Energy-Efficient Cross-layer Resource Allocation for Heterogeneous Wireless Access

Lei Xu, *IEEE Member* and Weihua Zhuang, *IEEE Fellow*

Abstract

In this paper, an uplink cross-layer resource allocation problem based on imperfect channel state information (CSI) is modeled as min-max fractional stochastic programming for heterogeneous wireless access. The resource allocation is subject to constraints in delay, service outage probability, system radio bandwidth, and total power consumption. The joint bandwidth and power allocations are based on CSI at the physical layer and queue state information (QSI) at the link layer. In order to determine the transmission rate of each mobile terminal (MT) according to the queue buffer occupancy, a probability upper bound of exceeding the maximum packet delay in terms of a required transmission rate is presented based on M/D/1 model. Then, the bandwidth and power allocation problem is transformed into bi-convex programming, and an optimal distributed bandwidth and power allocation (ODBPA) algorithm is proposed. To reduce computational complexity, a suboptimal distributed bandwidth and power allocation (SDBPA) algorithm is presented. Simulation results demonstrate that the proposed algorithms improve the energy efficiency greatly.

Index Terms

Heterogeneous wireless access, energy efficiency, cross-layer resource allocation, min-max fractional stochastic programming.

Lei Xu is with School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China (E-mail: xulei_marcus@126.com).

W. Zhuang is with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. (wzhuang@uwaterloo.ca).

I. INTRODUCTION

A future wireless communication system is expected to integrate different radio access technologies, leading to a heterogeneous wireless access [1]. For example, a cellular network will include macrocells to support low-to-medium rate services with a large coverage area, and femtocells to support high rate services in hotspots. Integrating macrocells and femtocells can help to provide various classes of service to MTs and to support seamless user roaming [2]. In order to maximize user experience in a heterogeneous wireless access, extensive research works have been carried out to take advantage of multi-homing capability, where the data stream from an MT is split into multiple sub-streams and transmitted over multiple wireless media by different radio interfaces simultaneously in the uplink [3].

Existing studies in resource allocation for a heterogeneous wireless access can be divided into three categories [2], [4]–[13]. The first category is bandwidth allocation at the network layer, which aims to provide call-level quality-of-service (QoS) guarantee [2], [4], [5]. The available radio resources from multiple wireless access media can be aggregated to support services requiring a high transmission rate and to reduce call blocking probability [2]. The second category is packet scheduling of video traffic at the link layer, which provides packet-level QoS guarantee [6], [7]. The packet scheduler determines which packet should be assigned to which radio interface of an MT based on CSI, available radio resources, and video traffic characteristics [6]. The third category is joint bandwidth and power allocation at the physical layer, to meet bit-level QoS requirements [8]–[13]. Compared with the call-level and packet-level resource allocation, the bit-level resource allocation needs to jointly allocate the radio bandwidth and energy resources simultaneously. It not only exploits the multiuser diversity in wireless transmission among different MTs, but also takes advantage of the disparity of available resources among different wireless access media [13].

In [4], a bandwidth allocation algorithm with fairness consideration is proposed. In [5], a bandwidth allocation problem of video traffic is solved by convex optimization theory and the Karush-Kuhn-Tucher (KKT) condition. Further, a distributed prediction-based resource allocation algorithm for video traffic is presented to achieve an acceptable call blocking probability [2]. For

packet scheduling, an energy management algorithm is proposed in [6], to support a sustainable video transmission over the call duration and to guarantee a target video quality lower bound. Also, a novel scheduling framework with delay-constrained high definition video transmission featured by frame-level data protection is proposed [7].

For joint bandwidth and power allocation at the physical layer, there exist energy-efficient and spectral-efficient resource allocation algorithms. In [8], a joint link selection and resource allocation algorithm using the branch and bound method is proposed. In [9], an energy-per-bit minimized joint power, subchannel, and time allocation scheme for heterogeneous wireless networks is proposed with a double-loop iteration method. On the other hand, joint bandwidth and power allocation algorithms are proposed to maximize the spectral efficiency [10], [12], or to achieve max-min fairness and proportional fairness in resource allocation [11], [13].

The existing resource allocation mechanisms mainly limit to a particular networking protocol layer. However, cross-layer resource allocation has been shown to be beneficial in homogeneous wireless access [14], [15]. Extending cross-layer design to a heterogeneous wireless access requires further studies. In this paper, we investigate cross-layer resource allocation[1] based on CSI and QSI for uplink energy-efficient video transmission in a heterogeneous wireless access. Specially, we summarize our contributions as follows: (i) An uplink energy-efficient cross-layer resource allocation problem is formulated as min-max fractional stochastic programming, and the probability upper bound of exceeding the maximum packet delay at the link layer is transformed to a required transmission rate based on an M/D/1 queue system; (ii) Using the Dinkelbach-type method, the min-max fractional stochastic programming problem is transformed into a bi-convex optimization problem, and an optimal distributed energy-efficient bandwidth and power allocation algorithm based on the dual decomposition method is proposed; (iii) In order to reduce the computational complexity, a suboptimal distributed algorithm is developed. Simulation results demonstrate that the proposed algorithms greatly improve the energy efficiency of video packet transmission in a heterogeneous wireless access.

---

[1]In this work, the main design objective of the solution maximizes the energy efficiency of heterogeneous wireless networks with the QoS constraints. Additionally, the uplink energy-efficient cross-layer resource allocation can be extended to the downlink transmission easily [16].

The rest of this paper is organized as follows. The system model and problem formulation are described in Sections II. Optimal and suboptimal bandwidth and power allocation algorithms are presented in Section III. The computational complexity and signaling overhead analysis is given in Section IV. Finally, performance evaluation and conclusions are given in Sections V and VI, respectively.

II. System Model and Problem Formulation

In this section, the system and video traffic models are described. Then, the transmission rate based on imperfect CSI, and power consumption models are presented. Finally, an uplink energy-efficient cross-layer resource allocation problem is formulated.

A. System Description

Consider a geographical region covered by heterogeneous wireless access. The coverage areas of BSs/APs can overlap as shown in Fig. 1. In Fig. 1, there exist $\mathcal{N} = \{1, 2, \cdots, N\}$ wireless access networks, e.g., Macrocell, Femtocell, and Wi-Fi, based on different access technologies and operated by different service providers[2]. There is a set, $\mathcal{S}_n = \{1, 2, \cdots, S_n\}$, of base stations (BSs)/access points (APs) for network n. There is a set, $\mathcal{M} = \{1, 2, \cdots, M\}$, of MTs in the geographical region, and $\mathcal{M}_{ns} = \{1, 2, \cdots, M_{ns}\} \in \mathcal{M}$ is a subset of MTs which reside in the coverage area of network n BS/AP s. Using multi-homing functions and multiple radio interfaces, an MT can communicate with multiple BSs/APs within its transmission range simultaneously. For example, an MT in the overlapped area with one macrocell and one femtocell can communicate with one macrocell BS and one femtocell BS simultaneously. In the heterogeneous wireless access, cooperation of different networks has a potential to improve service quality for MTs and enhance network performance. Hence, different networks cooperate in resource allocation via signaling exchanges over a wireline backbone. Time is partitioned into time slots, $\mathcal{T} = \{1, 2, \cdots\}$, of equal duration τ. The resource allocation is performed at the beginning of each time slot,

---

[2]Assuming different wireless access technologies are operated by different service providers is only an assumption. We can assume different wireless access technologies are operated by the same service provider. However, some literatures adopt the assumption that different wireless access technologies are operated by different service providers, e.g. [2], [6], [13], and we follow this assumption.
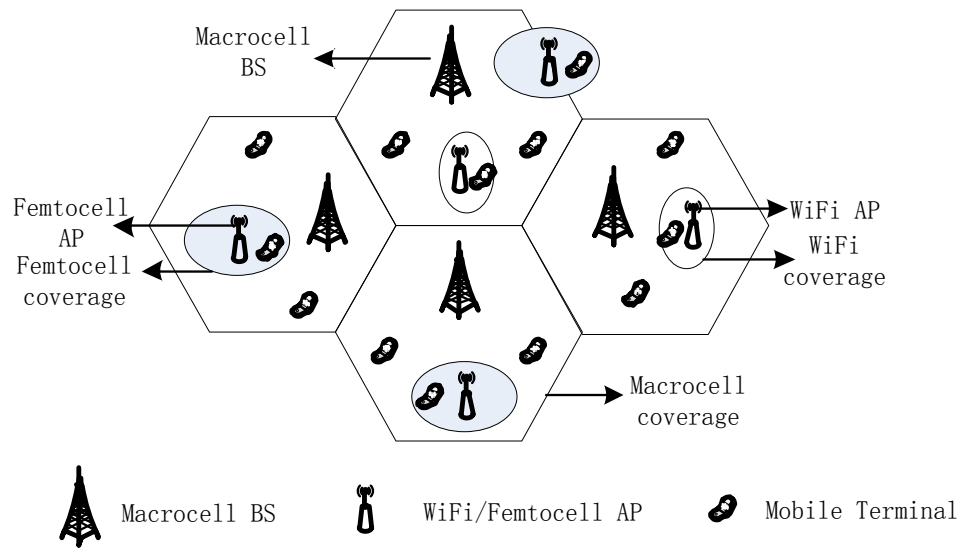
Fig. 1.   Heterogeneous wireless access.

remains constant within one time slot and varies from one time slot to another time slot. Consider video applications [17], [18]. In order to improve the coding efficiency of multiview video coding, variable block-size motion estimation, disparity estimation, and multiple reference frames selection are adopted [19]–[21]. Each MT has a video packet flow to transmit via nearby BSs/APs, which are the BSs/APs within MT's transmission range. For video applications, each packet should be transmitted before a deadline. If the transmission delay exceeds the bound, the packet is dropped from its queue at the source. At heterogeneous wireless access, we have considered the co-existence of LTE and WiFi. In Fig. 1, there are Macrocell, Femtocell, and Wi-Fi, based on different access technologies and operated by different service providers. Additionally, we follow the previous literatures of heterogeneous wireless access in the system model, i.e., the co-existence of LTE and WiFi e.g., [2], [10], [11], [13]. They study the bandwidth and power allocation for heterogeneous wireless access. On the other hand, the video traffic transmission for heterogeneous wireless access has been studied in [6] for heterogeneous wireless access.

There are some assumptions in this work, i.e.,

1) In the same network, different BSs/APs reuse the same spectral bands, and interference mitigation is achieved by interference management schemes [22]. Consequently, we do not consider the co-channel interference. On the other hand, at the different networks, the cross-tier interference is also not considered for bandwidth allocation at heterogeneous wireless access, e.g., [2], [13]. Therefore, we follow the interference assumptions for the heterogeneous wireless access in the previous literatures [2], [13].

2) Each MT has a packet queueing buffer and the buffer size is infinite.

3) The channel power gains remain constant within one time slot, and vary from one time slot to another time slot and from one link to another link, independently.

4) The video packet arrivals at the transmission buffer of MT m follow a Poisson process with an average arrival rate λ_m [23] with a constant packet length of L bits [6].

The cross-layer resource allocation design is considered in this work. We consider the queueing buffer for each MT at the link layer, and joint bandwidth and power allocation at the physical layer. For each MT, the packet delay requirement for video traffic[3] and the occupancy at the the queueing buffer influence the required minimum transmission rate at the physical layer. On the other hand, the joint bandwidth and power allocation determines the practical physical transmission rate, and influences the performance of the packet delay at the link layer. Consequently, it is necessary to design the joint bandwidth and power allocation at the physical layer incorporating the delay requirement information at the link layer.

B. Transmission Rate Based on Imperfect CSI

In wireless communications, the transmission rate should be determined based on CSI. BSs/APs estimate CSI for uplink transmission, and send it to MTs. Channel estimation algorithms can be used for heterogeneous radio networks based on machine learning techniques, e.g., support vector machine [26]–[29]. Because there exist channel estimation errors and feedback delays,

---

[3]Although it is difficult to determine user experience from video traffic only with delay, delay is an important parameter of video traffic, e.g., [24], [25]. In this work, we select delay at link layer to design the cross-layer resource allocation algorithm.

the CSI available at the transmitter of an MT is usually imperfect [30]. Using imperfect CSI can result in a scheduled transmission rate greater than what the system can support (e.g., Shannon capacity for simplicity), resulting in outage events. The outage probability, P_{nsm}^o, for MT $m \in \mathcal{M}_{ns}$ based on imperfect CSI, is

$$P_{nsm}^o = \Pr\left[R_{nsm} > B_{nsm}\log_2\left(1 + \frac{P_{nsm}|\alpha_{nsm}|^2}{B_{nsm}n_0}\right)\right] \tag{1}$$

where B_{nsm} is the allocated bandwidth from network n BS/AP s to MT m, P_{nsm} is the allocated power of MT m for transmission to network n BS/AP s, R_{nsm} is the scheduled transmission rate for the MT m uplink, n_0 is the one-sided noise power spectral density, and α_{nsm} is the complex uplink channel gain [24]. Since, in the same network, different BSs/APs reuse the same spectral bands and interference mitigation is achieved by interference management schemes [22], no interference is considered in Eq. (1), as in some existing studies, e.g., [8], [12]. In each cell, frequency division duplexing (FDD) is adopted for uplink and downlink.

Due to channel estimate errors, the estimate complex uplink channel gain $\hat{\alpha}_{nsm}$, e.g., using a minimum mean square error (MMSE) channel estimator [31], is not equal to α_{nsm}. Let $\alpha_{nsm} = \hat{\alpha}_{nsm} + \Delta\alpha_{nsm}$, where the estimation error, $\Delta\alpha_{nsm}$, is assumed to be a zero-mean complex Gaussian random variable with variance σ_{nsm}^2, and is independent for different radio interfaces. Therefore, $|\alpha_{nsm}|$ follows a Rice distribution with the cumulative distributed function (CDF) given by

$$F_{|\alpha_{nsm}|}\left(x, |\hat{\alpha}_{nsm}|, \sigma_{nsm}^2\right) = 1 - Q_1\left(\frac{|\hat{\alpha}_{nsm}|}{\sigma_{nsm}}, \frac{x}{\sigma_{nsm}}\right) \tag{2}$$

where $Q_1\left(|\hat{\alpha}_{nsm}|/\sigma_{nsm}, x/\sigma_{nsm}\right)$ is the Marcum Q-function.

According to (1), the outage probability is required to satisfy

$$P_{nsm}^o = \Pr\left[|\alpha_{nsm}| < \sqrt{\frac{\left(2^{\frac{R_{nsm}}{B_{nsm}}} - 1\right)B_{nsm}n_0}{P_{nsm}}}\right] \leq \varepsilon_{nsm} \tag{3}$$

where ε_{nsm} is an upper bound of the outage probability.

In a practical wireless communication system, rate R_{nsm} is a discrete value. However, in most

existing studies on resource allocation at the physical layer, R_{nsm} is assumed to be a continuous value in bandwidth and power allocations, e.g., [2], [9], [13]. Under this assumption, from (3), the scheduled transmission rate to satisfy the required outage probability for MT m uplink to network n BS/AP s is

$$\begin{cases} R_{nsm} = B_{nsm} \log_2 \left(1 + SNR_{nsm}\right) \\ SNR_{nsm} = \frac{P_{nsm} \left[F_{|\alpha_{nsm}|}^{-1} \left(\varepsilon_{nsm}, |\hat{\alpha}_{nsm}|, \sigma_{nsm}^2\right)\right]^2}{B_{nsm} n_0} \end{cases} \tag{4}$$

where $F_{|\alpha_{nsm}|}^{-1}$ is the inverse CDF of $|\alpha_{nsm}|$, and SNR_{nsm} is the signal noise ratio for MT $m \in \mathcal{M}_{ns}$.

C. Power Consumption Model

The power consumption at each MT consists of two components. The first component is a fixed power consumption P_c, which captures the power consumption of hardware at transmitter. The second component is a transmission power consumption. Consequently, the total power consumption, P_m, for MT m is

$$P_m = P_c + \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} P_{nsm}. \tag{5}$$

In Eq. (5), if an MT does not transmit to all nearby BSs/APs, some P_{nsm} terms in $\sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} P_{nsm}$ are equal to zero.

D. Problem Formulation

The total allocated bandwidth by network n BS/AP s should not be larger than its total available bandwidth, i.e.,

$$\sum_{m \in \mathcal{M}_{ns}} B_{nsm} \leq B_{ns} \tag{6}$$

where B_{ns} is the total available bandwidth at network n BS/AP s.

The total power consumption at MT m should satisfy its maximum power constraint, i.e.,

$$P_m \leq P_m^T \tag{7}$$

where P_m^T is the total available power at MT m and is assumed to be a constant.

The packet delay for MT m should probabilistically satisfy the maximum packet delay constraint[4], given by

$$\Pr\left(D_m > D_m^{\max}\right) \leq \gamma_m \tag{8}$$

where D_m is the time of a packet from its generation to its transmission, $D_m^{\max}$ is the maximum packet delay after which the packet will be dropped at the transmitter, and γ_m is the probability upper bound of exceeding the maximal packet delay.

Define the consumed energy per bit of MT m, η_m, as the ratio of the total power consumption to the total achieved transmission rate, i.e., $\eta_m = P_m/R_m$. $R_m = \sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}_n} R_{nsm}$ is the achieved transmission rate for MT m. To guarantee the consumed energy fairness for all MTs, we minimize the maximum consumed energy per bit, η_m, for all MTs. Consequently, the resource allocation problem is formulated as

$$OP1 \min_{B_{nsm},P_{nsm}} \{\max \eta_m\}$$
$$\text{s.t. } (3), (6) - (8), \tag{9}$$
$$B_{nsm} \geq 0, P_{nsm} \geq 0.$$

Problem (9) is a min-max fractional stochastic programming, where (8) is a stochastic constraint. Additionally, constraint (8) on the delay requirement is not directly related to the bandwidth and power allocations at the physical layer. In order to solve problem (9), we first analyze the relationship between the packet delay at the link layer and then the transmission rate at the physical layer, and transform the stochastic programming problem (9) into a min-max fractional deterministic programming problem.

III. JOINT BANDWIDTH AND POWER RESOURCE ALLOCATION

In this section, we first analyze the packet delay. Then, we propose an optimal distributed bandwidth and power allocation algorithm based on the dual decomposition method. Finally,

---

[4]The probabilistic delay constraint is very useful for wireless communication systems, e.g., [32], [33].

a suboptimal distributed bandwidth and power allocation algorithm is presented to reduce the computational complexity.

A. Packet delay Analysis

Although advanced video codecs are developed, the resource at each wireless network is limited and video packet will be dropped at the transmitter when exceeding a delay threshold, e.g., [34]. Given the service rate[5], the service time for each packet at each MT is deterministic. Hence, the packet buffer is modeled as an M/D/1 queueing system[6]. For the stability of the queue, $\rho_m = \lambda_m L / R_m < 1$. Let $\pi = (\pi_0, \pi_1, \cdots)$ denote the stationary distribution of the number of packets at the queueing system. By the Pollaczek-Khinchin formula [35], we have the probability generating function $\pi(z)$, i.e.,

$$\pi(z) = \frac{(1 - \rho_m)(1 - z) e^{\rho_m(z-1)}}{e^{\rho_m(z-1)} - z}. \tag{10}$$

Utilizing the Taylor expansion of $\pi(z)$ [36], the stationary probability, π_i, is given by

$$
\begin{aligned}
\pi_0 &= 1 - \rho_m \\
\pi_1 &= (1 - \rho_m)(e^{\rho_m} - 1) \\
\pi_i &= (1 - \rho_m)(e^{i\rho_m} + \varpi_i), i \geq 2 \\
\varpi_i &= \sum_{j=1}^{i-1} e^{j\rho_m}(-1)^{i-j}\left[\frac{(j\rho_m)^{i-j}}{(i-j)!} + \frac{(j\rho_m)^{i-j-1}}{(i-j-1)!}\right].
\end{aligned} \tag{11}
$$

Since the service time for all packets in the M/D/1 queueing system is the same given the packet length and service rate, the upper bound for the number of packets in the queue buffer to meet the maximum delay requirement is $Q_m^{\max} = \lfloor D_m^{\max} R_m / L \rfloor$, where $\lfloor x \rfloor$ is the floor function. For each new packet arrival, the packet delay is the ratio of the number of packets at the queueing system to the service rate R_m. Therefore, we have the following relationship for the queueing system,

$$\Pr(Q_m > Q_m^{\max}) = \Pr(D_m > D_m^{\max}). \tag{12}$$

---

[5]In this work, the service rate of the packet M/D/1 queueing system at MT m is defined as the transmission rate, R_m.

[6]In OFDMA system, the cross-layer resource allocation is modeled as M/G/1 queueing system for average delay requirement, e.g., [14], while our work adopt M/D/1 queueing system to solve the probabilistic delay requirement.

where Q_m is the number of packets at the queueing system for MT m.

On the other hand, we have

$$\Pr\left(Q_m > Q_m^{\max}\right) = 1 - \sum_{i=0}^{Q_m^{\max}} \pi_i. \tag{13}$$

From (12)-(13), constraint (8) can be rewritten as

$$1 - \sum_{i=0}^{Q_m^{\max}} \pi_i \leq \gamma_m. \tag{14}$$

From (11) and (14), we can obtain the minimum transmission rate, $\psi\left(D_m^{\max}, \gamma_m, \lambda_m, L\right)$, as a function of the delay requirement, the packet average arrival rate, and the packet length, by the binary search method over a range $\left[R_m^{\min}, R_m^{\max}\right]$, where $R_m^{\min}$ and $R_m^{\max}$ are the search lower and upper bounds for MT m, e.g., $R_m^{\min} = \lambda_m L$ and $R_m^{\max} = 10\lambda_m L$. The transmission rate for MT m should be at least the required minimum transmission rate in order to satisfy the packet delay requirement, i.e.,

$$R_m \geq \psi\left(D_m^{\max}, \gamma_m, \lambda_m, L\right). \tag{15}$$

B. Optimal Distributed Bandwidth and Power Allocation

With the delay requirement in (8) equivalently represented by the minimum transmission rate constraint (15), problem (9) is transformed into the min-max fractional deterministic programming. Solving the problem mean first solving the fractional programming problem and then solving the min-max programming problem. For the first subproblem, the Dinkelbach-type method is adopted [37], and $\xi = \max_{m \in \mathcal{M}} \eta_m$ is defined. For a given parameter, $\xi = \max_{m \in \mathcal{M}} \eta_m$, the objective function in (9) is rewritten as

$$F(\xi) = \min_{B_{nsm}, P_{nsm}} \left\{ \max_{m \in \mathcal{M}} \left(P_m - \xi R_m\right) \right\}. \tag{16}$$

In order to obtain the optimal resource allocation solution, we find a root of equation $F(\xi) = 0$. This is the common procedure to solve the fractional programming problem [37]. For the second subproblem, $\vartheta = \max_{m \in \mathcal{M}} \left(P_m - \xi R_m\right)$, and add the constraint $P_m - \xi R_m \leq \vartheta$ to transform the

min-max programming problem to minimize the programming problem. Additionally, problem (9) is rewritten as

$$
\begin{aligned}
OP2 \min_{B_{nsm}, P_{nsm}, \vartheta} \quad & \vartheta \\
\text{s.t. } \quad & P_m - \xi R_m \leq \vartheta, \\
& (6), (7), (15), \\
& B_{nsm} \geq 0, P_{nsm} \geq 0, \vartheta \geq 0.
\end{aligned}
\tag{17}
$$

In problem (17), we solve the variables ϑ, P_{nsm}, and B_{nsm}, separately. Firstly, the variables P_{nsm}, and B_{nsm} are solved via the dual decomposition method with the fixed variable ϑ. Then, ϑ is obtained via binary search method with the given variables P_{nsm} and B_{nsm}.

Proposition 1: Problem (17) is a bi-convex optimization problem with the variables B_{nsm} and P_{nsm}.

Proof: See Appendix A.

Problem (17) is a bi-convex optimization problem, and the optimal solutions for the primal and dual problems are equal [38]. The Lagrangian function for (17) is

$$
\begin{aligned}
L(\vartheta, u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm}) = {} & \vartheta - \sum_{m \in \mathcal{M}} \alpha_m \left(P_m^T - P_m \right) \\
& - \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} v_{ns} \left(B_{ns} - \sum_{m \in \mathcal{M}_{ns}} B_{nsm} \right) - \sum_{m \in \mathcal{M}} u_m \left(\vartheta - P_m + \xi R_m \right) \\
& - \sum_{m \in \mathcal{M}} \beta_m \left(R_m - \psi \left(D_m^{\max}, \gamma_m, \lambda_m, L \right) \right)
\end{aligned}
\tag{18}
$$

where u_m is a Lagrangian multiplier for the first constraint condition in (16), v_{ns}, α_m and β_m are the Lagrangian multipliers for (6), (7), and (15), respectively.

The dual function $h\left(\vartheta, u_m, v_{ns}, \alpha_m, \beta_m \right)$ is

$$
\min_{B_{nsm}, P_{nsm}} L(\vartheta, u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm}).
\tag{19}
$$

The dual problem is

$$OP3 \max_{u_m, v_{ns}, \alpha_m, \beta_m} h(\vartheta, u_m, v_{ns}, \alpha_m, \beta_m) \tag{20}$$

$$\text{s.t. } u_m \geq 0, v_{ns} \geq 0, \alpha_m \geq 0, \beta_m \geq 0.$$

For each MT, the Lagrangian function in (18) can be simplified to

$$L_m(u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm}) = u_m (\xi R_m - P_m)$$
$$-\alpha_m P_m + \beta_m \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} R_{nsm} - \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} v_{ns} B_{nsm}. \tag{21}$$

Consequently, each MT solves its own utility minimization problem

$$OP4 \min_{B_{nsm}, P_{nsm}} L_m(u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm}) \tag{22}$$

$$\text{s.t. } B_{nsm} \geq 0, P_{nsm} \geq 0.$$

Problem (22) has the bandwidth allocation variable (B_{nsm}) and the power allocation variable (P_{nsm}), and is a bi-convex optimization problem [39]. If we want to obtain the optimal solution of bandwidth allocation variable B_{nsm}, we need to fix the power allocation variable P_{nsm}, and vice versa. By iteratively updating the Lagrangian multipliers, we can design the optimal distributed bandwidth and power allocation algorithm in a recursive manner. The optimal bandwidth allocation B_{nsm}, given P_{nsm}, u_m, v_{ns}, α_m and β_m, is calculated by applying the KKT condition on (22),

$$\frac{\partial L_m(u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm})}{\partial B_{nsm}} = 0 \tag{23}$$

which results in

$$\log_2 (1 + SNR_{nsm}) - \frac{SNR_{nsm}}{(1 + SNR_{nsm}) \ln 2} = \frac{v_{ns}}{\xi u_m + \beta_m}. \tag{24}$$

Using the Newtons method on (24), the optimal bandwidth solution B_{nsm} is denoted by

$$B_{nsm} = [\phi (P_{nsm}, u_m, v_{ns}, \alpha_m, \beta_m)]^+ \tag{25}$$

where $[x]^+$ is a projection of x on the positive orthant.

The optimal power allocation P_{nsm}, given B_{nsm}, u_m, v_{ns}, α_m and β_m, is calculated with (26)

by applying the KKT condition on (21),

$$\frac{\partial L_m(u_m, v_{ns}, \alpha_m, \beta_m, B_{nsm}, P_{nsm})}{\partial P_{nsm}} = 0. \tag{26}$$

From (26), we obtain

$$P_{nsm} = \left[\frac{B_{nsm}(\xi u_m + \beta_m)}{\alpha_m} - \frac{B_{nsm} n_0}{\left[F_{|\alpha_{nsm}|}^{-1}\left(\varepsilon_{nsm}, |\hat{\alpha}_{nsm}|, \sigma_{nsm}^2 \right) \right]^2} \right]^+. \tag{27}$$

The optimum values of u_m, v_{ns}, α_m and β_m are calculated by solving the dual problem (20), and a gradient descent method can be applied to calculate u_m, v_{ns}, α_m and β_m [38], i.e.,

$$u_m(i+1) = \left[u_m(i) + \Delta\kappa_1 \left(\vartheta - P_m + \xi R_m \right) \right]^+ \tag{28}$$

$$v_{ns}(i+1) = \left[v_{ns}(i) + \Delta\kappa_2 \left(B_{ns} - \sum_{m \in \mathcal{M}} B_{nsm} \right) \right]^+ \tag{29}$$

$$\alpha_m(i+1) = \left[\alpha_m(i) + \Delta\kappa_3 \left(P_m^T - P_m \right) \right]^+ \tag{30}$$

and

$$\beta_m(i+1) = \left[\beta_m(i) + \Delta\kappa_4 \left(R_m - \psi \left(D_m^{\max}, \gamma_m, \lambda_m, L \right) \right) \right]^+ \tag{31}$$

where i is the iteration index, and $\Delta\kappa_j$, $j = 1, \cdots, 4$, is the step size.

Given parameter ξ, the optimal distributed bandwidth and power allocations are presented in Algorithm 1. Although B_{nsm} and P_{nsm} are calculated in the first line, they may not satisfy the constraints in problem (17). Consequently, the proposed algorithm is designed in a recursive manner converging to the optimum solution by updating the Lagrangian multipliers with (28)-(31) [38]. Additionally, the remaining steps are to determine the resource allocation solution to satisfy the constraints. Then, ϑ is solved via the binary search method. Finally, the Dinkelbach-type method can be applied to find the root of equation $F(\xi) = 0$ in an iterative manner according to the output, $F(\xi)$ [37]. Note that $F_m(\xi)$ is a parameter of the Dinkelbach-type method for MT m, ε_p is a small positive number; $\vartheta(i-1)$ and $\vartheta(i)$ are the variable values at the $(i-1)$th and

---

Algorithm 1 Optimal Distributed Bandwidth and Power Allocation (ODBPA).

Input: $D_m^{\max}$, γ_m, λ_m, ϑ, L, ξ, B_{ns} and P_m^T.

Output: B_{nsm}, P_{nsm} and $F(\xi)$.

1: Initialize $\alpha_m(i)$, $\beta_m(i)$, $u_m(i)$, $v_{ns}(i)$, P_{nsm}, B_{nsm} and $i = 1$.

2: **repeat**

3:      Each MT calculates R_{nsm}, B_{nsm} and P_{nsm} according to (4), (25) and (27).

4:      Each BS/AP calculates $v_{ns}(i+1)$ with (29), and each MT calculates $u_m(i+1)$, $\alpha_m(i+1)$ and $\beta_m(i+1)$ based on (28), (30) and (31).

5:      **if** $R_m \geq \psi\left(D_m^{\max}, \gamma_m, \lambda_m, L\right)$ **then**

6:         **if** $|\eta_m(i) - \eta_m(i-1)| \leq \varepsilon_p$ **then**

7:            Each MT calculates $F_m(\xi) = P_m - \xi R_m$, and all BSs/APs find $F(\xi) = \min\limits_{m \in \mathcal{M}} F_m(\xi)$. Go to step 12.

8:         **end if**

9:      **end if**

10:     Set $i = i + 1$.

11: **until**

12: Output B_{nsm}, P_{nsm} and $F(\xi)$.

---

ith iterations, respectively; $\eta_m(i-1)$ and $\eta_m(i)$ are the consumed energy per bit for MT m at the $(i-1)$th and ith iterations, respectively; $\alpha_m(i)$, $\beta_m(i)$, $u_m(i)$ and $v_{ns}(i)$ are the Lagrangian multipliers at the ith iteration, and $\alpha_m(i+1)$, $\beta_m(i+1)$, $u_m(i+1)$ and $v_{ns}(i+1)$ are the updated Lagrangian multipliers at the $(i+1)$th iteration.

The ODBPA is implemented in the distributed manner by cooperation among BSs/APs of different networks and MTs. This is because different networks are operated by different service providers and each MT utilizes the multi-homing technology to obtain the aggregated bandwidth from different networks. If we design a centralized algorithm to solve the cross-layer resource allocation problem, a central controller from a third party collects the CSI and QSI for all MTs, runs the algorithm and feeds the results back to all MTs. However, the proposed distributed resource allocation algorithm (i.e., the ODBPA) updates the Lagrangian multipliers $\alpha_m(i)$, $\beta_m(i)$ and $u_m(i)$ at each MT, and $\vartheta(i)$ and $v_{ns}(i+1)$ at each BS/AP. Through the wireless broadcast, each MT can receive the updating variables from its serving BSs/APs, and the bandwidth and power allocations are calculated at each MT. Hence, it is desirable to design the distributed resource allocation algorithm for a heterogeneous wireless medium.

C. Suboptimal Algorithm

The ODBPA gives an upper bound of network performance, but its computational complexity is high due to its bandwidth and power allocations in the recursive manner. This motives us to develop a suboptimal distributed bandwidth and power allocation algorithm (SDBPA) with lower computational complexity[7], and its bandwidth allocation and power allocation are designed separately. Hence, the SDBPA includes a bandwidth allocation algorithm and a power allocation algorithm[8]. In the bandwidth allocation algorithm, the power across different radio interfaces at each MT is allocated equally[9], and the bandwidth is allocated, based on the greedy method, to different MTs and different radio interfaces at each MT. In the power allocation algorithm, the bandwidth allocation is fixed, and the power is allocated across different radio interfaces at each MT, by the greedy method to maximize the energy efficiency.

The bandwidth allocation algorithm in SDBPA is given in Algorithm 2, where B_{nsm}^t and R_{nsm}^t are the temporary bandwidth allocation and transmission rate variables for MT $m \in \mathcal{M}_{ns}$, respectively; ΔB is the bandwidth allocation increment, B_{ns}^r is the remaining bandwidth for network n BS/AP s; η_m^t and R_m^t are the temporary energy-efficient and transmission rate variables for MT m, respectively. In Algorithm 2, the MT with minimum transmission rate is selected, and the selected MT allocates its bandwidth increment to the radio interface with highest transmission rate. The iterative procedure is repeated until the minimum transmission rate for each MT is satisfied. If there is any unallocated bandwidth, the remaining bandwidth is allocated to improve the energy efficiency. The power allocation algorithm can be obtained via modifying Algorithm 2, by replacing the bandwidth allocation update and the remaining bandwidth check with the power allocation update and the remaining power check.

---

[7]The computational complexity analysis and the signaling overhead analysis for the ODBPA and SDBPA are given in Appendix B and Appendix C.

[8]If the bandwidth allocation algorithm and the power allocation algorithm for SDBPA are designed in a recursive manner, the performance can be improved further. For simplicity, we adopt one recursive.

[9]The equal power allocation assumption is inspired by the suboptimal energy-efficient subcarrier and power allocation algorithms for OFDMA system, e.g., [40].

Algorithm 2 Bandwidth Allocation Algorithm in SDBPA.

Input: $D_m^{\max}$, γ_m, λ_m, L, B_{ns} and P_m^T.

Output: B_{nsm}.

1: Initialize B_{nsm}, B_{nsm}^t, R_{nsm}^t, P_{nsm} and R_m.

2: **repeat**

3:    All BSs/APs find $m^* = \min\limits_{m \in \mathcal{M}} R_m$, and MT m^* calculates $B_{nsm^*}^t = B_{nsm^*} + \Delta B$, $R_{nsm^*}^t$, $(n^*, s^*) = \max\limits_{(n,s) \in \mathcal{N}, \mathcal{S}_n} R_{nsm^*}^t$, $B_{n^*s^*m^*}$, η_{m^*} and R_{m^*}.

4:    **if** $R_m \geq \psi\left(D_m^{\max}, \gamma_m, \lambda_m, L\right)$ **then**

5:        All BSs/APs find $m^* = \max\limits_{m \in \mathcal{M}} \eta_m$, and MT m^* calculates $B_{nsm^*}^t = B_{nsm^*} + \Delta B$, $R_{nsm^*}^t$ and $(n^*, s^*) = \max\limits_{(n,s) \in \mathcal{N}, \mathcal{S}_n} R_{nsm^*}^t$.

6:        Network n^* BS/AP s^* calculates $\eta_{m^*}^t$ and $B_{n^*s^*}^r = B_{n^*s^*} - \sum\limits_{m \in \mathcal{M}_{n^*s^*}} B_{n^*s^*m}$.

7:        **if** $B_{n^*s^*}^r > \Delta B$ and $\eta_{m^*}^t < \eta_{m^*}$ **then**

8:            MT m^* updates $B_{n^*s^*m^*}$, and go to step 5.

9:        **else**

10:           Go to step 14.

11:        **end if**

12:    **end if**

13: **until**

14: Output B_{nsm}.

IV. Performance Evaluation

This section presents simulation results for the proposed energy-efficient cross-layer resource allocation algorithms. The simulation tool is Matlab 9.0. The simulation scenario is a hexagon geographical region, which is covered by a macro BS and two femto APs and is divided into three service areas. In the first service area, MTs can obtain service only from the macro BS. In the second and third service areas, MTs can obtain service from both the macro BS and one femto AP. The estimate complex uplink channel gain, $\hat{\alpha}_{nsm}$, is generated by the complex uplink channel gain, $\alpha_{nsm} = d_{nsm}^{-\chi} h_{nsm}$, minus the complex Gaussian random variable, $\Delta\alpha_{nsm}$, where d_{nsm} is calculated by the distance between MT m and network n BS/AP s, and the fast fading process $|h_{nsm}|$ follows the Rayleigh distribution with the scale parameter 1. The macro BS is deployed at the center of the hexagon coverage area with the circumradius of 500 m. Each femto AP provides a circle service area with a radius 20 m. The number of MTs at each service area is $M_{ns} = 5$. The probability upper bound of exceeding the maximum packet delay for MT m is

$\gamma_m = 0.01$. The variance of the complex Gaussian random variable, $\Delta\alpha_{nsm}$, for MT $m \in \mathcal{M}_{ns}$ is $\sigma_{nsm}^2 = 0.03$ and the packet length is $L = 1024$ bits [24] [10]. The estimate complex uplink channel gain, $\hat{\alpha}_{nsm}$, is generated by the complex uplink channel gain, $\alpha_{nsm} = d_{nsm}^{-\chi} h_{nsm}$, minus the complex Gaussian random variable, $\Delta\alpha_{nsm}$, where d_{nsm} is calculated by the distance between MT m and network n BS/AP s, the path loss exponent χ is 4, and the fast fading process $|h_{nsm}|$ follows the Rayleigh distribution with the scale parameter 1. The other simulation parameters are $\delta = 0.01$, $P_c = 20$ dBm, and $n_0 = -174$ dBm/Hz [13]. Simulation results present the confidence interval for a 0.95 confidence level. We aim to evaluate the performance of the ODBPA and SDBPA compared with Benchmark algorithm and sum rate maximization (SRM) algorithm [41]. Benchmark algorithm is fixed bandwidth and power allocation, which allocates the equal bandwidth among different MTs and the equal power across different radio interfaces for each MT. The simulation results are averaged over 500 runs, and each run is operated based on a snapshot of the simulation scenario to capture the position change of each MT. The position of each MT is generated randomly at the beginning of each time slot, which is like the existing resource allocation algorithms [8], [9], [11].

Fig. 2 shows the impact of the maximum packet delay on the maximum consumed energy per bit for different algorithms. The simulation parameters are listed in Table I. With a lower maximum packet delay, more energy per bit is consumed for both ODBPA and SDBPA due to the higher required transmission rate for each MT. As shown in Fig. 2, the proposed ODBPA and SDBPA outperform the Benchmark and SRM. This is mainly due to the fact that the proposed ODBPA and SDBPA allocate different bandwidth among different MTs, and different power across different radio interfaces at each MT, to minimize the maximum consumed energy per bit. Benchmark allocates the bandwidth and power equally among different MTs and different radio interfaces, and SRM allocates the bandwidth and power to maximize the throughput. The optimization objective for ODBPA and SDBPA minimizes the maximal energy efficiency among different MTs, while that for SRM maximizes the total network throughput. Since increasing

---

[10] In order to analyze the influence of video traffic at the link layer to the required transmission rate at the physical layer conveniently, we adopt a simple video traffic model, i.e., the video packet arrivals at the transmission buffer of MT m follow a Poisson process with an average arrival rate λ_m [23] with a constant packet length of L bits [6].

TABLE I
SIMULATION PARAMETERS

Simulation Parameter	Value
Simulation Tool	Matlab 9.0
Confidence interval	0.95
Circumradius of macrocell	500
Circumradius of femtocell	20
Number of macrocell MTs	10
Numbers of femtocell MTs	3
Macrocell available bandwidth	30 MHz
Femtocell available bandwidth	5 MHz
Maximum power of femtocell MT	0.5 W
Maximum power of macrocell MT	3 W
γ_m	0.01
σ^2_{nsm}	0.03
ε_{nsm}	0.1
χ	4
L	1024 bit
P_c	20 dBm
n_0	-174 dBm/Hz
λ_m	1.5×10^4 packets/s

the maximum packet delay reduces the required minimum transmission rate, that means more bandwidth and power resources are allocated to maximize the total throughput for SRM, not to improve energy efficiency among different MTs. Additionally, maximizing the total throughput for SRM is at the cost of energy efficiency, and the maximum consumed energy per bit for SRM increases with the required maximum packet delay. When the maximum packet delay is 0.2 ms, the maximum consumed energy per bit for SDBPA exceeds 16.95% compared with that for ODBPA.

Fig. 3 depicts how the maximum consumed energy per bit for different algorithms changes with the average packet arrival rate. The maximum packet delay for MT m is $D_m^{\max} = 1$ ms. The other simulation parameters in Fig. 3 are the same with Table I. As the average packet arrival rate increases, the maximum energy efficiency among different MTs for the ODBPA is reduced due to the fact that more power is needed to guarantee the packet delay requirement. Compared with the Benchmark and SRM, the maximum consumed energy per bit for the ODBPA and SDBPA
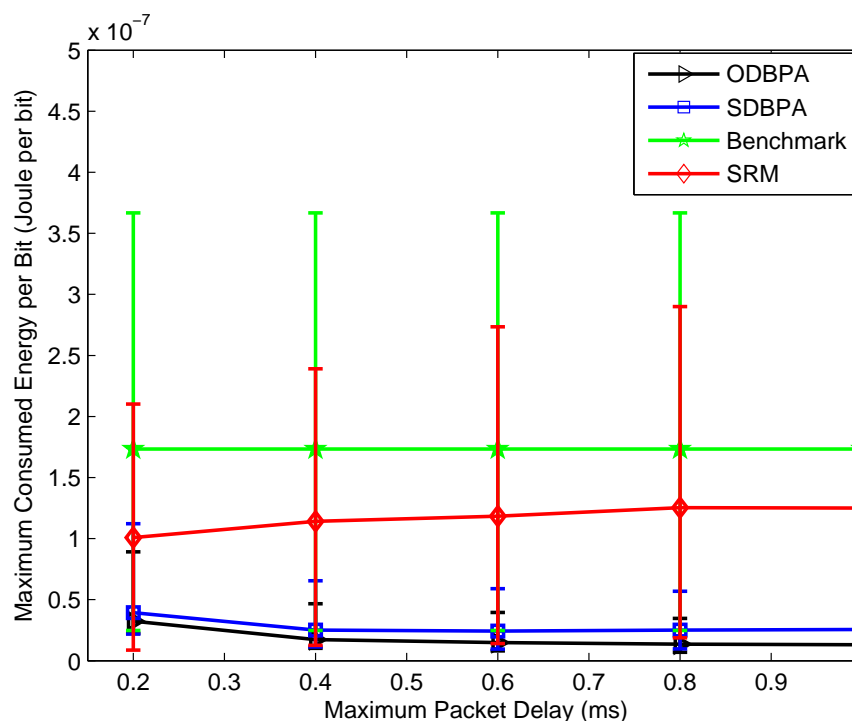
Fig. 2.   Maximum consumed energy per bit vs. the maximum packet delay.

is reduced. In Benchmark, the radio bandwidth and transmission power are allocated without considering the QSI at the buffer. This results in the unchanged maximum energy efficiency when increasing the average packet arrival rate. The ODBPA has an improved performance over the SDBPA due to the fact that the ODBPA jointly allocates the bandwidth and power, whereas the bandwidth and power allocations in the SDBPA are designed separately. For SDBPA, increasing the minimum transmission rate for each MT can make the bandwidth and power resources to be allocated fairly among different MTs for SDBPA. This leads to the maximization energy efficiency among different MTs for SDBPA to be improved firstly. When the average packet arrival rate exceeds a threshold, the maximization energy efficiency among different MTs for SDBPA decreases. Consequently, the maximum consumed energy per bit for SDBPA increases. It is also observed that the minimum energy efficiency for SRM increases with the average packet arrival rate, because the fact that increasing the average packet arrival rate means increasing the required minimum transmission rate and improves the minimum energy efficiency. When the
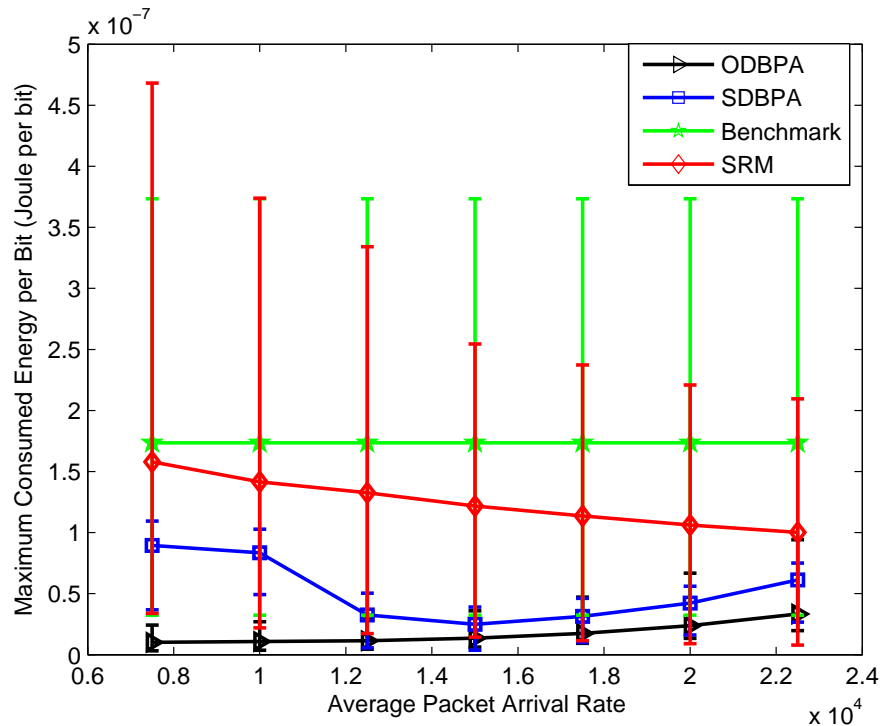
Fig. 3. Maximum consumed energy per bit vs. the average packet arrival rate.

average packet arrival rate is 1.5×10^4 packets/s, the maximum consumed energy per bit for SRM exceeds 477.25% compared with that for ODBPA.

Fig. 4 evaluates the impact of the total available bandwidth at the macro BS on the maximum consumed energy per bit for the different algorithms. The average packet arrival rate for MT m is $\lambda_m = 1 \times 10^4$ packets/s. The maximum packet delay for MT m is $D_m^{\max} = 1$ ms. The other simulation parameters in Fig. 4 are the same with Table I. The maximum consumed energy per bit decreases with the increased total available bandwidth at macro BS, and the energy efficiency for different algorithms improves. This is because macrocell MTs and femtocell MTs use the bandwidth at macro BS, increasing total available bandwidth resource at macro BS can reduce the power consumption for macrocell MTs and femtocell MTs. When the total available bandwidth at the macro BS is 20 MHz, the maximum consumed energy per bit for SRM exceeds 386.77% compared with that for SDBPA.

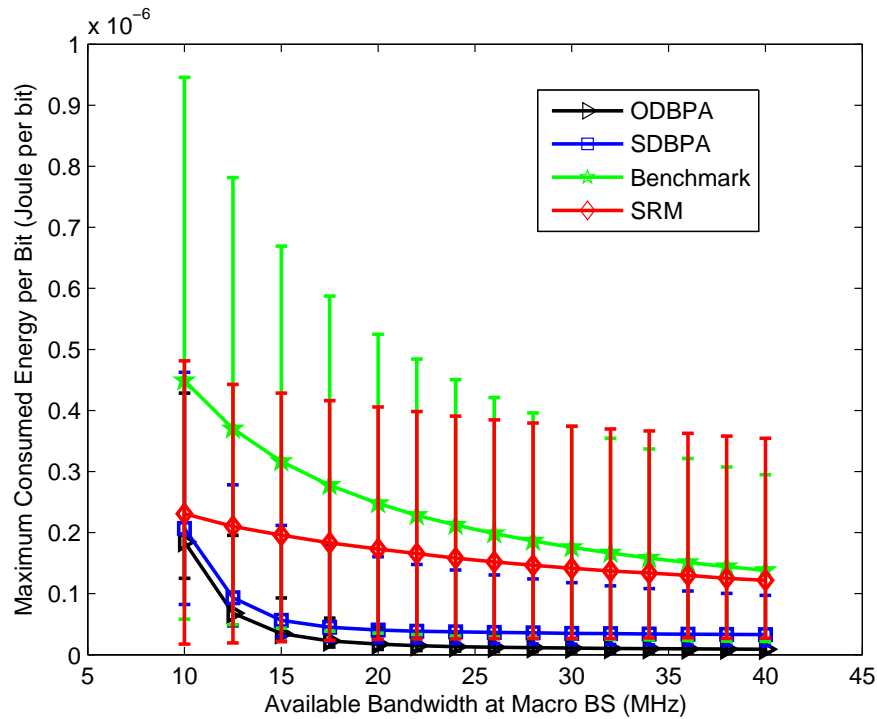Fig. 5 shows the effect of number of microcell BSs/APs on the maximum consumed energy

Fig. 4.   Maximum consumed energy per bit vs. the total available bandwidth at macro BS.

per bit for the algorithms. In the geographical region, there are one macrocell, two femtocells, and several microcells. The average packet arrival rate at MTs is $\lambda_m = 1.5 \times 10^4$ packets/s. The maximum packet delay for MT m is $D_m^{\max} = 1$ ms. Since each microcell has limited resources and can not satisfy the QoS requirement of MTs, each microcell needs 2 MHz bandwidth resource from the macrocell. The maximum consumed energy per bit for four algorithms increases with the number of microcell BSs/APs. This is due to the fact that increasing the number of microcell BSs/APs means more heavy traffic in heterogeneous wireless networks. This can leads to the decreasing energy efficiency.

Fig. 6 shows the effect of number of femtocell MTs on the maximum consumed energy per bit for the algorithms. The average packet arrival rate for MT m is $\lambda_m = 1.25 \times 10^4$ packets/s, and the maximum power of each MT at the macrocell and femtocell are both 5 W. The maximum packet delay for MT m is $D_m^{\max} = 1$ ms. The other simulation parameters in Fig. 5 are the same with Table I. The maximum consumed energy per bit for SDBPA, Benchmark, and SRM
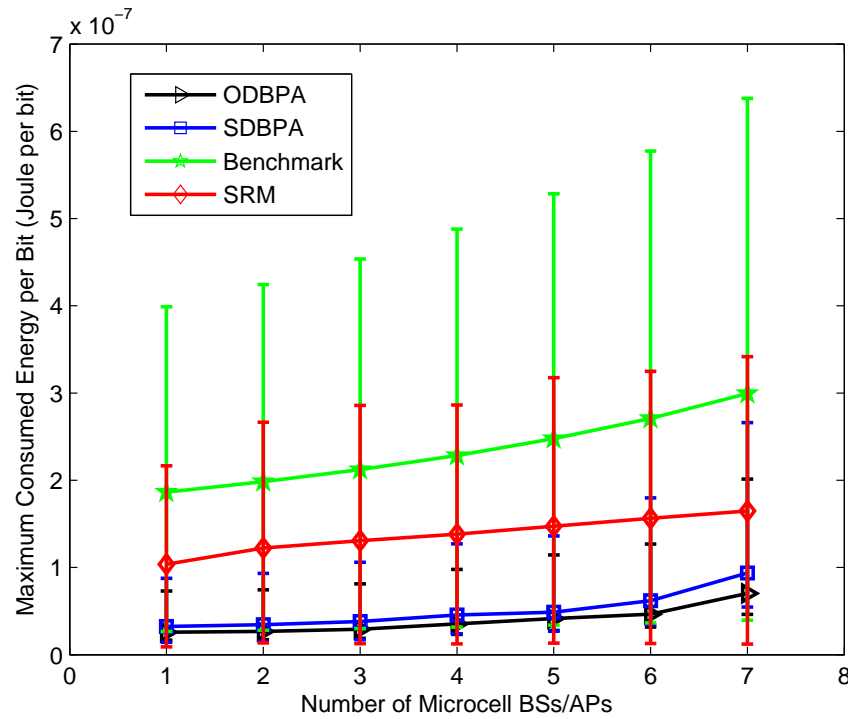
Fig. 5. Maximum consumed energy per bit vs. the number of microcell BSs/APs.

increases with the number of femtocell MTs. The maximum consumed energy per bit for ODBPA decreases first, and then increases with the number of femtocell MTs. When the number of femtocell MTs is small, the bandwidth and power resource is sufficient, and multiuser diversity can improve the energy efficiency slightly. However, the QoS requirement becomes more strict with the increasing number of MTs, which dominates the multiuser diversity effect.

From Fig. 2 to Fig. 6, it can be concluded that the proposed algorithms improve the energy efficiency as compared with the Benchmark and SRM. Although the SDBPA has some energy efficiency loss compared with the ODBPA, it achieves a tradeoff between system performance and complexity.

V. Conclusions

In this paper, we study the uplink energy-efficient cross-layer resource allocation problem for a heterogeneous wireless access. Each MT adjusts radio bandwidth and power based on the
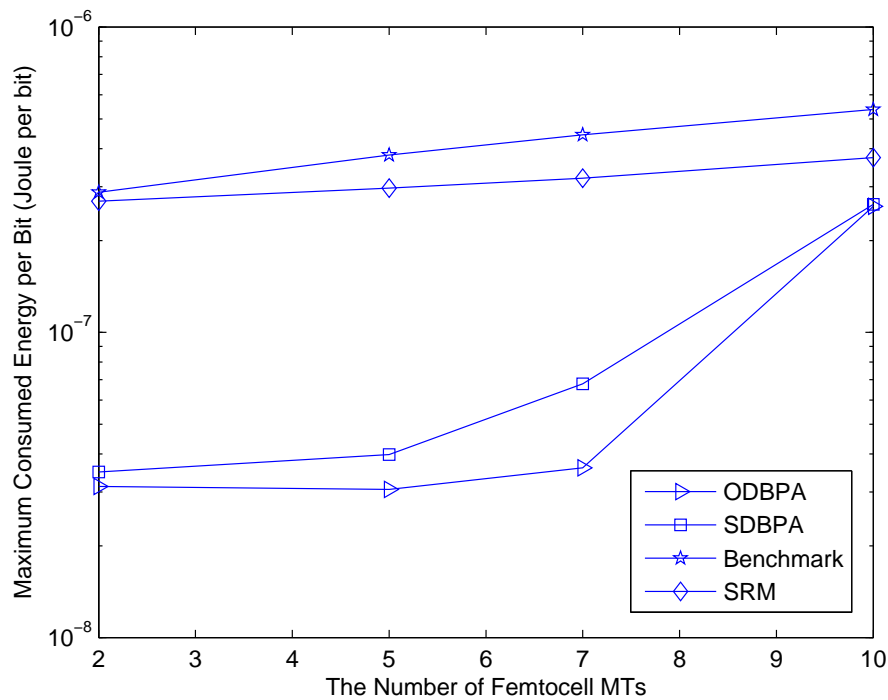
Fig. 6. Maximum consumed energy per bit vs. the number of femtocell MTs.

channel condition and the queue buffer occupancy, so as to minimize the maximum consumed energy per bit under the QoS constraints. In order to solve the above resource allocation problem, we model the energy-efficient bandwidth and power allocation problem as min-max fractional stochastic programming, and analyze the packet delay. Then, the Dinkelbach-type and dual decomposition methods are utilized to design the ODBPA, and the SDBPA is proposed to reduce the computational complexity. Simulation results demonstrate proposed algorithms can improve the energy efficiency significantly.

APPENDIX A PROOF OF PROPOSITION 1

In order to prove that (17) is a bi-convex optimization problem, we prove the convexity of the constraints and the objective function in terms of decision variables B_{nsm} and P_{nsm}, respectively. Define the objective function and the constraints in (17) as the functions g_1, f_1, f_2, f_3, and f_4,

i.e.,

$$\begin{cases} g_1 = \vartheta \\ f_1 = \vartheta - P_m + \xi R_m \\ f_2 = B_{ns} - \sum_{m \in \mathcal{M}_{ns}} B_{nsm} \\ f_3 = P_m^T - P_m \\ f_4 = R_m - \psi\left(D_m^{\max}, \gamma_m, \lambda_m, L\right) \end{cases} \tag{32}$$

With the fixed variable P_{nsm}, the second derivatives of g_1, f_1, f_2, f_3 and f_4 with respective to B_{nsm} are

$$\begin{cases} \frac{\partial^2 g_1}{\partial (B_{nsm})^2} = 0 \\ \frac{\partial^2 f_1}{\partial (B_{nsm})^2} = -\frac{\xi (SNR_{nsm})^2}{(1+SNR_{nsm})^2 B_{nsm} \ln 2} \le 0 \\ \frac{\partial^2 f_2}{\partial (B_{nsm})^2} = 0 \\ \frac{\partial^2 f_3}{\partial (B_{nsm})^2} = 0 \\ \frac{\partial^2 f_4}{\partial (B_{nsm})^2} = -\frac{(SNR_{nsm})^2}{(1+SNR_{nsm})^2 B_{nsm} \ln 2} \le 0 \end{cases} \tag{33}$$

Given the variable P_{nsm}, the objective function is concave on B_{nsm}, and the constraints are concave according to (33). Hence, (17) is a convex programming problem with the variable B_{nsm}.

On the other hand, with the fixed variable B_{nsm}, the second derivatives of g_1, f_1, f_2, f_3 and f_4 with respective to P_{nsm} are

$$\begin{cases} \frac{\partial^2 g_1}{\partial (P_{nsm})^2} = 0 \\ \frac{\partial^2 f_1}{\partial (P_{nsm})^2} = -\frac{\xi \left[F_{|\alpha_{nsm}|}^{-1}\left(\varepsilon_{nsm}, |\hat\alpha_{nsm}|, \sigma_{nsm}^2\right) \right]^4}{(1+SNR_{nsm})^2 (n_0)^2 B_{nsm}} \le 0 \\ \frac{\partial^2 f_2}{\partial (P_{nsm})^2} = 0 \\ \frac{\partial^2 f_3}{\partial (P_{nsm})^2} = 0 \\ \frac{\partial^2 f_4}{\partial (P_{nsm})^2} = -\frac{\left[F_{|\alpha_{nsm}|}^{-1}\left(\varepsilon_{nsm}, |\hat\alpha_{nsm}|, \sigma_{nsm}^2\right) \right]^4}{(1+SNR_{nsm})^2 (n_0)^2 B_{nsm}} \le 0 \end{cases} \tag{34}$$

Given the variable B_{nsm}, the objective function is concave on P_{nsm}, and the constraints are concave according to (34). Hence, (17) is a convex programming problem with the variable

P_{nsm}. Consequently, problem (17) is a bi-convex optimization problem.

APPENDIX B SIGNALING OVERHEAD ANALYSIS

In the ODBPA, each MT broadcasts its $F_m(\xi)$, $u_m(i)$, and B_{nsm} to all serving BSs/APs. The BSs/APs broadcast their $v_{ns}(i+1)$ to all MTs, and exchange the information regarding $F_m(\xi)$ and $u_m(i)$. The signaling overhead is $O\left(D_I S_I O_I \left(\sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} M_{ns} + \sum_{n \in \mathcal{N}} S_n + 4M\right)\right)$, where D_I is the number of iterations for the convergence of the Dinkelbach-type method, S_I is the number of iterations for the convergence of binary search method, and O_I is the number of iterations for the convergence of the ODBPA. In the SDBPA, each BS/AP broadcasts to notice the selected MT, each MT broadcasts its η_m, P_{nsm}, B_{nsm}, and R_m to all serving BSs/APs, and the BSs/APs exchange the information regarding η_m. The overall signaling overhead is $O\left((B_I + P_I)\left(8M + \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} M_{ns}\right)\right)$, where B_I and P_I are the number of iterations for the bandwidth and power allocation, respectively.

APPENDIX C COMPUTATIONAL COMPLEXITY ANALYSIS

In the ODBPA, the computational complexity is determined by solving the dual problem [13], and the computational complexity of the ODBPA is given by $O\left(D_I S_I O_I M^2 \sum_{n \in \mathcal{N}} S_n\right)$. In the SDBPA, the computational complexity is determined by the bandwidth and power allocation. The computational complexities for bandwidth and power allocation are $O\left(B_I \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} M_{ns}\right)$ and $O\left(P_I \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} M_{ns}\right)$, respectively. The overall computational complexity of the SDBPA is $O\left((P_I + B_I) \sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}_n} M_{ns}\right)$.

ACKNOWLEDGMENTS

REFERENCES

[1] K. Son, S. Lee, Y. Yi, and S. Chong, "Refim: A practical interference management in heterogeneous wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1260–1272, Jun. 2011.

[2] M. Ismail, A. Abdrabou, and W. Zhuang, "Cooperative decentralized resource allocation in heterogeneous wireless access medium," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 714–724, Feb. 2013.

[3] L. Xu, P. Wang, Q. Li, and Y. Jiang, "Call admission control with inter-network cooperation for cognitive heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1963–1973, Mar. 2017.

[4] H. Chen, X. Ding, Z. Wang, and L. Xie, "A rate allocation scheme for multi-user over heterogeneous wireless access networks," in *Proc. 2010 IEEE VTC*, Sep. 2010, pp. 1–5.

[5] S. Wei and Q. Zhu, "Efficient and fair bandwidth allocation for multiuser multimedia communication over heterogeneous networks," in *Proc. 2013, IEEE CISP*, Dec. 2013, pp. 16–20.

[6] M. Ismail and W. Zhuang, "Mobile terminal energy management for sustainable multi-homing video transmission," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4616–4627, Aug. 2014.

[7] J. Wu, C. Yuen, N. M. Cheung, and J. Chen, "Delay-constrained high definition video transmission in heterogeneous wireless networks with multi-homed terminals," *IEEE Trans. Mobile Comput.*, vol. 15, no. 3, pp. 641–655, Mar. 2016.

[8] Q. D. Vu, L. N. Tran, M. Juntti, and E. K. Hong, "Energy-efficient bandwidth and power allocation for multi-homing networks," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1684–1699, Apr. 2015.

[9] S. Kim, B. G. Lee, and D. Park, "Energy-per-bit minimized radio resource allocation in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1862–1873, Apr. 2014.

[10] Y. Choi, H. Kim, S. wook Han, and Y. Han, "Joint resource allocation for parallel multi-radio access in heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3324–3329, Nov. 2010.

[11] P. Xue, P. Gong, J. H. Park, D. Park, and D. K. Kim, "Radio resource management with proportional rate constraint in the heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1066–1075, Mar. 2012.

[12] J. Miao, Z. Hu, C. Wang, R. Lian, and H. Tian, "Optimal resource allocation for multi-access in heterogeneous wireless networks," in *Proc. 2012 IEEE VTC*, May. 2012, pp. 1–5.

[13] M. Ismail, A. Gamage, W. Zhuang, X. Shen, E. Serpedin, and K. Qaraqe, "Uplink decentralized joint bandwidth and power allocation for energy-efficient operation in a heterogeneous wireless medium," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1483–1495, Apr. 2015.

[14] C. C. Zarakovitis, Q. Ni, D. E. Skordoulis, and M. G. Hadjinicolaou, "Power-efficient cross-layer design for OFDMA systems with heterogeneous QoS, imperfect CSI, and outage considerations," *IEEE Trans. Veh. Tech.*, vol. 61, no. 2, pp. 781–798, Feb. 2012.

[15] X. Zhu, B. Yang, C. Chen, L. Xue, X. Guan, and F. Wu, "Cross-layer scheduling for OFDMA-based cognitive radio systems with delay and security constraints," *IEEE Trans. Veh. Tech.*, vol. 64, no. 12, pp. 5919–5934, Dec. 2015.

[16] L. Xu and A. Nallanathan, "Energy-efficient chance-constrained resource allocation for multicast cognitive OFDM network," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1298–1306, May. 2016.

[17] L. Xu, A. Nallanathan, and X. Song, "Joint video packet scheduling, subchannel assignment and power allocation for cognitive heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1703–1712, Mar. 2017.

[18] L. Xu, N. Arumugam, and Y. Yang, "Video packet scheduling with stochastic QoS for cognitive heterogeneous networks," *IEEE Trans. Veh. Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[19] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcasting*, vol. 61, no. 2, pp. 166–176, Jun. 2015.

[20] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Trans. Broadcasting*, vol. 62, no. 3, pp. 675–684, Sep. 2016.

[21] Z. Pan, P. Jin, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast reference frame selection based on content similarity for low complexity HEVC encoder," *Journal of Visual Communication and Image Representation*, vol. 40, no. B, pp. 516–524, Oct. 2016.

[22] L. Perez, X. Chu, A. V. Vasilakos, and H. Claussen, "Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 333–344, Feb. 2014.

[23] S. Tanwir and H. Perros, "A survey of VBR video traffic models," *IEEE Commun. Surveys Tutorials*, vol. 15, no. 4, pp. 1778–1802, Nov. 2013.

[24] N. Mokari, M. R. Javan, and K. Navaie, "Cross-layer resource allocation in OFDMA systems for heterogeneous traffic with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 1011–1017, Feb. 2010.

[25] G. Saleh, A. El-Keyi, and M. Nafie, "Cross-layer minimum-delay scheduling and maximum-throughput resource allocation for multiuser cognitive networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 4, pp. 761–773, Apr. 2013.

[26] B. Gu and V. S. Sheng, "A robust regularization path algorithm for nu -support vector classification," *IEEE Trans. Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1241–1248, May. 2017.

[27] M. J. F. G. Garcia, J. L. Rojo-Alvarez, F. Alonso-Atienza, and M. Martinez-Ramon, "Support vector machines for robust channel estimation in OFDM," *IEEE Signal Processing Lett.*, vol. 13, no. 7, pp. 397–400, Jul. 2006.

[28] B. Gu, X. Sun, and V. S. Sheng, "Structural minimax probability machine," *IEEE Trans. Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–11, 2016.

[29] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.

[30] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, Apr. 2015.

[31] I. Ngebani, Y. Li, X. G. Xia, and M. Zhao, "EM-based phase noise estimation in vector OFDM systems using linear MMSE receivers," *IEEE Trans. Veh. Tech.*, vol. 65, no. 1, pp. 110–122, Jan. 2016.

[32] Q. Du and X. Zhang, "QoS-aware base-station selections for distributed MIMO links in broadband wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1123–1138, Jun. 2011.

[33] Y. Wang, P. Ren, Q. Du, and L. Sun, "Optimal power allocation for underlay-based cognitive radio networks with primary user's statistical delay QoS provisioning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6896–6910, Dec. 2015.

[34] S. E. Ghoreishi and A. H. Aghvami, "Power-efficient QoE-aware video adaptation and resource allocation for delay-constrained streaming over downlink OFDMA," *IEEE Commun. Letters*, vol. 20, no. 3, pp. 574–577, Mar. 2016.

[35] J. F. Hayes and G. Babu., *Modeling and Analysis of Telecommunications Networks*. Wiley, 2004.

[36] J. M. T. C. M. H. Donald Gross, John F. Shortie, *Fundamentals of Queueing Theory*. Wiley, 2008.

[37] E. Bedeer, A. Alorainy, M. J. Hossain, O. Amin, and M. S. Alouini, "Fairness-aware energy-efficient resource allocation for AF co-operative OFDMA networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2478–2493, Dec. 2015.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[39] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Math Meth. Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.

[40] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Commun. Surveys Tutorials*, vol. 14, no. 2, pp. 322–337, May. 2012.

[41] S. Kim, B. G. Lee, and D. Park, "Energy-per-bit minimized radio resource allocation in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1862–1873, Apr. 2014.