Energy Efficient Multiuser Scheduling: Statistical Guarantees on Bursty Packet Loss

M. Majid Butt[*], Eduard A. Jorswieck[†] and Amr Mohamed[*]
[*]Department of Computer Science and Engineering, Qatar University, Qatar
Email: {majid.butt, amrm}@qu.edu.qa
[†]Department of Electrical Engineering and Information Technology, Dresden University of Technology, Germany
Email: eduard.jorswieck@tu-dresden.de

Abstract—In this paper, we consider energy efficient multiuser scheduling. Packet loss tolerance of the applications is exploited to minimize average system energy. There is a constraint on average packet drop rate and maximum number of packets dropped successively (bursty loss). A finite buffer size is assumed. We propose a scheme which schedules the users opportunistically according to the channel conditions, packet loss constraints and buffer size parameters. We assume imperfect channel state information at the transmitter side and analyze the scheme in large user limit using stochastic optimization techniques.

First, we optimize system energy for a fixed buffer size which results in a corresponding statistical guarantee on successive packet drop. Then, we determine the minimum buffer size to achieve a target (improved) energy efficiency for the same (or better) statistical guarantee. We show that buffer size can be traded effectively to achieve system energy efficiency for target statistical guarantees on packet loss parameters.

Index Terms—Energy efficiency, opportunistic scheduling, stochastic optimization, simulated annealing.

I. Introduction

Radio resource allocation in wireless networks is becoming more complex due to stringent constraints imposed by Quality of Service (QoS) requirements. The QoS requirements vary from application to application in terms of average/maximum packet delay and average packet loss rate. Energy efficiency is another important factor in today's wireless networks design due to increasing cost of network operation. The radio resource algorithms should aim at utilizing every soft QoS requirement to improve the energy efficiency of the network, e.g., the authors in [1] investigate power-delay tradeoff and propose resource allocation schemes to minimize power consumption subject to a delay QoS constraint where the delay constraint is in terms of queue-length decay rate. Energy-performance tradeoffs have been addressed in different settings in [2], [3]. Most of the work focuses on exploiting delay tolerance to optimize the system energy. However, data loss tolerance is another aspect that can be exploited to save system energy.

We consider a multiple access system where the users have soft average packet drop constraint and a hard constraint on the maximum number of packet allowed to be dropped *successively*. This constraint is referred to as continuity constraint [4]. The work in [4], [5] takes this constraint into account for packet scheduling. The idea is to allow intentional (but bounded) packet drop to save energy if application can tolerate it without a major deterioration in quality of experience. The

authors assume a perfect channel state information (CSI) on the transmitter and receiver sides. A channel threshold based scheduling scheme is introduced and analyzed. However, perfect CSI condition is hard to achieve in practical networks due to limitation on feedback information. This paper extends the work for imperfect CSI case and addresses the resulting effects on the system energy and scheduling process. The effect of imperfect channel information is modeled through a corresponding packet drop probability.

We have two reasons for packet drop in our problem settings:

1) Intentional packet drop at the transmitter depending on the application loss tolerance to save energy if applications's loss tolerance permits.
2) Packet drop due to imperfect CSI at the transmitter side which implies that channel state is worse than the estimated one and results in packet loss after transmission.

The scheduling algorithm design for the packet loss tolerant application should take the packet loss due to imperfect CSI into account and adapt its intentional packet drop rate accordingly to maintain a bounded average packet drop rate. The main contribution of this work is to evaluate this effect through packet level channel model. Moreover, we extend the results to the (more practical) case of statistical guarantees on continuity constraint as compared to hard guarantees in [5].

The rest of the paper is structured as follows. Section II introduces the system model and preliminaries. We model the proposed scheme using Markov chain in Section III. The optimization problem is formulated in Section IV. We evaluate the scheme numerically in Section V and conclude with the main contributions in Section VI.

II. System Model and Preliminaries

We follow the system model used in [4], [6] and consider a multiple-access system with K users randomly placed within a certain area. Every scheduled user requires an average rate $R_k = \frac{C}{K}$ where C denotes the system spectral efficiency.

A. Propagation Channel Model

We consider an uplink scenario where time is slotted such that each user k experiences a channel gain $h_k(t)$ in a time slot t. The channel gain $h_k(t)$ comprises of path loss component s_k and small-scale fading $f_k(t)$ such that $h_k(t) = s_k f_k(t)$. The

path loss is a function of the distance between the transmitter and the receiver and remains constant within the time scale considered in this work. Small-scale fading depends on the scattering environment. It changes from slot to slot for every user and is independent and identically distributed (i.i.d) across both users and slots, but remains constant during the time span of a single time slot. The multi-access channel is described by the input (X) and output (Y) relation as

$$Y_k(t) = \sum_{k=1}^{K} \sqrt{h_k(t)} X_k(t) + Z(t) \qquad (1)$$

where Z represents additive i.i.d. complex Gaussian random variable with zero mean and unit variance. The distribution of $h_k(t)$ differs from user to user.

B. Packet level channel Model

The CSI is assumed to be known at the transmitter, but it is not perfect. As a result of imperfect CSI, the transmitter is not able to compute correct power level for the assigned rate. This could result in packet loss. We model the effect of imperfect CSI by a probability ν_d that a resulting transmission is not successful. Furthermore, we assume if the estimated channel was not good enough to support the rate, all the packets transmitted in a single transmission are lost. The information about the dropping of the packet(s) is conveyed by the receiver to the transmitter through a perfect feedback channel. This is termed as packet level channel modeling in the literature. We assume that the feedback arrives at the transmitter by the next scheduling instance. The unsuccessfully transmitted packet(s) is buffered for (possible) retransmission if the buffer has the capacity to store it for the next time slot, dropped otherwise.

C. Statistical Guarantees on Continuity Constraint

The model considered in [5] assumes that continuity constraint can be met with probability one. It is not practicable to assume that a packet can be transmitted with probability one when N packets have been dropped successively; where N is termed as continuity constraint parameter. We extend our framework in the direction of providing statistical guarantees on continuity constraint, i.e., a user violates the continuity constraint with a probability γ. If channel conditions are not good after dropping N packets successively, the user is still allowed to drop a finite amount of packets corresponding to γ. We define the event of violation of continuity constraint as the number of time slots a packet is dropped after successively dropping N packets already.

We allow multiple users to be scheduled in a single time slot to minimize γ. If only a single user is scheduled per time slot, all the users other than the scheduled one may have to drop the packets (intentionally) which results in increase in γ rapidly. We have no control over the packets dropped due to channel impairments, but packet scheduler can be designed such that γ is bounded by facilitating maximum scheduling of the users who already have dropped N packets successively.

The analysis of the scheme is based on asymptotic user case which implies that there is no limit on the number of users scheduled simultaneously. We use superposition coding and successive interference cancelation (SIC) mechanism for successful transmission of data streams of simultaneously scheduled users. Let $\mathcal{K}$ denote the set of users to be scheduled and Φ_k be the permutation of the scheduled user indices that sorts the channel gains in increasing order, i.e. $h_{\Phi_1} \leq \cdots \leq h_{\Phi_k} \leq \cdots \leq h_{\Phi_{|\mathcal{K}|}}$. Then, the energy of the scheduled user Φ_k with rate R_{Φ_k}, is given by [6], [7]

$$E_{\Phi_k} = \frac{Z_0}{h_{\Phi_k}} \left(2^{\sum_{i \leq k} R_{\Phi_i}} - 2^{\sum_{i < k} R_{\Phi_i}} \right). \qquad (2)$$

where Z_0 denotes the noise power spectral density.

III. ANALYSIS OF THE SCHEDULING SCHEME

In this section, we briefly review the scheduling scheme presented in [5] for perfect CSI scenario. Then, we model the scheme for the imperfect CSI case and analyze the effect on the scheduling decisions and average system energy.

A. Review of the Scheme with Perfect CSI

We consider a constant arrival of a single packet in the buffer of each user in each time slot for simplicity. However, this model can be extended to the random arrival case where multiple packets arriving in the same time slot are treated as a single packet [4]. Every arriving packet is queued in the user's buffer after arrival.

The design of the scheme is based on the asymptotic case when $K \to \infty$. In this case, multiuser scheduling problem can be broken into a single user scheduling problem such that every user takes the scheduling decision independent of the other users [4]. The scheduling decisions for every user in each time slot are based on the instantaneous channel condition and the scheduling thresholds; which are optimized by taking into consideration the continuity constraint parameter N, maximum buffer size B, average packet dropping probability θ_{tar} and user's small scale fading distribution. The number of thresholds equals the number of buffered packets and the scheduler decides how many packets are scheduled in a single time slot based on the channel conditions. If no packet is scheduled, all the packets (including the recently arrived packet) are buffered if buffer size allows. If the buffer is full, the oldest packet in the buffer is dropped. When the user has dropped N packets successively, the head of line (HOL) packet has to be scheduled regardless of the channel conditions.

B. Finite State Markov Chain Model

We extend the results for the scheduling scheme proposed in [5] to the case of imperfect CSI. As explained in Section II-B, the effect of imperfect CSI is modeled through packet level description such that ν_d denotes packet drop probability and $\nu_s = 1 - \nu_d$ is the probability of a successful transmission.

We model the scheme discussed in Sec. III-A using a finite state Markov chain (FSMC). Let $i \leq B$ and $j \leq N$ denote the number of packets buffered and dropped successively at time t. Then, state p at time t is defined by the summation

$i + j$. At the start of the process, p equals zero. If a packet is not scheduled, it is buffered and $i = 1$ (while $j = 0$), thereby the system makes transition to next state $q = 1$. Remember $p(t + 1) = q(t)$ in FSMC. When the buffer is full, an event of not scheduling a packet results in a packet drop, thereby j starts increasing and $i = B$ remains fixed until there is a room in the buffer for unscheduled packets due to scheduling of previously buffered packets. The event of dropping/buffering of the packet results in forward state transition to next state $q = p + 1$. The size of FSMC is determined by the buffer size and continuity constraint parameters such that $M = B + N$.

We did not consider the event of packet drop due to imperfect CSI in the state space description yet. As we assume that feedback for the successful/unsuccessful transmission (ACK/NACK) arrives by the end of time slot, the system buffers the scheduled packet(s) by the end of time slot. If the transmitter receives an ACK, the packets are dropped from the buffer as they have been received successfully. In case of a NACK, the buffered packets are treated in the same way as intentional packet dropping, i.e. buffer if there is a room or drop otherwise. In contrast to the case of ACK, dropping a packet in case of a NACK occurs solely due to insufficient buffer capacity and affects system performance similar to intentional packet drop scenario. The packet drop due to imperfect CSI needs to be modeled in the system separately due to its different effect on system energy. Intentional packet dropping (without transmission) does not cost any energy to the system while packets dropped due to imperfect CSI result in waste of energy without transmitting any data successfully.

In an FSMC model, we define α_{pq} as

$$\alpha_{pq} = \Pr(S_{t+1} = q | S_t = p) = \begin{cases} \hat{\alpha}_{pq} & \forall p, q \leq \min(p, B) \\ \tilde{\alpha}_{pq} & \forall p, q = p + 1 \\ 0 & \text{else} \end{cases} \quad (3)$$

where

$$\begin{aligned} \alpha_{pq} &= \text{Transition probability from state } p \text{ to } q. \\ \hat{\alpha}_{pq} &= \text{Transition probability from state } p \text{ to } q \\ &\quad \text{when scheduling of one or more packets occurs.} \\ \tilde{\alpha}_{pq} &= \text{Transition probability from state } p \text{ to } q \\ &\quad \text{when no packet is scheduled.} \end{aligned}$$

α_{pq} is a function of $\hat{\alpha}_{pq}$ and $\tilde{\alpha}_{pq}$.

To define $\hat{\alpha}_{pq}$ and $\tilde{\alpha}_{pq}$, we define a scheduling threshold.

Definition 1 (Scheduling Threshold κ_{pq}): It is defined as the minimum small scale fading value f required to make a state transition from state p to q such that

$$\hat{\alpha}_{pq} = \Pr\big(\kappa_{pq} < f \leq \kappa_{p(q-1)}\big) \quad 0 \leq q \leq \min(p, B). \quad (4)$$

where κ_{p0-} is defined to be infinity with S_{0-} denoting a dummy state before S_0.

From scheduling point of view, it is advantageous to schedule more packets for good fading states. Therefore, the scheduling thresholds quantize the fading vector to optimize the number

of scheduled packets according to the fading. In a state $p \geq q$, the scheduler makes a state transition to state q such that [5]

$$q = \arg\min_{\acute{q}} \kappa_{p\acute{q}} < f \leq \kappa_{p(\acute{q}-1)} \quad 0 \leq \acute{q} \leq \min(p, B). \quad (5)$$

For a state transition from state p to q, the number of the scheduled packets is given by

$$L(p, f) = \min(p, B) - q + 1, \quad (6)$$

where q is determined uniquely by (5). Note that the number of scheduled packets cannot exceed $\min(p, B)$ because of finite capacity of buffer. We denote $\min(p, B)$ by μ in the rest of this article for convenience. The probability of not scheduling any packet for transmission is given by

$$\tilde{\alpha}_{pq} = \Pr(f \leq \kappa_{p\mu}), \quad 0 \leq p < M, q = p + 1 \quad (7)$$

$$= 1 - \sum_{q=0}^{\mu} \hat{\alpha}_{pq}. \quad (8)$$

where $\kappa_{p\mu}$ denotes the minimum thresholds to schedule at least one packet.

C. Modeling γ in FSMC

Ideally one would like to schedule a packet with probability one when $p = M$. As explained earlier, it is not practical to apply 'water filling' principle on any arbitrary channel due to power limitations of the transmitter. Thus, a packet is not scheduled if fading is worse than a minimum value even in state M which contributes to γ in addition to packet dropping due to imperfect CSI. To handle the event of unscheduled or/and lost HOL packet in state M, we define a self transition α_{MM} where no packet is scheduled in contrast to other self state transitions.

According to our FSMC model,

$$\gamma = \alpha_{MM} \pi_M = \left(\tilde{\alpha}_{MM} + \nu_d \sum_{q=0}^{B} \hat{\alpha}_{Mq} \right) \pi_M \quad (9)$$

$$= \left(1 - \nu_s \sum_{q=0}^{B} \hat{\alpha}_{Mq} \right) \pi_M \quad (10)$$

where π_M is steady state transition probability for state M. Thus, α_{pq} for any states p, q is modified as

$$\alpha_{pq} = \nu_s \hat{\alpha}_{pq}, \quad 0 \leq p \leq M, 0 \leq q \leq \mu \quad (11)$$

$$\alpha_{pq} = \tilde{\alpha}_{pq} + \nu_d \sum_{m=0}^{\mu} \hat{\alpha}_{pm}, \quad 0 \leq p < M, q = p + 1 \quad (12)$$

Example 1: Let us explain FSMC model with the help of an example with $B = 2$, $N = 1$ as in Fig. 1. For this example, we evaluate the transition probability matrix $\boldsymbol{Q}$.

For state 0, α_{00} is the probability that a packet is scheduled for transmission and received successfully.

$$\alpha_{00} = \nu_s \hat{\alpha}_{00} \quad (13)$$

Transition from state 0 to 1 is the result of un-scheduled; and scheduled but unsuccessful transmission events. Thus,

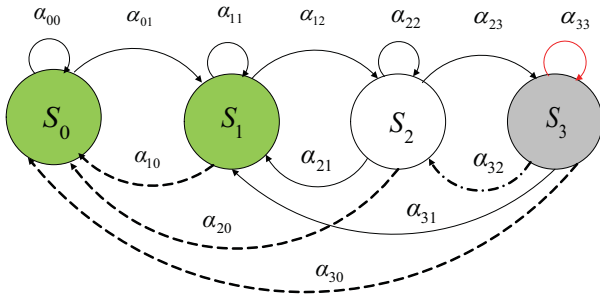$$\alpha_{01} = \tilde{\alpha}_{01} + \nu_d \hat{\alpha}_{00} \quad (14)$$

Fig. 1. State transition diagram of the scheme for the case $B = 2, N = 1$. α_{MM} represents state transition probability related to γ.

Similarly, for state 1,

$$\alpha_{1q} = \nu_s \hat{\alpha}_{1q} \qquad q = 0, 1 \tag{15}$$

$$\alpha_{12} = \tilde{\alpha}_{12} + \nu_d(\hat{\alpha}_{10} + \hat{\alpha}_{11}) \tag{16}$$

Following the same line of arguments, the matrix Q can be written as a summation of two matrices such that

$$Q = Q_s + Q_c \tag{17}$$

where

$$Q = \begin{pmatrix} \alpha_{00} & \alpha_{01} & 0 & 0 \\ \alpha_{10} & \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{20} & \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{30} & \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \tag{18}$$

$$Q_s = \begin{pmatrix} \nu_s \hat{\alpha}_{00} & \tilde{\alpha}_{01} & 0 & 0 \\ \nu_s \hat{\alpha}_{10} & \nu_s \hat{\alpha}_{11} & \tilde{\alpha}_{12} & 0 \\ \nu_s \hat{\alpha}_{20} & \nu_s \hat{\alpha}_{21} & \nu_s \hat{\alpha}_{22} & \tilde{\alpha}_{23} \\ \nu_s \hat{\alpha}_{30} & \nu_s \hat{\alpha}_{31} & \nu_s \hat{\alpha}_{32} & \tilde{\alpha}_{33} \end{pmatrix} \tag{19}$$

and

$$Q_c = \nu_d \begin{pmatrix} 0 & \sum_{q=0}^{0} \hat{\alpha}_{0q} & 0 & 0 \\ 0 & 0 & \sum_{q=0}^{1} \hat{\alpha}_{1q} & 0 \\ 0 & 0 & 0 & \sum_{q=0}^{2} \hat{\alpha}_{2q} \\ 0 & 0 & 0 & \sum_{q=0}^{2} \hat{\alpha}_{3q} \end{pmatrix}. \tag{20}$$

Q_c captures the effect of imperfect CSI while Q_s is optimized scheduling decision matrix. Note that this model implies that it is not possible to achieve continuity constraint with probability one if $\nu_d > 0$ and only statistical guarantees can be provided.

IV. Mathematical Formulation of the Problem

The objective of the optimization problem is to minimize the system energy for a soft average packet drop rate constraint and statistical guarantee on continuity constraint. We formulate the optimization problem using the FSMC model. Each scheduled packet is treated as a virtual user for the analysis purpose. The average system energy per transmitted information bit at the large system limit $K \to \infty$ is given by [6]

$$\frac{E_b}{N_0} = \log(2) \int_0^\infty \frac{2^{C \, P_{h,\mathrm{VU}}(x)}}{x} \, dP_{h,\mathrm{VU}}(x) \tag{21}$$

where $P_{h,\mathrm{VU}}(\cdot)$ denotes the cumulative distribution function (cdf) of the fading of the scheduled virtual users (VUs). In the large system limit, the state transitions depend only on the small scale fading distribution as the path loss for VUs follows the same distribution as the path loss of the users. Thus, the optimization problem[1] is formulated as

$$\min_{Q \in \Omega} \frac{E_b}{N_0} \tag{22}$$

$$\text{s.t. :} \begin{cases} \mathcal{C}_1 : 0 \leq \sum_{m=0}^{\mu} \alpha_{pm} \leq 1 & 0 \leq \alpha_{pm} \leq 1, \\ & \qquad 0 \leq p \leq M \\ \mathcal{C}_2 : \theta_r \leq \theta_{\mathrm{tar}} & Q \in \Omega \\ \mathcal{C}_3 : \sum_{q=0}^{M} \alpha_{pq} = 1 & 0 \leq p \leq M \\ \mathcal{C}_4 : B + N = M & B < \infty, N < \infty \end{cases} \tag{23}$$

where Ω denotes the set of permissible matrices for Q and θ_r is the average packet drop rate for a fixed Q and given by

$$\theta_r = \sum_{p=B}^{M-1} \alpha_{p(p+1)} \pi_p + \alpha_{MM} \pi_M \tag{24}$$

$$= \sum_{p=B}^{M} \left(1 - \nu_s \sum_{m=0}^{B} \hat{\alpha}_{pm}\right) \pi_p . \tag{25}$$

Equation (25) is a result of combining $\mathcal{C}_1$ and $\mathcal{C}_3$ in (23). The forward transition for the states $B \leq p < M$ and self state transition α_{MM} represent the events of packet drop and the summation over the corresponding transition probabilities gives the average dropping probability in (25). The summation starts from state B as the unscheduled packets are buffered for $p < B$. For a fixed p, the corresponding channel-dependent optimal scheduling thresholds can be computed from the optimized $\vec{\alpha}_p^* = [\alpha_{p0}^*, \dots \alpha_{p\mu}^*]$ using (4). The violation probability on continuity constraint γ for fixed B and N parameters is computed from Q^* using (10). Let us denote γ for this special case by γ_m where the *maximum* energy efficiency can be achieved for fixed $B, N, \theta_{\mathrm{tar}}$ parameters and relaxing γ further does not help to improve energy efficiency due to coupling of γ with N and θ_{tar}.

If the statistical guarantees has to be improved further, we apply an upper bound on γ such that $\gamma \leq \epsilon$ where ϵ is a small constant representing the target statistical guarantee. This constraint appears as an additional constraint in (23) such that

$$\mathcal{C}_5 : \gamma \leq \epsilon \quad 0 \leq \epsilon \leq \theta_{\mathrm{tar}} \tag{26}$$

because $\theta_r = \sum_{p=B}^{M-1} \alpha_{p(p+1)} \pi_p + \gamma$. Consequently, the improved γ is achieved at the increased energy cost. Theoretically ϵ is upper bounded by θ_{tar}; but γ_m upper bounds ϵ (tightly) at a value lower than θ_{tar} due to the reasons explained above.

The probability density function (pdf) of the small scale fading of the scheduled VUs is given by

$$p_{f,\mathrm{VU}}(y) = \sum_{p=0}^{M} c_p \pi_p L(p, y) \, p_f(y) \tag{27}$$

[1]This section follows the work in [5] closely, but developments of the next section require its inclusion for completeness and clarity of the discussion.

where $\mathrm{p}_f(y)$ and c_p denote the small scale fading distribution and a normalization constant respectively while $L(p, y)$ is given by (6). The derivation of the cdf of VUs can be found in [5]. The channel distribution for the scheduled VUs can be computed using fading and the path loss distributions which in turns is used to compute system energy in (21).

A. Trading Buffer for Improved Guarantees on γ

We would like to achieve $\epsilon \leq \gamma_m$ at improved energy by increasing the value of B for a fixed N. Let us denote the optimal solution of the programming problem in previous section by $Q^*(B, \theta_{\mathrm{tar}}, \epsilon)$ as a function of B, θ_{tar} and target violation probability on continuity constraint ϵ. Let $\frac{E_b}{N_0}(Q^*(B, \theta_{\mathrm{tar}}, \epsilon))$ be the corresponding system energy and ΔE represents the target energy gain. Now, the optimization is performed over $B \in \Phi$ where Φ is a set of possible buffer sizes. For every candidate $B \in \Phi$, optimization in (22) and (23) is performed again with inclusion of $\mathcal{C}_5$. The aim of the optimization is to find minimum value of B which gives energy less than $\left(\frac{E_b}{N_0}(Q^*(B, \theta_{\mathrm{tar}}, \epsilon)) - \Delta E \right)$ at ϵ:

$$\text{Find } B^* \in \Phi \quad \text{s.t.} \quad \gamma(Q^*(B^*, \theta_{\mathrm{tar}})) \leq \epsilon \quad \text{and} \quad (28)$$
$$\frac{E_b}{N_0}(Q^*(B^*, \theta_{\mathrm{tar}}, \epsilon)) - \frac{E_b}{N_0}(Q^*(B, \theta_{\mathrm{tar}}, \epsilon)) \geq \Delta E, \quad B \in \Phi$$

The suitable value of B is highly dependent on the application. For example, wireless sensor networks would prefer large B due to battery requirements whereas multimedia applications prefer small B due to stringent delay requirements on data delivery.

B. Stochastic Optimization

The optimization problem formulated in (22) and (23) is not convex and belongs to class of problems called stochastic optimization problems. There are a few heuristic techniques in literature to solve such problems like genetic algorithm, Q-learning, neural networks, etc. We use Simulated Annealing (SA) algorithm to solve the problem. As the name suggests, the algorithm originates from statistical mechanics area and has been found quite useful to solve different combinatorial optimization problems like traveling salesman.

In SA algorithm, a random configuration in terms of transition probability matrix Q is presented in each step and system energy as an objective function is evaluated only if Q fulfills all the constraints in (23). If system energy improves the previous best solution, the candidate configuration is selected as the best available solution. However, a candidate configuration can be treated as best solution with a certain temperature dependent probability even if the new solution is worse than the best known solution. This step is called *muting* and helps the system to avoid local minima. The muting step occurs frequently at the start of the process as temperature is selected very high and decrease as temperature is decreased gradually. Thus, the term temperature determines the rate of muting process. In literature, different cooling temperature schedules have been employed according to the problem requirements. In this work, we employ the following cooling schedule, called

TABLE I
γ_m AND SYSTEM ENERGY

N	γ_m	E_b/N_0
1	0.09	-3.63 dB
2	0.032	-3.63 dB
3	0.014	-3.61 dB

fast annealing (FA) [8]. In FA, it is sufficient to decrease the temperature linearly in each step b such that,

$$T_b = \frac{T_0}{c_{\mathrm{sa}} * b + 1} \quad (29)$$

where T_0 is a suitable starting temperature and c_{sa} is a constant which depends on the requirements of the problem. We skip details of the SA scheme due to space limitations. The interested reader is referred to [9] for details of the algorithm.

V. NUMERICAL RESULTS

We assume that the users are placed uniformly at random in a circular cell except for a forbidden region around the access point of radius $\delta = 0.01$. The path loss exponent equals 2 and the path loss distribution follows the model in [6]. All the users experience independent small-scale fading with exponential distribution with mean one. Spectral efficiency is 0.5 bits/s/Hz for all simulations. In SA algorithm, 100 temperature values are simulated according to FA temperature schedule while $50(M + 1)$ random configuration of transition probability matrix are generated for a single temperature iteration.

To compute γ_m, we perform optimization in (23) without applying constraint in (26) and the *best*[2] solution matrix Q^* is obtained. The value of γ computed via (10) for Q^* gives us γ_m and upper bounds ϵ. Table I shows numerical values of system energy and γ_m for different N and fixed $B = 0, \theta_{\mathrm{tar}} = 0.3$ values while ν_d equals 0.02. Based on numerical results in Table I, we evaluate the tradeoffs addressed in Section IV-A.

Fig. 2 demonstrates the effect of imposing constraint $\epsilon \leq \gamma_m$ on system performance when $\theta_{\mathrm{tar}} = 0.3$. We evaluate $\mathcal{C}_5$ alongwith $\mathcal{C}_1 - \mathcal{C}_4$ in (23) for the candidate Q before evaluation of (21) in SA algorithm. We observe in Fig. 2(a) that decreasing ϵ has an associated energy cost and the solution becomes suboptimal by energy point of view. Also, γ can never approach zero as long as $\nu_d > 0$ and packet dropping due to imperfect CSI cannot be completely eliminated. For a given set of parameters and fixed ν_d, the minimum value of achievable ϵ is denoted by γ_0 which lower bounds ϵ such that $\gamma_0 \leq \epsilon \leq \gamma_m$. The greater the value of ν_d, the greater is γ_0. For instance, increasing ν_d from 0.02 to 0.1 for the case $N = 2$ raises γ_m from 0.001 to 0.002 while system energy increases for all values of ϵ as well. We observe that bounds on ϵ (in the form of γ_0 and γ_m) become tighter as N increases for the fixed θ_{tar}. This is due to the fact that allowing large N increases degrees of freedom (DoF) for the system and the effect of parameter ϵ on system energy is minimized.

[2]We avoid using term energy optimal here as SA is a heuristic algorithm and solution cannot be proven optimal.

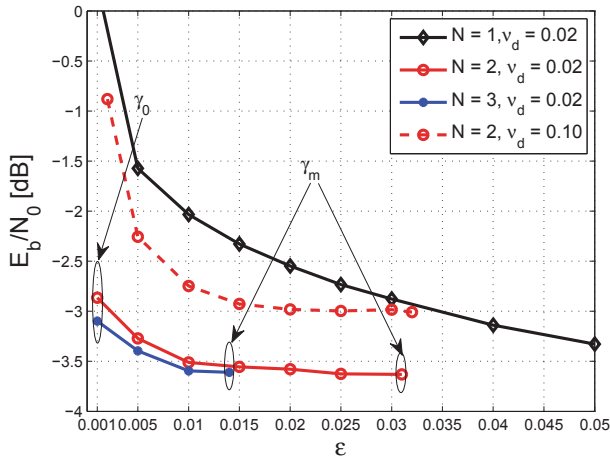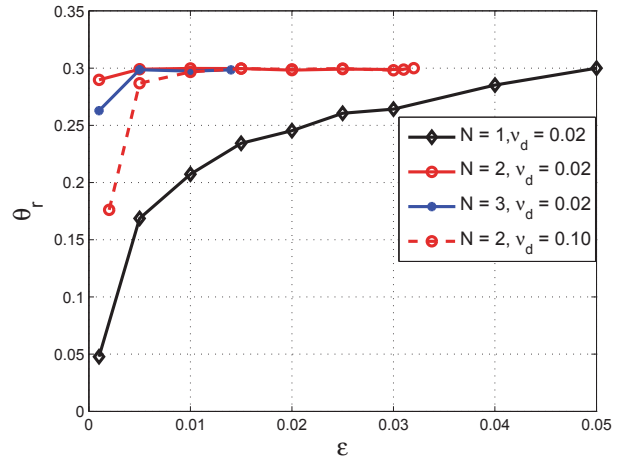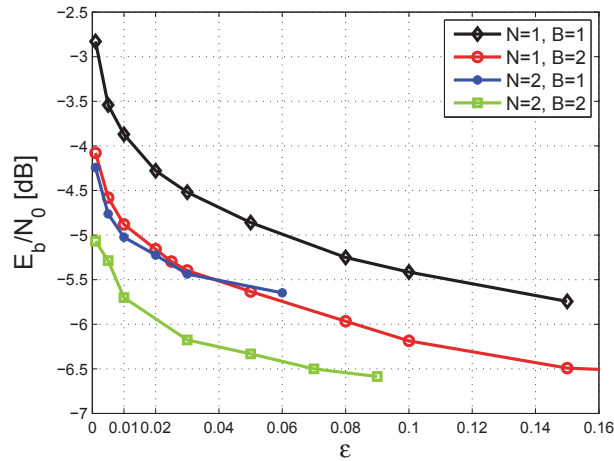(a) System energy as a function of ϵ for a system with fixed $B = 0$.

(b) Achieved packet drop rate θ_r from (25) as a function of ϵ for the same parameters as in Fig. 2(a).

(c) System energy as a function of ϵ when $B > 0$. All other parameters are the same as in Fig. 2(a)

Fig. 2. System energy and packet drop behavior as a function of ϵ. θ_{tar} is fixed to 0.3 for all simulations.

Correspondingly, Fig. 2(b) demonstrates that achieved average packet drop rate θ_r (calculated via (25)) approaches θ_{tar} for large ϵ and remains almost identical thereafter. This implies that all the extra energy cost is contributed by strict guarantees on continuity constraint. When ϵ is very small, the energy optimal Q^* provides a θ_r which is much less that θ_{tar} and severely sub optimal. We conclude that a strict statistical guarantee on continuity constraint has a severe plenty in terms of energy and even other DoF (like relaxed θ_{tar}) cannot be utilized efficiently.

Fig. 2(c) demonstrates the energy gain achieved by increasing buffer size as described in Section IV-A. First, we observe that increasing the value of B for a fixed N increases γ_m, i.e. more flexibility in ϵ. Secondly, an energy gain by increasing B for all ϵ and a fixed N is evident. It depends on the system design that which B needs to be employed for a particular performance guarantee. Let us discuss the case for parameters $N = 1, \theta_{\mathrm{tar}} = 0.3, \epsilon = 0.01$. The system with $B = 0$ provides system energy of almost -2 dB as shown in Fig. 2(a). If we want the same performance at reduced energy, $B = 1$ provides a gain of $\Delta E = 1.9$ dB. If $\Delta E > 1.9$ dB, $B > 1$ is required. For the same set of parameters, $B = 2$ provides ΔE equal to 3.1 dB. A similar comparison can be drawn for $N = 2$ and $B > 0$.

A comparison of the curves for the cases $N = 2, B = 1$ and $N = 1, B = 2$ (with same $M = 3$) shows that increasing DOF in any parameter (B, N) is energy efficient as compared to the case $N = 1, B = 1$ but the effect differs widely in many ways, e.g., value of γ_m for both cases. Similarly, increasing B to reduce system energy affects system cost while increasing N costs performance loss in terms of jitter. Thus, system's energy, packet loss and latency requirements

determine the parameters required to achieve performance in terms of statistical guarantee on continuity constraint.

VI. Conclusions

We consider energy efficient multiuser scheduling over fading channels for packet loss tolerant applications. Packet loss is modeled by an average packet drop rate and *continuity constraint* on successive dropping of packets. The proposed scheme is analyzed for imperfect CSI case using packet level channel model where the effect of imperfect CSI is modeled by fixed packet drop and success probabilities. We formulate the optimization problem for achieving minimum system energy for a target statistical guarantee on continuity constraint. Through stochastic optimization framework, we characterize the limits on achievable statistical guarantees on continuity constraint. We evaluate the effect of buffer size on the problem settings and validate numerically that buffer size can be traded to achieve better energy efficiency for a given statistical guarantee on continuity constraint and average packet drop rate. We conclude that application's energy and latency requirements are important to determine preferable buffer size to achieve system performance in terms of protection to bursty packet loss.

References

[1] X. Zhang and J. Tang, "Power-delay tradeoff over wireless networks," *IEEE Trans. Wireless Communications*, vol. 61, no. 9, pp. 3673–3684, 2013.

[2] M. M. Butt, "Energy-performance trade-offs in multiuser scheduling: Large system analysis," *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 217–220, June 2012.

[3] M. J. Neely, "Optimal energy and delay tradeoffs for multiuser wireless downlinks," *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3095–3113, September 2007.

[4] M. M. Butt and E. A. Jorswieck, "Maximizing system energy efficiency by exploiting multiuser diversity and loss tolerance of the applications," *IEEE Trans. Wireless Communications*, vol. 12, no. 9, pp. 4392–4401, 2013.

[5] ——, "Maximizing energy efficiency for loss tolerant applications: The packet buffering case," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, London, UK, Sep 2013.

[6] G. Caire, R. Müller, and R. Knopp, "Hard fairness versus proportional fairness in wireless communications: The single-cell case," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1366–1385, April 2007.

[7] D. Tse and S. Hanly, "Multi-access fading channels-part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.

[8] H. Szu and R. Hartley, "Fast simulated annealing," *Physics Letters A*, vol. 122, no. 3, 1987.

[9] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.