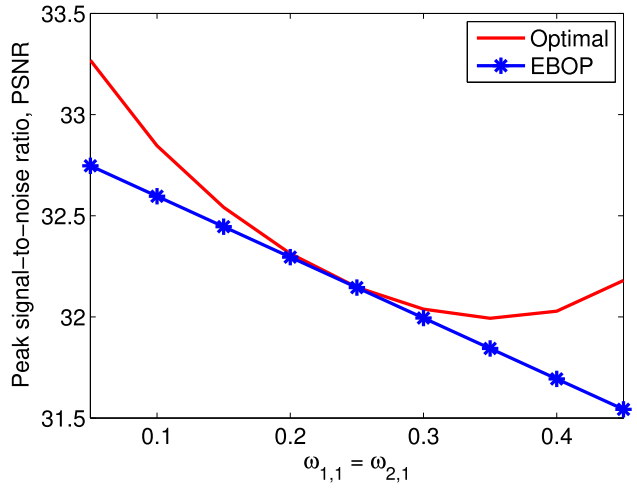


Fig. 12. Quality of video sequences as a function of $\omega_{1,1} = \omega_{2,1}$. Both optimal and equal bandwidth allocation are considered.

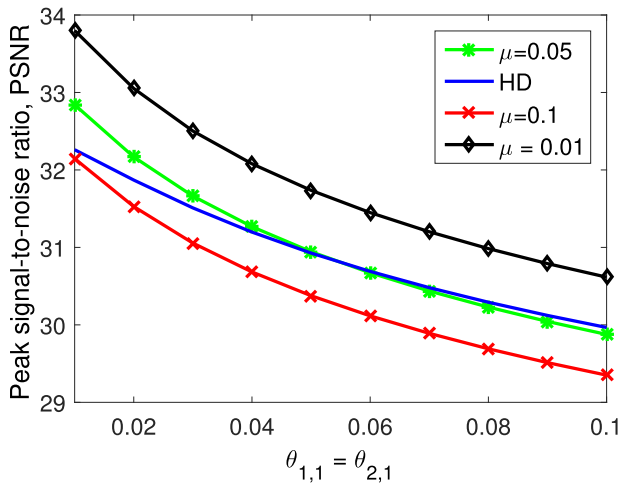


Fig. 13. PSNR vs. QoS exponents $\theta_{1,1} = \theta_{2,1}$ for both full-duplex and half-duplex (HD) operations. Self-interference suppression factors are $\mu = 0.1$, $\mu = 0.05$, and $\mu = 0.01$ in the three full-duplex curves, respectively.

and $\omega_{1,4} = \omega_{2,4} = 0.2$. The total bandwidth is $B = 0.4$MHz $= 400$kHz. It is further assumed that $\theta_{1,1} = \theta_{2,1} = 0.1$, $\theta_{1,2} = \theta_{2,2} = 0.07$, $\theta_{1,3} = \theta_{2,3} = 0.04$ and $\theta_{1,4} = \theta_{2,4} = 0.01$. The weighted sum quality of video sequences achieved with optimal allocations is 36.8243dB.

VI. Conclusion

In this paper, we have addressed the maximization of the weighted sum quality of received video sequences under total bandwidth, minimum video quality, maximum transmission

power, and delay QoS constraints in a full-duplex wireless model. LTPRS model is employed as the full-duplex model and the self-interference is measured by multiplying a self-interference factor with the transmission power. We have reformulated the original nonconvex optimization problem as a monotonic optimization problem, and developed algorithms to determine the optimal bandwidth and power allocation levels in an efficient manner using this framework.

We have gleaned several practical insights from our analysis. We have shown analytically that at least one of the transmission powers of the full-duplex user pair should be at its peak level in order to maximize the performance and minimize the bandwidth requirements. We have also noted that larger values of the QoS exponent θ lead to lower PSNR levels since more stringent delay constraints result in smaller video rates, lowering the quality. We have demonstrated that the user with a larger θ is allocated smaller transmission power and bandwidth. We have seen that video quality parameters have influence on optimal resource allocation policies, e.g., if the video quality increases faster with increased source rate (i.e., $a_{i,k}$ is larger for a video sequence), transmission power is higher. We have also shown that optimal bandwidth and power allocation has better performance than the equal bandwidth and optimal power (EBOP) allocation scheme, and the performance gap widens as the weight differences among the transmitted videos grow.

APPENDIX

A. Proof of Theorem 1

Let us assumer $P_1 \leq P^{\max}$ and $P_2 \leq P^{\max}$, and consider the function

$$V_1(P_1,P_2,B) = \left(\mathbb{E}_\gamma\left\{e^{-\theta B T_c \log\left(1+\frac{P_1\gamma}{N_0\,B+\mu P_2}\right)}\right\}\right)^{-1}. \tag{25}$$

We first show that V_1 is maximized if $P_1 = P^{\max}$ or $P_2 = P^{\max}$. Hence, at least one power value should be at the maximum level. Consider two power values strictly less than the maximum level, i.e., $P_1 < P^{\max}$ and $P_2 < P^{\max}$. Then, there exists some $\tau > 1$ such that $\tau P_1 \leq P^{\max}$ and $\tau P_2 \leq P^{\max}$. Then, considering the fraction in the exponent in (25), we can easily see for $\tau > 1$ that

$$\frac{\tau P_1\gamma}{N_0\,B+\mu\tau P_2} = \frac{P_1\gamma}{\frac{N_0\,B}{\tau}+\mu P_2} > \frac{P_1\gamma}{N_0\,B+\mu P_2}, \tag{26}$$

which leads to the result that

$$V_1(\tau P_1,\tau P_2,B) > V_1(P_1,P_2,B). \tag{27}$$

Hence, for given $P_1 < P^{\max}$ and $P_2 < P^{\max}$, we can increase the value of V_1 by increasing the power values to τP_1 and τP_2 for some $\tau > 1$ (with which the maximum power constraint $P^{\max}$ is still satisfied). Therefore, with this characterization, we conclude that in order to achieve the maximum value of V_1, we should have P_1 or P_2 attain its maximum value.

Next, we prove that V_1 is an increasing function of bandwidth B. Let us define $\chi = e^{-\theta T_c B \log\left(1+\frac{P_1\gamma}{N_0\,B+\mu P_2}\right)}$. Taking the first derivative of $V_1(P_1,P_2,B)$ with respect to B, we obtain (28), as shown at the bottom of this page,

Let us also define

$$g(x) = \ln\left(1+\frac{1}{x}\right) - \frac{1}{1+x}. \tag{29}$$

The first derivative of $g(x)$ with respect to x is

$$\frac{dg(x)}{dx} = -\frac{1}{x(1+x)^2} < 0, \tag{30}$$

and hence $g(\cdot)$ is a decreasing function of $x \geq 0$. Moreover, $\lim_{x\to 0} g(x) = \infty$ and $\lim_{x\to\infty} g(x) = 0$. Thus, $g(x) \geq 0$ for all $x \geq 0$, which also implies that

$$\ln\left(1+\frac{1}{x}\right) \geq \frac{1}{1+x} \quad \text{for } x \geq 0. \tag{31}$$

Now, assume $x = \frac{N_0\,B+\mu P_2}{P_1\gamma}$. Then, we have

$$\ln\left(1+\frac{P_1\gamma}{N_0B+\mu P_2}\right) \geq \frac{P_1\gamma}{N_0B+\mu P_2+P_1\gamma} \tag{32}$$

$$> \frac{P_1\gamma}{N_0B+\mu P_2+P_1\gamma}\frac{N_0B}{N_0B+\mu P_2}, \tag{33}$$

where (32) follows from (31), and (33) is due to the fact that $\frac{N_0\,B}{N_0\,B+\mu P_2} \leq 1$. The lower bound in (33) shows that the derivative in (28) is greater than zero because the numerator is greater than zero. Therefore, we conclude that V_1 is an increasing function of B.

Note that these derivations immediately apply to

$$V_2(P_1,P_2,B) = \left(\mathbb{E}_\gamma\left\{e^{-\theta B T_c \log\left(1+\frac{P_2\gamma}{N_0B+\mu P_1}\right)}\right\}\right)^{-1} \tag{34}$$

due to the symmetry and similarity in the formulations.

Finally, we consider two target values V_1^* and V_2^* for the functions V_1 and V_2, respectively, i.e., $V_1(P_1,P_2,B) = V_1^*$ and $V_2(P_1,P_2,B) = V_2^*$, and show that the minimum bandwidth B required to achieve these target values is attained if $P_1 = P^{\max}$ or $P_2 = P^{\max}$. Assume that both power values are strictly less than the maximum level, i.e., $P_1 < P^{\max}$ and $P_2 < P^{\max}$, and B_a is the bandwidth value with which we satisfy $V_1(P_1,P_2,B_a) = V_1^*$ and $V_2(P_1,P_2,B_a) = V_2^*$. Then, as also discussed above, there exists $\tau > 1$ such that $P_{1a} = \tau P_1 \leq P^{\max}$ and $P_{2a} = \tau P_2 \leq P^{\max}$. With these increased power levels, we now have $V1(P_{1a},P_{2a},B_a) > V1^*$ and $V2(P_{1a},P_{2a},B_a) > V2^*$ as shown in (26) and (27). Since V_1 and V_2 are increasing functions of B, there exists $B_b < B_a$, such that $V_1(P_{1a},P_{2a},B_b) = V_1^*$ and $V_2(P_{1a},P_{2a},B_b) = V_2^*$. Therefore, if both $P_1 < P^{\max}$ and $P_2 < P^{\max}$, we can

$$\frac{\partial V_1}{\partial B} = \frac{\theta T_c \mathbb{E}_\gamma\left\{\chi\left(\ln\left(1+\frac{P_1\gamma}{N_0\,B+\mu P_2}\right) - \frac{P_1\gamma_k N_0\,B}{(N_0\,B+\mu P_2+P_1\gamma)(N_0\,B+\mu P_2)}\right)\right\}}{(\mathbb{E}_\gamma\{\chi\})^2 \ln 2}. \tag{28}$$

always increase the power values and lower the bandwidth requirement while attaining the target levels V_1^* and V_2^*. Hence, the minimum required bandwidth is achieved if $P_1 = P^{\max}$ or $P_2 = P^{\max}$.

B. Proof of the Required Conditions for Obtaining the Upper Bound $\partial^+\mathcal{G}$

Assume that there exists an upper boundary point $\mathbf{V}^u$ such that $\sum_{k=1}^{K} B_k < B$, i.e., $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k) = V^u_{(i-1)K+k}$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}$. From Theorem 1 and its proof in Appendix VI-A, we know that $V_{(i-1)K+k}$ is an increasing function of B_k. Then, there exists a small positive δ such that $\sum_{k=1}^{K}(B_k+\delta) < B$ and $V_{(i-1)K+k}(P_{1,k}, P_{2,k}, B_k+\delta) > V^u_{(i-1)K+k}$, which implies that $\mathbf{V}^u$ is not a upper boundary point. Similarly, assume that there exists a upper boundary point $\mathbf{V}^u$ such that $P_{1,k} < P_{1,k}^{max}$ and $P_{2,k} < P_{2,k}^{max}$ for some $k \in \mathcal{K}$. Again, from the proof in Appendix VI-A, we know that we can find a $\tau > 1$ such that $\tau P_{1,k} < P_{1,k}^{max}$ and $\tau P_{2,k} < P_{2,k}^{max}$, and with these increased power values, we have $V_{(i-1)K+k}(\tau P_{1,k}, \tau P_{2,k}, B_k) > V^u_{(i-1)K+k}$. This also means that $\mathbf{V}^u$ is not a upper boundary point. Therefore, the upper boundary point $\mathbf{V}^u$ only occurs when $\sum_{k=1}^{K} B_k = B$ and at least one power value is at its maximum level, i.e., $P_{1,k} = P_{1,k}^{max}$ or $P_{2,k} = P_{2,k}^{max}$, for all $k \in \mathcal{K}$.

REFERENCES

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021," Cisco, San Jose, CA, USA, White Paper, Feb. 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] A. A. Khalek, C. Caramanis, and R. W. Heath, Jr., "Delay-constrained video transmission: Quality-driven resource allocation and scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 60–75, Feb. 2015.

[3] Y. Xu, R. Q. Hu, L. Wei, and G. Wu, "QoE-aware mobile association and resource allocation over wireless heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4695–4701.

[4] Y. Ma, H. Zhang, D. Yuan, and D. Jiang, "Power-efficient resource allocation with QoS guarantees for TDMA fading channels," *Wireless Commun. Mobile Comput.*, vol. 12, no. 11, pp. 1023–1036, Aug. 2012. [Online]. Available: http://dx.doi.org/10.1002/wcm.1036

[5] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation for multichannel communications over wireless links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4349–4360, Dec. 2007.

[6] Y. Wang, P. Ren, and Q. Du, "Statistical QoS driven power allocation for cognitive networks under primary user's outage probability constraint," in *Proc. IEEE 23rd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 167–172.

[7] A. Sadeghi, M. Luvisotto, F. Lahouti, S. Vitturi, and M. Zorzi, "Statistical QoS analysis of full duplex and half duplex heterogeneous cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[8] W. Cheng, X. Zhang, and H. Zhang, "QoS driven power allocation over full-duplex wireless links," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 5286–5290.

[9] Y. Wang and S. Mao, "Distributed power control in full duplex wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1165–1170.

[10] S. Buzzi, G. Colavolpe, D. Saturnino, and A. Zappone, "Potential games for energy-efficient power control and subcarrier allocation in uplink multicell OFDMA systems," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 89–103, Apr. 2012.

[11] J. Deng, R. Zhang, L. Song, Z. Han, G. Yang, and B. Jiao, "Joint power control and subchannel allocation for OFDMA femtocell networks using distributed auction game," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2012, pp. 1–6.

[12] T. Thanabalasingham, S. V. Hanly, L. L. H. Andrew, and J. Papandriopoulos, "Joint allocation of subcarriers and transmit powers in a multiuser OFDM cellular network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1. Jun. 2006, pp. 269–274.

[13] D. Wang, Z. Li, and X. Wang, "Joint optimal subcarrier and power allocation for wireless cooperative networks over OFDM fading channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 249–257, Jan. 2012.

[14] C. Ye, M. C. Gursoy, and S. Velipasalar, "Optimal resource allocation for full-duplex wireless video transmissions under delay constraints," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[15] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[16] Q. Du and X. Zhang, "Statistical QoS provisionings for wireless unicast/multicast of multi-layer video streams," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 420–433, Apr. 2010.

[17] C. Chen, X. Zhu, G. de Veciana, A. C. Bovik, and R. W. Heath, Jr., "Rate adaptation and admission control for video transmission with subjective quality constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 22–36, Feb. 2015.

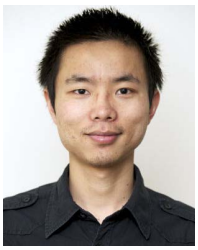
[18] L. Teixeira, "Rate-distortion analysis for H.264/AVC video statistics," in *Recent Advances on Video Coding*, J. D. S. Lorente, Ed. Oxford, U.K.: InTech, 2011.

[19] Y. J. Zhang, L. P. Qian, and J. Huang, "Monotonic optimization in communication and networking systems," *Found. Trends Netw.*, vol. 7, no. 1, pp. 1–75, Oct. 2013.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[21] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM J. Optim.*, vol. 11, no. 2, pp. 464–494, Nov. 2000. [Online]. Available: http://dx.doi.org/10.1137/S1052623499359828

[22] *Video Trace Library*. Accessed: Apr. 23, 2018. [Online]. Available: http://trace.eas.asu.edu/yuv/



Chuang Ye received the B.S. degree in information engineering and the M.S. degree in circuit and system from Zhejiang University, Zhejiang, China, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, Syracuse University. His research interests are in the fields of wireless communications, energy-efficient multimedia transmission techniques, cognitive radio systems, and smart camera systems.

M. Cenk Gursoy (S'01–M'04–SM'16) received the B.S. degree (Hons.) in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 1999, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2004. From 2004 to 2011, he was a Faculty Member with the Department of Electrical Engineering, University of Nebraska-Lincoln. He is currently a Professor with the Department of Electrical Engineering and Computer Science, Syracuse University. His research interests are in the general areas of wireless communications, information theory, communication networks, and signal processing. He is currently a member of the Editorial Boards of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was a recipient of the Gordon Wu Graduate Fellowship from Princeton University from 1999 to 2003, an NSF CAREER Award in 2006, the *EURASIP Journal of Wireless Communications and Networking* Best Paper Award, the 2017 IEEE PIMRC Best Paper Award, the 2017 IEEE Green Communications and Computing Technical Committee Best Journal Paper Award, the UNL College Distinguished Teaching Award, and the Maude Hammond Fling Faculty Research Fellowship. He was a Co-Chair of the Communication QoS and System Modeling Symposium, the 2017 International Conference on Computing, and the Networking and Communications. He is the Aerospace/Communications/Signal Processing Chapter Co-Chair of IEEE Syracuse Section. He also served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2010 to 2015, *Physical Communication* (Elsevier) from 2010 to 2017, the IEEE COMMUNICATIONS LETTERS from 2012 to 2014, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS-Series on Green Communications and Networking from 2015 to 2016.

Senem Velipasalar (M'04–SM'14) received the B.S. (Hons.) degree in electrical and electronic engineering from Bogazici University, Istanbul, Turkey, in 1999, the M.S. degree in electrical sciences and computer engineering from Brown University, Providence, RI, USA, in 2001, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 2004 and 2007, respectively. From 2001 to 2005, she was with the Exploratory Computer Vision Group, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. From 2007 to 2011, she was an Assistant Professor with the Department of Electrical Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA. She joined Syracuse University, Syracuse, NY, USA, in 2011, where she is currently an Associate Professor with the Department of Electrical Engineering and Computer Science. The focus of her research has been on mobile camera applications, wireless embedded smart cameras, multi-camera tracking and surveillance systems, and automatic event detection from videos. Her current research interests include embedded computer vision, video/image processing, distributed multi-camera systems, pattern recognition, and signal processing.

Dr. Velipasalar is a member of the Editorial Board of the *Journal of Signal Processing Systems* (Springer). She was a recipient of the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2006, the Faculty Early Career Development Award from the National Science Foundation in 2011, the Excellence in Graduate Education Faculty Recognition Award, the EPSCoR First Award, two Layman Awards, the IBM Patent Application Award, and Princeton and Brown University Graduate Fellowships. She has co-authored the paper, which received the third place award at the 2011 ACM/IEEE International Conference on Distributed Smart Cameras.