Optimal Wireless Networks Based on Local Channel State Information

Yichuan Hu, *Student Member, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

***Abstract*—This paper considers distributed algorithms to optimize random access multihop wireless networks in the presence of fading. Since the associated optimization problem is neither convex nor amenable to distributed implementation, a problem approximation is introduced. This approximation is still not convex but it has zero duality gap and can be solved and decomposed into local subproblems in the dual domain. The solution method is through a stochastic subgradient descent algorithm that operates without knowledge of the fading's probability distribution and leads to an architecture composed of layers and layer interfaces. With limited amount of message passing among terminals and small computational cost, the proposed algorithm converges almost surely in an ergodic sense. Numerical results on a randomly generated network corroborate theoretical results.**

***Index Terms*—Cross-layer design, random access, wireless networking.**

I. Introduction

OPTIMAL design is emerging as the future paradigm for wireless networking. The fundamental idea is to select operating points as solutions of optimization problems, which, inasmuch as optimization criteria are properly chosen, yield the best possible network. Results in this field include architectural insights, e.g., [1], and protocol design, e.g., [2] and [3], but a drawback shared by most of these works is that they rely on global channel state information (CSI); i.e., the optimal variables of a terminal depend on the channels between all pairs of terminals in the network. While availability of global CSI is plausible in certain situations, it is unlikely to hold if time varying fading channels are taken into account.

We consider optimal design of wireless networks in the more practical situation where, due to the presence of random fading, only local CSI is available. This restriction implies that operating variables of each terminal are selected as functions of the channels linking the terminal with neighboring nodes and further leads to the selection of random access as the natural medium access choice. Indeed, if transmission decisions depend on local channels only and these channels are random and independent for different terminals, transmission decisions can be viewed as random and resultant link capacities as limited by collisions. Thus, we can restate our goal as the development of algorithms to find optimal operating points of wireless random access networks in the presence of fading. Operating points are characterized by external arrival rates, routes, link capacities, average power consumptions, instantaneous channel access decisions, and power allocations. Our goal is to select these variables to be optimal in terms of ergodic averages.

Optimal design of multihop random access networks has been considered in [4]–[10]. Assuming that capacities of links in the network are fixed and that terminals transmit with certain probabilities without coordination, these works focus on computing terminal transmission probabilities that are optimal in some sense. For example, distributed algorithms are proposed in [4] and [5] for achieving proportionally fair utility, and in [6]–[8] for general utility functions. To reduce algorithm complexity and increase convergence speed, several enhancements are discussed in [9] and [10]. However, optimization across fading states is not considered in any of these works.

Adapting transmission decisions to random fading states has been considered in the particular case of random multiple access protocols [11]–[18]. In this case it is known that a threshold-based policy in which terminals transmit when their channels exceed a threshold and stay silent otherwise is optimal. This was originally proved for simple collision models [11], and later extended to other scenarios with different packet reception assumptions [12]–[18]. Since these works consider single hop wireless networks they do not apply directly to the multihop wireless fading networks considered here. An existing approach to optimal multihop random access is [19] where threshold-based policies are applied in multihop random access networks. Our work differs from [19] in that i) While routes are fixed in [19] we consider them as variables to be optimized and ii) while terminals in [19] are assumed to have access to the channels' probability distributions, we develop online algorithms that operate without this prior knowledge.

This paper builds on recent results showing that non-convex wireless networking optimization problems have null duality gap as long as the probability distributions of underlying fading channels have no points of strictly positive measure [20]. Given this result it is possible to develop stochastic subgradient descent algorithms in the dual domain that have been proven optimal in an ergodic almost sure sense [21]. While global CSI is assumed available in general wireless networking problems considered in [20] and [21], our goal is to find solutions for a specific scenario where only local CSI is available and random access is used at the physical layer. To do so we begin by introducing an optimization problem that defines the optimal random access network (Section II). Since

Manuscript received May 18, 2011; revised February 17, 2012; accepted May 04, 2012. Date of publication June 01, 2012; date of current version August 07, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David Love. This work was supported by ARO W911NF-10-1-0388 and NSF CAREER CCF-0952867. Part of the results in this paper were presented at the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, May 22–27, 2011, and the Allerton Conference on Communication, Control and Computing (Allerton), Monticello, IL, September 28–30, 2011.

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: yichuan@seas.upenn.edu; aribeiro@seas.upenn.edu).

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

Digital Object Identifier 10.1109/TSP.2012.2202110

this problem is not amenable to distributed implementation we proceed to a suboptimal approximation through a problem that while still not convex has zero duality gap [20] (Section II-B). We further observe that solution is simpler in the dual domain—and equivalent because of the lack of duality gap—and proceed to develop stochastic dual descent algorithms that converge to the optimal operating point (Section III). The resultant algorithm decomposes in a layered architecture and is computationally tractable in that iterations require a few simple algebraic operations (Section III-B). We also explain a decentralized implementation based on information exchanges with neighboring terminals (Section III-C). Results on ergodic stochastic optimization from [21] are finally leveraged to show that the proposed algorithm yields operating points that are almost surely close to optimal (Section IV). Numerical results and concluding remarks are presented in Sections V and VI.

II. Problem Formulation

Consider an ad-hoc wireless network consisting of J terminals indexed as $i = 1, \dots J$. Network connectivity is modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $v \in \mathcal{V} := \{1, \dots, J\}$ representing the J terminals and edges $e = (i,j) \in \mathcal{E}$ connecting pairs of terminals that can communicate with each other. Denote the neighborhood of terminal i as $\mathcal{N}(i) := \{j \,|\, (i,j) \in \mathcal{E}\}$ and define the interference neighborhood of the link (i,j) as the set of nodes $\mathcal{M}_i(j) := \mathcal{N}(j) \cup \{j\} \backslash \{i\}$ whose transmission can interfere with a transmission from i to j. The network supports a set $\mathcal{K} := \{1, \dots, K\}$ of end-to-end flows through multihop transmission. The average rate at which k-flow packets are generated at i is denoted by a_i^k. Terminal i transmits these packets to neighboring terminals at average rates r_{ij}^k and, consequently, receives k-flow packets from neighbors at average rates r_{ji}^k. To conserve flow, exogenous rates a_i^k and endogenous rates r_{ij}^k at terminal i must satisfy

$$a_i^k \leq \sum_{j \in \mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k \right), \quad \text{for all } i \in \mathcal{V}, \text{ and } k \in \mathcal{K}. \tag{1}$$

Further denote the capacity of the link from $i \to j$ as c_{ij}. Since packets of different flows k are transmitted from i to j at rates r_{ij}^k it must be

$$\sum_{k \in \mathcal{K}} r_{ij}^k \leq c_{ij}, \quad \text{for all } (i,j) \in \mathcal{E}. \tag{2}$$

Unlike wireline networks where c_{ij} are fixed, link capacities in wireless networks are dynamic. Let time be divided into slots indexed by t and denote the channel between i and j at time t as $h_{ij}(t)$. The channel is assumed to be block fading—for this to be true the length of a time slot has to be comparable to the coherence time of the channel. As a result, $h_{ij}(t)$ remains constant within a time slot and changes randomly in subsequent time slots. Channel gains $h_{ij}(t)$ of link (i,j) are assumed independent and identically distributed with probability distribution function (pdf) $m_{h_{ij}}(\,\cdot\,)$. We assume no channel realization has nonzero probability, something that is true for models used in practice ([22], Chapter 3). For reference, define the vector of terminal i outgoing channels $\mathbf{h}_i(t) := \{h_{ij}(t) \,|\, j \in \mathcal{N}(i)\}$ and the vector of all channels $\mathbf{h}(t) := \{h_{ij}(t) \,|\, (i,j) \in \mathcal{E}\}$. Denote their pdfs as $m_{\mathbf{h}_i}(\,\cdot\,)$ and $m_{\mathbf{h}}(\,\cdot\,)$, respectively.

Based on the channel state $\mathbf{h}_i(t)$ of his outgoing links, terminal i decides whether to transmit or not on link (i,j) in time slot t by determining the value of a scheduling function $q_{ij}(t) := Q_{ij}(\mathbf{h}_i(t)) \in \{0,1\}$. If $q_{ij}(t) = 1$, terminal i transmits on link (i,j) and remains silent otherwise. Further define $q_i(t) := Q_i(\mathbf{h}_i(t)) := \sum_{j \in \mathcal{N}(i)} Q_{ij}(\mathbf{h}_i(t))$ to indicate a transmission from i to any of his neighbors. We restrict i to communicate with, at most, one neighbor per time slot implying that we must have $q_i(t) \in \{0,1\}$. We emphasize that $q_{ij}(t) := Q_{ij}(\mathbf{h}_i(t))$ depends on local outgoing channels only and not on global CSI. Further note that terminals have access to instantaneous local CSI $\mathbf{h}_i(t)$ but underlying pdfs $m_{\mathbf{h}_i}(\,\cdot\,)$ are unknown.

Besides channel access decisions, terminals also adapt transmission power to local CSI through a power control function $p_{ij}(t) := P_{ij}(\mathbf{h}_i(t))$ taking values in $[0, p_{ij}^{\max}]$. Here, $p_{ij}^{\max}$ represents the maximum allowable instantaneous power on link (i,j). The average power consumed by terminal i is then given as the expected value over channel realizations of the sum of $P_{ij}(\mathbf{h}_i)$ over all $j \in \mathcal{N}(i)$, i.e.,

$$p_i \geq \mathbb{E}_{\mathbf{h}_i}\left[\sum_{j \in \mathcal{N}(i)} P_{ij}(\mathbf{h}_i) Q_{ij}(\mathbf{h}_i) \right] \tag{3}$$

where we also relaxed the equality constraint to an inequality, which can be done without loss of optimality. If terminal i transmits to node j in time slot t, $p_{ij}(t)$ and $h_{ij}(t)$ determine the transmission rate through a function $C_{ij}(h_{ij}(t)p_{ij}(t))$ whose form depends on modulation and coding. To keep analysis general, we do not restrict $C_{ij}(h_{ij}(t)p_{ij}(t))$ to a specific form. We just assume that it is a nonnegative increasing function of the signal to noise ratio (SNR) $h_{ij}(t)p_{ij}(t)$ taking finite values for finite arguments. This restriction is lax enough to allow for discontinuous rate functions that arise in, e.g., adaptive modulation and coding.

Due to contention, a transmission from i to j at time t succeeds if a collision does not occur. In turn, this happens if i) terminal i transmits to j, i.e., $q_{ij}(t) = 1$; ii) terminal j is silent, i.e., $q_j(t) = 0$; iii) no other neighbor of j transmits, i.e., $q_l(t) = 0$ for all $l \in \mathcal{N}(j)$ and $l \neq i$. Recalling the definition of interference neighborhood $\mathcal{M}_i(j)$ and that if a transmission occurs its rate is $C_{ij}(h_{ij}(t)p_{ij}(t))$ we express the instantaneous transmission rate from i to j in time slot t as $c_{ij}(t) := c_{ij}(\mathbf{h}_i(t)) = C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t) \prod_{l \in \mathcal{M}_i(j)} [1 - q_l(t)]$. Assuming an ergodic mode of operation, the capacity of link $i \to j$ can then be written as

$$c_{ij} = \mathbb{E}_{\mathbf{h}}\left[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i)) Q_{ij}(\mathbf{h}_i) \prod_{l \in \mathcal{M}_i(j)} [1 - Q_l(\mathbf{h}_l)] \right]. \tag{4}$$

Because terminals are required to make channel access and power control decisions independently of each other, $Q_{ij}(\mathbf{h}_i)$ and $P_{ij}(\mathbf{h}_i)$ are independent of $Q_{lm}(\mathbf{h}_l)$ and $P_{lm}(\mathbf{h}_l)$ for all $i \neq l$. Since $Q_l(\mathbf{h}_l) := \sum_{m \in \mathcal{N}(l)} Q_{lm}(\mathbf{h}_l(t))$ by definition, it follows that $Q_{ij}(\mathbf{h}_i)$ is also independent of $Q_l(\mathbf{h}_l)$ for all $i \neq l$. This allows us to write the expectation of the product on the right-hand side of (4) as a product of expectations,

$$c_{ij} \leq \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i)) Q_{ij}(\mathbf{h}_i) \right] \times \prod_{l \in \mathcal{M}_i(j)} [1 - \mathbb{E}_{\mathbf{h}_l}[Q_l(\mathbf{h}_l)]] \tag{5}$$

where we also relaxed the equality constraint to an inequality, which can be done without loss of optimality.[1]

The operating point of a wireless network is characterized by variables a_i^k, r_{ij}^k, c_{ij}, p_i and functions $P_{ij}(\mathbf{h}_i)$, $Q_{ij}(\mathbf{h}_i)$. Besides, (1)–(3) and (5) these variables are subject to certain box constraints. Admission variables, have lower and upper bounds due to application layer requirements, i.e., $a_i^{\min} \le a_i^k \le a_i^{\max}$. Similarly, routing variables, link capacities, and terminal power budgets cannot be negative and are also subject to given upper bounds, i.e., $0 \le r_{ij}^k \le r_{ij}^{\max}$, $0 \le c_{ij} \le c_{ij}^{\max}$, and $0 \le p_i \le p_i^{\max}$. Furthermore, according to definition, $P_{ij}(\mathbf{h}_i)$ and $Q_{ij}(\mathbf{h}_i)$ can only take values from $[0, p_{ij}^{\max}]$ and $\{0,1\}$, respectively. For notational simplicity, we define vectors $\mathbf{x}_i := \{p_i, a_{ij}^k, r_{ij}^k, c_{ij} : \forall j \in \mathcal{N}(i)\}$ and $\mathbf{P}_i(\mathbf{h}_i) := \{P_{ij}(\mathbf{h}_i), Q_{ij}(\mathbf{h}_i) : \forall j \in \mathcal{N}(i)\}$ to group all the variables related to terminal i and summarize these box constraints as $\{\mathbf{x}_i, \mathbf{P}_i(\mathbf{h}_i)\} \in \mathcal{B}_i$ with

$$\begin{aligned}\mathcal{B}_i := \big\{\mathbf{x}_i, \mathbf{P}_i(\mathbf{h}_i) \,|\, & a_i^{\min} \le a_i^k \le a_i^{\max}, 0 \le r_{ij}^k \le r_{ij}^{\max}, \\ & 0 \le c_{ij} \le c_{ij}^{\max}, 0 \le p_i \le p_i^{\max}, 0 \le P_{ij}(\mathbf{h}_i) \le p_{ij}^{\max}, \\ & Q_{ij}(\mathbf{h}_i) \in \{0,1\}, Q_i(\mathbf{h}_i) \in \{0,1\}\big\}. \qquad (6)\end{aligned}$$

A. Optimal Operating Point

As network designers, we wish to find the optimal operating point of the wireless network defined as a set of variables a_i^k, r_{ij}^k, c_{ij}, p_i and functions $Q_{ij}(\mathbf{h}_i)$, $P_{ij}(\mathbf{h}_i)$ that satisfy constraints (1)–(3), (5), and (6) and are optimal according to certain criteria. In particular, we are interested in large rates a_i^k and low power consumptions p_i. Define then increasing concave functions $U_i^k(\cdot)$ representing rewards for accepting a_i^k units of information for flow k at terminal i and increasing convex functions $V_i(\cdot)$ typifying penalties for consuming p_i units of power at i. The optimal network based on local CSI is then defined as the solution of

$$\begin{aligned}\mathsf{P} = \max_{\{\mathbf{x}_i, \mathbf{P}_i(\mathbf{h}_i)\} \in \mathcal{B}_i} & \sum_{i \in \mathcal{V}, k \in \mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i \in \mathcal{V}} V_i(p_i) \\ & \text{s.t. constraints (1), (2), (3), (5).} \qquad (7)\end{aligned}$$

Our goal is to develop a distributed algorithm to solve (7) without accessing the channel pdf $m_{\mathbf{h}}(\cdot)$. This is challenging because i) the optimization space in (7) includes functions $Q_{ij}(\mathbf{h}_i)$ and $P_{ij}(\mathbf{h}_i)$ implying that the dimension of the problem is infinite; ii) since the capacity constraint (5) is non-convex and the capacity function may be even discontinuous, (7) is a non-convex optimization problem; iii) constraints (3) and (5) involve expectations over channel states $\mathbf{h}$ whose pdf is unknown; and iv) the fact that the transmission rate c_{ij} is determined not only by the transmitter but also by the receiver and his neighbors [cf. (5)] hinders the development of distributed optimization algorithms.

Notice that the number of constraints in (7) is finite. This implies that while there are infinite number of variables in the primal domain, there are a finite number of variables in the dual domain. Thus, while working in the dual domain may entail some loss of optimality due the non-convex constraints in (7), it does overcome challenge i) because the dual function is finite dimensional. It also overcomes challenge ii) since the dual function is always convex, while challenge iii) can be solved by using stochastic subgradient descent algorithms on the dual function; see, e.g., [18] and [21]. However, working with the dual problem of (7) does not conduce to a distributed optimization algorithm due to the coupling introduced by constraint (5). This prompts the introduction of a decomposable approximation that we pursue in the next section.

B. Problem Approximation

For reasons that will become clear in Section III, a distributed solution of the problem in (7) is not possible because scheduling functions $Q_{ij}(\mathbf{h}_i)$ and $Q_l(\mathbf{h}_l)$ are coupled as a product in constraint (5). If we reformulate this constraint into an expression in which the terms $C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i))Q_{ij}(\mathbf{h}_i)$ and $1 - Q_l(\mathbf{h}_l)$ appear as summands instead of as factors of a product the problem will become decomposable in the dual domain. This reformulation can be accomplished by taking logarithms on both sides of (5), yielding

$$\begin{aligned}\tilde{c}_{ij} := \log c_{ij} \le \log \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i))Q_{ij}(\mathbf{h}_i)\right] \\ + \sum_{l \in \mathcal{M}_i(j)} \log\left[1 - \mathbb{E}_{\mathbf{h}_l}[Q_l(\mathbf{h}_l)]\right] \qquad (8)\end{aligned}$$

where we defined $\tilde{c}_{ij} := \log c_{ij}$. While scheduling functions of different terminals now appear as summands on the right-hand side of (8), the link capacity constraint (2) mutates into the non-convex constraint $\sum_{k \in \mathcal{K}} r_{ij}^k \le e^{\tilde{c}_{ij}}$. To avoid this issue, we use the linear lower bound $1 + \tilde{c}_{ij} \le e^{\tilde{c}_{ij}}$ and approximate this constraint as $\sum_{k \in \mathcal{K}} r_{ij}^k \le 1 + \tilde{c}_{ij}$. Upon defining the average attempted transmission rate of link (i,j) as

$$x_{ij} := \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i))Q_{ij}(\mathbf{h}_i)\right], \qquad (9)$$

and the transmission probability of terminal i as

$$y_i := \mathbb{E}_{\mathbf{h}_i}[Q_i(\mathbf{h}_i)], \qquad (10)$$

the original optimization problem P is approximated by

$$\begin{aligned}\mathsf{P} \ge \tilde{\mathsf{P}} = & \max_{\{\tilde{\mathbf{x}}_i, \mathbf{P}_i(\mathbf{h}_i)\} \in \mathcal{B}_i} \sum_{i \in \mathcal{V}, k \in \mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i \in \mathcal{V}} V_i(p_i) \\ & \text{s.t. } a_i^k \le \sum_{j \in \mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k\right), \quad \sum_{k \in \mathcal{K}} r_{ij}^k \le 1 + \tilde{c}_{ij}, \\ & \tilde{c}_{ij} \le \log x_{ij} + \sum_{l \in \mathcal{M}_i(j)} \log(1 - y_l), \\ & x_{ij} \le \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i))Q_{ij}(\mathbf{h}_i)\right], \\ & y_i \ge \mathbb{E}_{\mathbf{h}_i}[Q_i(\mathbf{h}_i)], \\ & p_i \ge \mathbb{E}_{\mathbf{h}_i}\left[\sum_{j \in \mathcal{N}(i)} P_{ij}(\mathbf{h}_i)Q_{ij}(\mathbf{h}_i)\right] \qquad (11)\end{aligned}$$

where we defined $\tilde{\mathbf{x}}_i := [\mathbf{x}_i, x_{ij}, y_i]$ and relaxed the definitions of attempted transmission rate and transmission probability, which we can do without loss of optimality. Problems (7) and (11) are *not* equivalent because of the linear approximation to the link capacity constraint. However, since $1 + \tilde{c}_{ij}$ is a lower bound on $e^{\tilde{c}_{ij}}$, any operating point that satisfies the constraints

[1]If we have channel reciprocity $h_{ij}(t) = h_{ji}(t)$, the derivation of (5) from (4) is no longer valid since power control and channel access functions of neighboring nodes will have common arguments implying that $Q_{ij}(\mathbf{h}_i)$ and $Q_{ji}(\mathbf{h}_j)$ would not be independent. The general methodology used here seems applicable but is beyond the scope of the present paper.

in (11) also satisfies the constraints in (7). In particular, the solution of (11) is feasible in (7), although possibly suboptimal. Further note that variables associated with different terminals appear as different summands of the objective and constraints in (11). This is the signature of optimization problems amenable to distributed implementations as we explain in the next section.

III. Distributed Stochastic Learning Algorithm

To define the dual of the optimization problem in (11) introduce Lagrange multipliers $\mathbf{\Lambda}_i$, associated with terminal i where $\mathbf{\Lambda}_i := \{\lambda_i^k, \mu_{ij}, \nu_{ij}, \alpha_{ij}, \beta_i, \xi_i : \forall j \in \mathcal{N}(i)\}$. The dual variable λ_i^k is associated with the flow conservation constraint in (1), the multiplier μ_{ij} with the reformulated rate constraint $\sum_{k\in\mathcal{K}} r_{ij}^k \leq 1 + \tilde{c}_{ij}$, the variable ν_{ij} with the link capacity constraint $\tilde{c}_{ij} \leq \log x_{ij} + \sum_{l\in\mathcal{M}_i(j)} \log(1-y_l)$, multiplier α_{ij} with the attempted transmission rate constraint in (9), β_i with the transmission probability constraint in (10), and ξ_i with the average power constraint in (3). The Lagrangian for the optimization problem in (11) is given by the sum of the objective and the products of the constraints with their respective multipliers

$$\begin{aligned}
\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda})
&= \sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i\in\mathcal{V}} V_i(p_i) \\
&\quad + \sum_{i\in\mathcal{V},k\in\mathcal{K}} \lambda_i^k \left[\sum_{j\in\mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k\right) - a_i^k\right] \\
&\quad + \sum_{(i,j)\in\mathcal{E}} \mu_{ij}\left[(1+\tilde{c}_{ij}) - \sum_{k\in\mathcal{K}} r_{ij}^k\right] \\
&\quad + \sum_{(i,j)\in\mathcal{E}} \nu_{ij}\left[\log x_{ij} + \sum_{l\in\mathcal{M}_i(j)} \log(1-y_l) - \tilde{c}_{ij}\right] \\
&\quad + \sum_{(i,j)\in\mathcal{E}} \alpha_{ij}\left[\mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\left(h_{ij}P_{ij}(\mathbf{h}_i)\right)Q_{ij}(\mathbf{h}_i)\right] - x_{ij}\right] \\
&\quad + \sum_{i\in\mathcal{V}} \beta_i\left[y_i - \mathbb{E}_{\mathbf{h}_i}\left[Q_i(\mathbf{h}_i)\right]\right] \\
&\quad + \sum_{i\in\mathcal{V}} \xi_i\left[p_i - \mathbb{E}_{\mathbf{h}_i}\left[\sum_{j\in\mathcal{N}(i)} P_{ij}(\mathbf{h}_i)Q_{ij}(\mathbf{h}_i)\right]\right] \qquad (12)
\end{aligned}$$

where we introduced vectors $\tilde{\mathbf{x}}$, $\mathbf{P}(\mathbf{h})$, and $\mathbf{\Lambda}$ grouping $\tilde{\mathbf{x}}_i$, $\mathbf{P}_i(\mathbf{h}_i)$, and $\mathbf{\Lambda}_i$ for all nodes $i \in \mathcal{V}$. The dual function is now defined as the maximum of the Lagrangian in (12) over the set of feasible $\tilde{\mathbf{x}}_i$ and $\mathbf{P}_i(\mathbf{h}_i)$ and the dual problem as the minimum of $g(\mathbf{\Lambda})$ over positive dual variables, i.e.,

$$\tilde{\mathsf{D}} = \min_{\mathbf{\Lambda}\geq 0} g(\mathbf{\Lambda}) = \min_{\mathbf{\Lambda}\geq 0} \max_{\{\tilde{\mathbf{x}}_i, \mathbf{P}_i(\mathbf{h}_i)\}\in\mathcal{B}_i} \mathcal{L}(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}). \qquad (13)$$

Despite being non-convex, the structure of the problem in (11) is such that $\tilde{\mathsf{P}} = \tilde{\mathsf{D}}$ as long as the fading distribution has no realization of nonzero probability; see [20]. This lack of duality gap implies that the finite dimensional and convex dual problem is equivalent to the infinite dimensional and nonconvex primal problem.

Further note that the Lagrangian in (12) exhibits a separable structure because all summands involve a single primal variable. Consider all summands of (12) that involve network variables associated with terminal i and define the local Lagrangian at terminal i as

$$\begin{aligned}
&\mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}) \\
&:= \sum_k U_i^k\left(a_i^k\right) - \lambda_i^k a_i^k + \sum_{j\in\mathcal{N}(i)} \left(\lambda_i^k - \lambda_j^k - \mu_{ij}\right) r_{ij}^k \\
&\quad + \sum_{j\in\mathcal{N}(i)} (\mu_{ij} - \nu_{ij})\tilde{c}_{ij} + (\xi_i p_i - V_i(p_i)) \\
&\quad + \sum_{j\in\mathcal{N}_i} \left[\nu_{ij}\log x_{ij} - \alpha_{ij}x_{ij}\right] \\
&\quad + \beta_i y_i + \left[\sum_{k\in\mathcal{N}(i)} \left(\nu_{ki} + \sum_{l\in\mathcal{M}_i(j)} \nu_{lk}\right)\right] \log(1-y_i). \qquad (14)
\end{aligned}$$

Define also the local per channel Lagrangian $\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda})$ grouping all summands of (12) that involve resource allocations of a given terminal i and a given channel realization $\mathbf{h}_i$, i.e.,

$$\begin{aligned}
&\mathcal{L}_i^{(2)}\left(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}\right) \\
&:= \sum_{j\in\mathcal{N}(i)} Q_{ij}(\mathbf{h}_i)\left[\alpha_{ij}C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i)) - \beta_i - \xi_i P_{ij}(\mathbf{h}_i)\right]. \qquad (15)
\end{aligned}$$

It is easy to see by reordering summands in (12) that we can rewrite the Lagrangian as a sum of the local terms $\mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda})$ and an expectation of the local per channel components $\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda})$,

$$\begin{aligned}
&\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}) \\
&\quad = \sum_{i\in\mathcal{V}} \mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}) + \mathbb{E}_{\mathbf{h}_i}\left[\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda})\right]. \qquad (16)
\end{aligned}$$

This separability on per-terminal terms $\mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda})$ and per-terminal and per-channel elements $\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda})$ is exploited in the next section to develop a distributed stochastic subgradient descent algorithm on the dual domain that solves the dual problem (13) and, indirectly, the primal problem (11).

A. Stochastic Subgradient Descent

The dual stochastic subgradient descent algorithm consists of recursive updates of dual variables along stochastic subgradient directions $\mathbf{s}(t)$ moderated by a constant stepsize ϵ,

$$\mathbf{\Lambda}(t+1) = \left[\mathbf{\Lambda}(t) - \epsilon\mathbf{s}(t)\right]^+ \qquad (17)$$

where the operator $[\,\cdot\,]^+$ denotes projection to the nonnegative quadrant. The stochastic subgradient $\mathbf{s}(t)$ in (17) is a vector whose expectation is a descent direction of the dual function.

The important observation is that a stochastic subgradient $\mathbf{s}(t)$ can be computed from primal maximizers of the Lagrangian $\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}(t))$. At time t terminal i proceeds to compute primal variables $\tilde{\mathbf{x}}_i(t) = [a_i^k(t), r_{ij}^k(t), \tilde{c}_{ij}(t), p_i(t), x_{ij}(t), y_i(t)]$ that maximize the local Lagrangian $\mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda})$,

$$\tilde{\mathbf{x}}_i(t) = \underset{\tilde{\mathbf{x}}_i}{\operatorname{argmax}}\, \mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}(t)). \qquad (18)$$

It then observes local channel realizations $\mathbf{h}_i(t)$ and determines instantaneous resource allocation variables $\mathbf{P}_i(t) = [p_{ij}(t), q_{ij}(t)]$ that optimize the local per-channel Lagrangian $\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i(t)), \mathbf{h}_i(t), \mathbf{\Lambda})$ associated with the observed channel realization $\mathbf{h}_i(t)$, i.e.,

$$\mathbf{P}_i(t) = \operatorname*{argmax}_{\mathbf{P}_i} \mathcal{L}_i^{(2)}(\mathbf{P}_i, \mathbf{h}_i(t), \mathbf{\Lambda}(t)). \tag{19}$$

Based on the primal Lagrangian maximizers $\tilde{\mathbf{x}}_i(t)$ and $\mathbf{P}_i(t)$ defined in (18)–(19), a stochastic subgradient $\mathbf{s}(t)$ is obtained by evaluating the resultant constraint slack; see e.g., [21]. E.g., the multiplier λ_i^k is associated with the flow conservation constraint $\sum_{j\in\mathcal{N}(i)}(r_{ij}^k - r_{ji}^k) - a_i^k$. Consequently, the stochastic subgradient component $s_{\lambda_i^k}(t)$ along the λ_i^k direction is given by the constraint slack

$$s_{\lambda_i^k}(t) = \sum_{j\in\mathcal{N}(i)} \left(r_{ij}^k(t) - r_{ji}^k(t)\right) - a_i^k(t). \tag{20}$$

Likewise, components $s_{\mu_{ij}}(t)$ along the μ_{ij} direction and $s_{\nu_{ij}}(t)$ along the ν_{ij} direction can be obtained as

$$\begin{aligned} s_{\mu_{ij}}(t) &= (1 + \tilde{c}_{ij}(t)) - \sum_{k\in\mathcal{K}} r_{ij}^k(t), \\ s_{\nu_{ij}}(t) &= \log x_{ij}(t) + \sum_{l\in\mathcal{M}_i(j)} \log(1 - y_l(t)) - \tilde{c}_{ij}(t). \end{aligned} \tag{21}$$

For the components $s_{\alpha_{ij}}(t)$, $s_{\beta_i}(t)$, and $s_{\xi_i}(t)$ along the α_{ij}, β_i, and ξ_i directions the corresponding constraints involve expectation with respect to the channel distribution. Since we implement stochastic subgradient descent algorithm, we compute instantaneous constraint slacks where the expectation is replaced by the values associated with the current channel realizations $\mathbf{h}_i(t)$

$$\begin{aligned} s_{\alpha_{ij}}(t) &= C_{ij}\left(h_{ij}(t)p_{ij}(t)\right) q_{ij}(t) - x_{ij}(t), \\ s_{\beta_i}(t) &= y_i(t) - q_i(t), \\ s_{\xi_i}(t) &= p_i(t) - \sum_{j\in\mathcal{N}(i)} p_{ij}(t) q_{ij}(t). \end{aligned} \tag{22}$$

Further note that since network variables $\tilde{\mathbf{x}}_i = [a_i^k, r_{ij}^k, \tilde{c}_{ij}, p_i, x_{ij}, y_i]$ appear as separate summands in $\mathcal{L}_i^{(1)}(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}(t))$ [cf. (16)], the maximization in (18) can be carried out separately with respect to individual variables. Specifically, $r_{ij}^k(t)$ and $\tilde{c}_{ij}(t)$ are obtained by solving the following maximization problems

$$\begin{aligned} r_{ij}^k(t) &= \operatorname*{argmax}_{0\le r_{ij}^k \le r_{ij}^{\max}} \left(\lambda_i^k(t) - \lambda_j^k(t) - \mu_{ij}(t)\right) r_{ij}^k, \\ \tilde{c}_{ij}(t) &= \operatorname*{argmax}_{0\le \tilde{c}_{ij} \le \tilde{c}_{ij}^{\max}} \left(\mu_{ij}(t) - \nu_{ij}(t)\right) \tilde{c}_{ij}. \end{aligned} \tag{23}$$

Notice that the maximands in (23) are linear functions of bounded variables which therefore have trivial solutions. E.g., $r_{ij}^k(t) = r_{ij}^{\max}$ if $\lambda_i^k(t) - \lambda_j^k(t) - \mu_{ij}(t) > 0$ and $r_{ij}^k(t) = 0$ otherwise. Solving for $a_i^k(t)$, $p_i(t)$, $x_{ij}(t)$ and $y_i(t)$ is also easy as it involves maximizing concave functions over convex sets of variables,

$$\begin{aligned} a_i^k(t) &= \operatorname*{argmax}_{a_i^{\min}\le a_i^k \le a_i^{\max}} U_i^k\left(a_i^k\right) - \lambda_i^k(t) a_i^k, \\ p_i(t) &= \operatorname*{argmax}_{0\le p_i \le p_i^{\max}} \xi_i(t) p_i - V_i(p_i), \\ x_{ij}(t) &= \operatorname*{argmax}_{x_{ij}\ge 0} \nu_{ij}(t) \log x_{ij} - \alpha_{ij}(t) x_{ij}, \\ y_i(t) &= \operatorname*{argmax}_{0\le y_i\le 1} \beta_i(t) y_i \\ &\quad + \left[\sum_{j\in\mathcal{N}(i)} \left(\nu_{ji}(t) + \sum_{l\in\mathcal{M}_i(j)} \nu_{lj}(t)\right)\right] \log(1 - y_i). \end{aligned} \tag{24}$$

Closed-form solutions for the maximizations in (24) can be easily obtained by solving for the zero of the derivative with respect to the optimization variable, and projecting the result on the feasible set. For example, the solution for the attempted transmission rate is $x_{ij}(t) = \nu_{ij}(t)/\alpha_{ij}(t)$.

The maximization in (19) can be written explicitly as

$$\begin{aligned} \{p_{ij}(t), q_{ij}(t)\} &= \operatorname*{argmax} \sum_{j\in\mathcal{N}(i)} q_{ij}\left[\alpha_{ij}(t) C_{ij}(h_{ij}(t) p_{ij})\right. \\ &\qquad \left. - \beta_i(t) - \xi_i(t) p_{ij}\right] \\ \text{s.t. } & p_{ij} \in \left[0, p_{ij}^{\max}\right],\ q_{ij} \in \{0, 1\}, \quad \sum_{j\in\mathcal{N}(i)} q_{ij} \in \{0, 1\}. \end{aligned} \tag{25}$$

Different from the maximizations in (23)–(24), the one in (25) is a non-convex problem because $C_{ij}(h_{ij} p_{ij})$ may be a non-convex function of p_{ij} and in any event the channel access indicator q_{ij} is an integer variable. Solving (25) is still simple, however, as it involves just two variables; see Remark 1.

To complete the definition of the stochastic subgradient descent algorithm we need an expression for $c_{ij}(t)$. Recall that in formulating (11) we made $c_{ij} = e^{\tilde{c}_{ij}} \ge 1 + \tilde{c}_{ij}$, which implies that at time t we should set

$$c_{ij}(t) = 1 + \tilde{c}_{ij}(t). \tag{26}$$

While the sequence of primal variables $\tilde{\mathbf{x}}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ is a byproduct of the dual stochastic subgradient descent algorithm, it is the optimality of these sequences, not $\mathbf{\Lambda}(\mathbb{N})$, that we want to study. In general, individual primal iterates $\tilde{\mathbf{x}}_i(t)$ and $\mathbf{P}_i(t)$ may not be optimal but sequences $\tilde{\mathbf{x}}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ have ergodic limits that are almost surely feasible and give a utility yield close to $\tilde{\mathsf{P}}$; see Section IV. In order to simplify upcoming discussions, define the ergodic limit of the sequence of operating points $\mathbf{x}_i(\mathbb{N})$ as

$$\bar{\tilde{\mathbf{x}}}_i := \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \mathbf{x}_i(u). \tag{27}$$

Note that subsumed in the definition in (27) are corresponding definitions for each of the individual sequences of admission rates $\bar{a}_i^k := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} a_i^k(u)$, routes, $\bar{r}_{ij}^k := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} r_{ij}^k(u)$, link capacities $\bar{c}_{ij} :=$

$\lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} c_{ij}(u)$, powers $\bar{p}_i := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} p_i(u)$, attempted transmission rates $\bar{x}_{ij} := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} x_{ij}(u)$, and transmission probabilities $\bar{y}_i := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} y_i(u)$.

Remark 1: To find $p_{ij}(t)$ and $q_{ij}(t)$ that solve (25) observe that since $q_{ij} \geq 0$ and the constraints on p_{ij} are separate for different j, the optimal selection for p_{ij} is

$$p_{ij}(t) = \underset{p_{ij}\in\left[0,p_{ij}^{\max}\right]}{\operatorname{argmax}} \alpha_{ij}(t)C_{ij}(h_{ij}(t)p_{ij}) - \beta_i(t) - \xi_i(t)p_{ij}. \tag{28}$$

Also note that q_{ij} can only take values from $\{0,1\}$ and that only one of the q_{ij} variables can be set to 1. If all the optimal objectives computed by (28) are negative, i.e., $\alpha_{ij}(t)C_{ij}(h_{ij}(t)p_{ij}(t)) - \beta_i(t) - \xi_i(t)p_{ij}(t) \leq 0$, the optimal solution for (25) is $q_{ij}(t) = 0$ for all neighbors. Otherwise, the optimal solution for (25) is obtained by setting $q_{ij}(t) = 1$ for the neighbor with the largest objective in (28). In summary, we determine

$$j^*(t) = \underset{j\in\mathcal{N}(i)}{\operatorname{argmax}} \alpha_{ij}(t)C_{ij}(h_{ij}(t)p_{ij}(t)) - \beta_i(t) - \xi_i(t)p_{ij}(t) \tag{29}$$

and set $q_{ij}(t) = 0$ for $j \neq j^*(t)$. For $j = j^*(t)$ we set $q_{ij}(t) = q_{ij^*(t)}(t) = 1$ as long as $\alpha_{ij}(t)C_{ij}(h_{ij}(t)p_{ij}(t)) - \beta_i(t) - \xi_i(t)p_{ij}(t) > 0$ or we make $q_{ij^*(t)}(t) = 0$ otherwise.

Remark 2: If the channel probability distribution is known we can compute powers corresponding not only to $\mathbf{h}(t)$ as in (19), but to generic channel realization $\mathbf{h}$

$$\mathbf{P}_i(\mathbf{h}, \mathbf{\Lambda}(t)) = \underset{\mathbf{P}_i}{\operatorname{argmax}} \mathcal{L}_i^{(2)}(\mathbf{P}_i, \mathbf{h}_i, \mathbf{\Lambda}(t)). \tag{30}$$

We can then use knowledge of the channel distribution to compute not instantaneous constraint slacks as in (22) but actual (average) constraint slacks

$$\begin{aligned} \tilde{s}_{\alpha_{ij}}(t) &= \mathbf{E}\left[C_{ij}\left(h_{ij}p_{ij}(\mathbf{h}, \mathbf{\Lambda}(t))\right) q_{ij}(\mathbf{h}, \mathbf{\Lambda}(t))\right] - x_{ij}(t), \\ \tilde{s}_{\beta_i}(t) &= y_i(t) - \mathbf{E}\left[q_i(\mathbf{h}, \mathbf{\Lambda}(t))\right], \\ \tilde{s}_{\xi_i}(t) &= p_i(t) - \mathbf{E}\left[\sum_{j\in\mathcal{N}(i)} p_{ij}(\mathbf{h}, \mathbf{\Lambda}(t))q_{ij}(\mathbf{h}, \mathbf{\Lambda}(t))\right]. \end{aligned} \tag{31}$$

The constraint slacks $\tilde{s}_{\alpha_{ij}}(t)$, $\tilde{s}_{\beta_i}(t)$, and $\tilde{s}_{\xi_i}(t)$ are gradients of the dual function and can be used in the descent (17) in lieu of the stochastic subgradients $s_{\alpha_{ij}}(t)$, $s_{\beta_i}(t)$, and $s_{\xi_i}(t)$. This will result in faster convergence but necessitates estimation of the channel probability distribution. The use of stochastic subgradients not only avoids this estimation problem but is also less computationally demanding and makes it easier to adapt to changes in channel statistics.

B. Network Operation, Layers, and Layer Interfaces

To describe the role of different variables as computed in (23)–(26) in the network's operation it is convenient to think in terms of a layered architecture with $a_i^k(t)$ associated with the transport layer, $r_{ij}^k(t)$ with the network layer, $c_{ij}(t)$ with the link layer, $x_{ij}(t)$, $y_i(t)$, and $p_i(t)$ with the medium access (MAC) layer, and $p_{ij}(t)$ and $q_{ij}(t)$ with the physical layer; see Figs. 1 and 2.

Variables $a_i^k(t)$, $r_{ij}^k(t)$, $c_{ij}(t)$, $p_{ij}(t)$ and $q_{ij}(t)$ determine network operation by controlling the flow of packets through queues associated with their corresponding layers; see Fig. 2. In the transport and network layers there are queues associated with each of the $|\mathcal{K}|$ flows. In the link and physical layers, queues for each of the $|\mathcal{N}(i)|$ outgoing links (i,j) are maintained. The value of $a_i^k(t)$ determines how many packets are moved from the k-flow queue in the transport layer to the k-flow queue at the network layer at time t. The number of packets transferred at time t from the k-flow network layer queue to the (i,j) queue at the link layer is determined by $r_{ij}^k(t)$. Notice that packets of a particular queue in the network layer may be distributed to different queues in the link layer. Conversely, packets in a particular queue in the link layer may come from different network layer queues, i.e., they may belong to different flows. At time t there are $c_{ij}(t)$ packets moved from the (i,j) queue at the link layer to the (i,j) queue at the physical layer.

At the physical layer queues are emptied through transmission to neighboring terminals. Resource allocation variables $q_{ij}(t)$ and $p_{ij}(t)$ determine the scheduling and transmitted power of link (i,j). If a transmission is scheduled and successful, i.e., a collision does not occur, $C_{ij}(h_{ij}(t)p_{ij}(t))$ units of information are transferred to terminal j from the (i,j) physical layer queue at terminal i. If a collision occurs, they stay at the same queue awaiting retransmission in a future time slot. When a packet is successfully decoded by terminal j it determines which flow they belong to and what destination they are heading for. If the terminal happens to be the destination, packets are forwarded to the application layer. If the terminal is not the designated destination, packets are put into a network layer queue according to their flow identifications.

Besides administering queues, layers are also responsible for updating the values of their corresponding primal variables according to (23)–(26); see Fig. 1. The transport layer updates $a_i^k(t)$ as in (24), the network layer keeps track of $r_{ij}^k(t)$ as per (23), while the link layer computes $\tilde{c}_{ij}(t)$ as in (23) and $c_{ij}(t)$ using (26). The MAC layer updates $p_i(t)$, $x_{ij}(t)$, and $y_i(t)$ according to the expressions in (24), while the physical layer determines $p_{ij}(t)$ and $q_{ij}(t)$ as dictated by (25).

Computation of these primal per layer updates necessitates access to Lagrange multipliers motivating the introduction of layer interfaces to maintain and update their values. E.g., since $\lambda_{ij}^k(t)$ is associated with the flow conservation constraint that relates transport variables $a_i^k(t)$ and network variables $r_{ij}^k(t)$ it provides a natural interface between the transport and network layers. Thus, we introduce a transport-network interface tasked with computing the dual stochastic subgradient component $s_{\lambda_i^k}(t)$ in (20) and executing the update $\lambda_i^k(t+1) = [\lambda_i^k(t) - \epsilon s_{\lambda_i^k}(t)]^+$. Similarly, a network-link interface is introduced to keep track of multipliers $\mu_{ij}(t)$, compute the dual stochastic subgradient component $s_{\mu_{ij}}(t)$ in (21), and execute the corresponding update. A link-MAC interface does the proper for multipliers $\nu_{ij}(t)$ and dual stochastic subgradient components $s_{\nu_{ij}}(t)$ in (21). The remaining multipliers $\alpha_{ij}(t)$, $\beta_i(t)$, and $\xi_i(t)$ provide a MAC-physical interface with stochastic subgradient components $s_{\alpha_{ij}}(t)$, $s_{\beta_i}(t)$, and $s_{\xi_i}(t)$ as given in (22). Observe that primal variables are updated with information available at adjacent interfaces, while dual variable

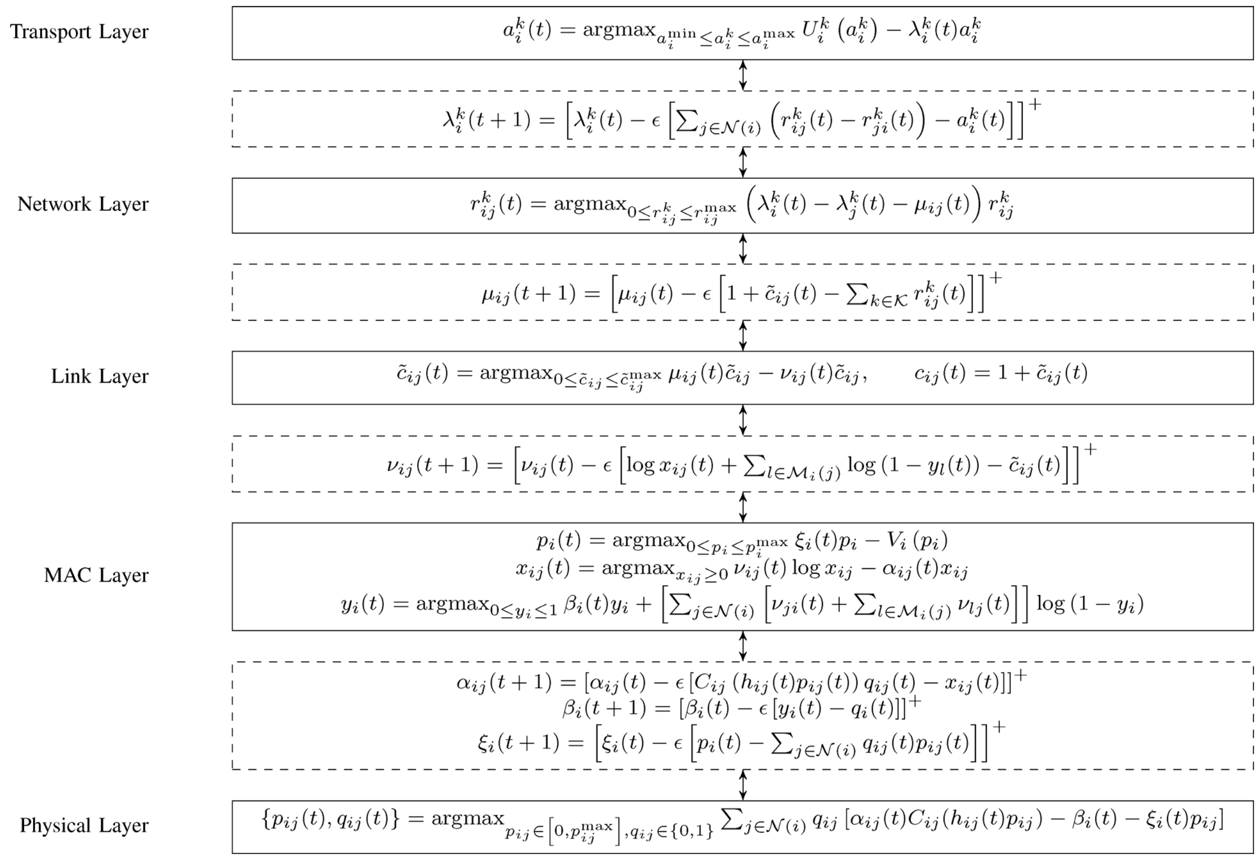

Fig. 1. Layers and layer interfaces. The stochastic subgradient descent algorithm in terms of layers and layer interfaces. Layers maintain primal variables $a_i^k(t)$, $r_{ij}^k(t)$, $\tilde{c}_{ij}(t)$, $p_{ij}(t)$, $q_{ij}(t)$ as well as auxiliary variables $p_i(t)$, $x_{ij}(t)$, and $y_i(t)$ while multipliers $\lambda_i^k(t)$, $\mu_{ij}(t)$, $\nu_{ij}(t)$, $\alpha_{ij}(t)$, $\beta_i(t)$ and $\xi_i(t)$ are associated with interfaces between adjacent layers. Primal variables can be easily computed based on multipliers from interfaces to adjacent layers and dual variables are updated using information from adjacent layers.

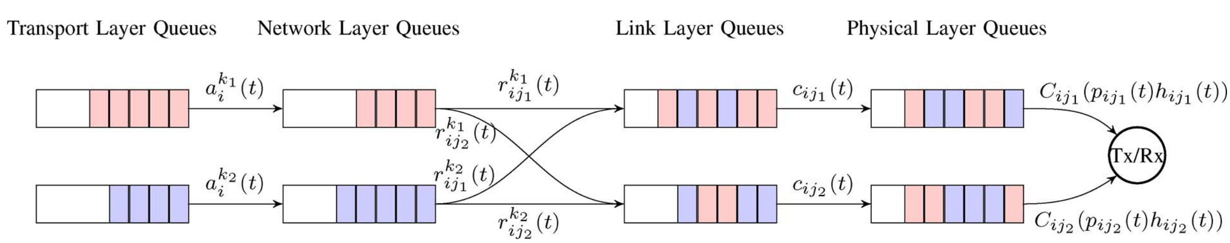

Fig. 2. Queue dynamics. Terminal i operates by controlling queues in different layers based on operating points $a_i^k(t)$, $r_{ij}^k(t)$, $c_{ij}(t)$, $p_{ij}(t)$, and $q_{ij}(t)$. In the transport layer and the network layer, each flow k has a queue. In the link layer and the physical layer, each outgoing link (i,j) maintains a queue. In this particular example, there are two flows k_1 and k_2, and there are two neighboring nodes j_1 and j_2. Packets for flow k_1 are marked red while packets for k_2 are in blue.

updates are undertaken with information available at adjacent layers. Their definition is thereby justified, because information is exchanged only between adjacent layers and interfaces.

We remark that MAC layer variables $x_{ij}(t)$, $y_i(t)$, and $p_i(t)$ do not affect network operation, i.e., queue dynamics, at time t. The role of these variables is to record average behaviors of the terminal to affect determination of $c_{ij}(t)$, $p_{ij}(t)$, and $q_{ij}(t)$ in subsequent time slots. This role is consistent with the definitions of p_i as the average transmitted power [cf. (3)], x_{ij} as the average attempted transmission rate [cf. (9)], and y_i as the (average) transmission probability [cf. (10)].

C. Message Passing

Most primal and dual variable updates in Fig. 1 can be done locally at terminal i. E.g., the physical layer update at terminal i requires access to multipliers $\alpha_{ij}(t)$, $\beta_i(t)$, and $\xi_i(t)$ which are available at the physical-MAC interface of terminal i. The updates for primal variables $r_{ij}^k(t)$ and $y_i(t)$, as well as duals $\lambda_{ij}^k(t)$ and $\nu_{ij}(t)$, however, necessitate access to variables of other terminals. The update of multiplier $\lambda_i^k(t)$ at the network-transport interface depends on network variables $r_{ij}^k(t)$ and $a_i^k(t)$ which are available at terminal i, but also on the variable $r_{ji}^k(t)$ available at (neighboring) terminal j. Similarly, the $r_{ij}^k(t)$ update at the network layer depends on locally available multipliers $\lambda_i^k(t)$ and $\mu_{ij}(t)$, but also on the neighboring multiplier $\lambda_j^k(t)$. The update of multiplier $\nu_{ij}(t)$ is somewhat more complex as it depends on local variables $x_{ij}(t)$ and $\tilde{c}_{ij}(t)$, 1-hop neighborhood variables $y_j(t)$, and 2-hop neighborhood variables $y_l(t)$ for all $l \in \mathcal{N}(j)$. Likewise, the update for $y_i(t)$ at the MAC layer depends on local dual variables $\beta_i(t)$, 1-hop neighborhood

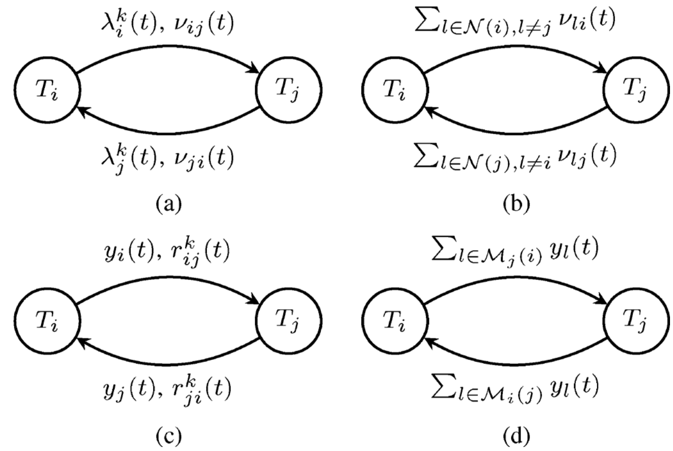

Fig. 3. Message passing. (a) Terminal i begins by transmitting dual variables $\lambda_i^k(t)$ and $\nu_{ij}(t)$ to all neighbors $j\in\mathcal{N}(i)$. (b) It then computes and shares $\sum_{k\in\mathcal{N}(i)}\nu_{ki}(t)$ with all $j\in\mathcal{N}(i)$. This information, along with locally available multipliers, is then used to perform the primal iterations associated with all the layers in Fig. 1. (c) Terminal i passes primal variables $y_i(t)$ and $r_{ij}^k(t)$ to all neighbors $j\in\mathcal{N}(i)$. (d) It then evaluates and broadcasts $\sum_{k\in\mathcal{N}(i)} y_k(t)$ to $j\in\mathcal{N}(i)$. Dual updates associated with the layer interfaces in Fig. 1 are now performed using these and locally accessible primal variables. We proceed to (a) for the next iteration.

variables $\nu_{ji}(t)$ for all $j\in\mathcal{N}(i)$, and 2-hop neighboring variables $\nu_{lj}(t)$ for all $l\in\mathcal{N}(j)$ in the neighborhood of j for some $j\in\mathcal{N}(i)$ in the neighborhood of i. Therefore, implementation of these four updates requires sharing appropriate variables with 1-hop and 2-hop neighbors.

Given that these four updates depend on quantities available at 1-hop and 2-hop neighbors it is necessary to devise a message passing mechanism among terminals to share the necessary values. For doing so we use the 4-step message passing mechanism illustrated in Fig. 3. At the beginning of primal iteration, terminal i transmits $\lambda_i^k(t)$ and $\nu_{ij}(t)$ to all his neighbors $j\in\mathcal{N}(i)$; Fig. 3(a). As a result, terminal i receives multipliers $\lambda_j^k(t)$ and $\nu_{ji}(t)$ from all of their neighbors $j\in\mathcal{N}(i)$. Terminal i follows by computing and broadcasting the term $\sum_{l\in\mathcal{N}(i)}\nu_{li}(t)$ to all his neighbors $j\in\mathcal{N}(i)$; Fig. 3(b). Upon receiving this information, terminal j subtracts $\nu_{ji}(t)$ from the received value to evaluate the expression $\sum_{l\in\mathcal{N}(i),l\neq j}\nu_{li}(t)$. The terms required for computing primal variables $r_{ij}^k(t)$ and $y_i(t)$ are now available at i. Since the variables necessary for the remaining primal updates are locally accessible the primal iterations associated with all the layers in Fig. 1 are performed at each terminal.

After completing the layer updates, primal iterates $r_{ij}^k(t)$ and $y_i(t)$ need to be exchanged between neighbors to perform the dual updates associated with the layer interfaces in Fig. 1. Terminal i starts passing variables $y_i(t)$ and $r_{ij}^k(t)$ to all his neighbors; Fig. 3(c). Having received $y_j(t)$ from all $j\in\mathcal{N}(i)$ terminal i computes and broadcasts the sum $\sum_{l\in\mathcal{N}(i)} y_l(t)$ to all his neighbors; Fig. 3(d). With this information in hand terminal j adds $y_i(t)$ and subtracts $y_j(t)$ from this value to evaluate $\sum_{l\in\mathcal{M}_j(i)} y_l(t)=\sum_{l\in\mathcal{N}(i)} y_l(t)+y_i(t)-y_j(t)$. Quantities necessary to update $\lambda_i^k(t)$ and $\nu_{ij}(t)$ are now available along with the terms necessary for the remaining dual updates that were locally available. The dual updates associated with the layer interfaces in Fig. 1 are now performed and we proceed to the next primal iteration.

We remark that $r_{ij}^k(t)$ and $\lambda_i^k(t)$ are transmitted to 1-hop neighbors, whereas $y_i(t)$ and $\nu_{ij}(t)$ are sent to 2-hop neighbors. This latter fact holds because transmissions of a given terminal can interfere with neighbors two hops away from her.

D. Successive Convex Approximation

As mentioned in the problem reformulation in Section II-B, we a use linear lower bound to approximate the capacity constraint. In general, we can use a concave function $f_{ij}(\tilde{c}_{ij})$ which is smaller than $e^{\tilde{c}_{ij}}$ to approximate $e^{\tilde{c}_{ij}}$. As a result, instead of directly computing link capacity variable $c_{ij}(t)$, an approximated version $\tilde{c}_{ij}(t)$ is calculated in the primal iteration. In the network operation, the link capacity $c_{ij}(t)=f_{ij}(\tilde{c}_{ij}(t))$ is used in the link layer. While this approximation convexifies the capacity constraint and provides a feasible solution to the original problem, it reduces the size of the feasible set of primal variables. This implies that this obtained link capacity $c_{ij}(t)$ may not be optimal to the original problem. To reduce its impact on optimality, we use different $f_{ij}(\tilde{c}_{ij})$ at different time slots and hope the approximations become better as time grows. Define then $\bar{\tilde{c}}_{ij}(t):=1/t\sum_{u=1}^{t}\tilde{c}_{ij}(u)$ and lower bound $e^{\tilde{c}_{ij}(t+1)}$ with the first order approximation

$$e^{\tilde{c}_{ij}(t+1)}\geq e^{\bar{\tilde{c}}_{ij}(t)}\tilde{c}_{ij}(t+1)+e^{\bar{\tilde{c}}_{ij}(t)}[1-\bar{\tilde{c}}_{ij}(t)]. \qquad (32)$$

Notice that the right-hand side of (32) is a linear function of $\tilde{c}_{ij}(t+1)$ and thus concave. We can then choose $f_{ij}^{(t+1)}(\tilde{c}_{ij})=e^{\bar{\tilde{c}}_{ij}(t)}\tilde{c}_{ij}+e^{\bar{\tilde{c}}_{ij}(t)}[1-\bar{\tilde{c}}_{ij}(t)]$ to approximate $e^{\tilde{c}_{ij}}$ at time slot $t+1$.

IV. Feasibility and Optimality

Solving the optimization problem in (7) entails finding optimal variables $\mathbf{x}_i^*$, and power allocations $\mathbf{P}_i^*(\mathbf{h}_i)$ that satisfy problem constraints and offer optimal yield P. This would require knowledge of the channels' probability distributions and a joint optimization among terminals. To overcome these restrictions and develop an adaptive distributed solution, we reformulated the problem as in (11) entailing a performance degradation to $\tilde{\mathsf{P}}\leq\mathsf{P}$. This reformulation permits introduction of the dual stochastic subgradient descent algorithm, defined by recursive application of (17)–(25), that produces a sequence of network operating points $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$—as well as sequences of auxiliary variables $x_{ij}(\mathbb{N})$ and $y_i(\mathbb{N})$—which given results in [21] are expected to be almost surely feasible and give a utility yield close to $\tilde{\mathsf{P}}$ in an ergodic sense. Notice however, that since (17)–(25) descends on the dual function of the reformulated problem, feasibility holds with respect to the constraints in (11). Our main intent here is to show that sequences of operating points $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ generated by (17)–(25) are also feasible for the optimization problem in (7). Specifically, our goal is to prove the following theorem.

Theorem 1: Consider a wireless network $\mathcal{G}(\mathcal{V},\mathcal{E})$ using random access at the physical layer so that ergodic link capacities are as given in (5). Let $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, $p_i(\mathbb{N})$, $q_{ij}(\mathbb{N})$ and $p_{ij}(\mathbb{N})$ be sequences of network operating points generated by the stochastic descent algorithm in (17)–(25) and denote as $\bar{a}_i^k$, $\bar{r}_{ij}^k$, $\bar{c}_{ij}$, and $\bar{p}_i$ the corresponding ergodic limits of $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, and $p_i(\mathbb{N})$. Assume the following hypotheses: (h1) The second moment of the norm of the stochastic subgradient $\mathbf{s}(t)$ is finite, i.e., $\mathbb{E}_{\mathbf{h}}[\|\mathbf{s}(t)\|^2\,|\,\boldsymbol{\Lambda}(t)]\leq\hat{S}^2$. (h2) There exists a set of strictly feasible primal variables that satisfy the constraints of the reformulated optimization problem in (11) with strict inequality. (h3) The dual function $g(\mathbf{\Lambda})$ of the reformulated problem as defined in (13) has a unique minimizer $\mathbf{\Lambda}^*$. It then holds:

i) Near feasibility of physical layer constraints. There exists a function $M(\epsilon)$ with $\lim_{\epsilon\to 0} M(\epsilon) = 0$ such that the average transmission rate constraint in (5) is almost surely satisfied with feasibility gap smaller than $M(\epsilon)$ in an ergodic sense, i.e.,

$$\bar{c}_{ij} \leq \lim_{t\to\infty} \frac{1}{t} \times \sum_{u=1}^{t} \left[C_{ij}(h_{ij}(u)p_{ij}(u))q_{ij}(u) \prod_{k\in\mathcal{M}_i(j)} [1-q_k(u)] \right] + M(\epsilon), \quad \text{a.s.} \tag{33}$$

ii) Feasibility of upper layer constraints. The flow conservation constraint in (1), the link capacity constraint in (2) and the average power constraint in (3) are almost surely satisfied in an ergodic sense, i.e.,

$$\bar{a}_i^k \leq \sum_{j\in\mathcal{N}(i)} \left[\bar{r}_{ij}^k - \bar{r}_{ji}^k\right], \quad \sum_{k\in\mathcal{K}} \bar{r}_{ij}^k \leq \bar{c}_{ij}, \quad \text{a.s.} \tag{34}$$

$$\bar{p}_i \geq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \sum_{j\in\mathcal{N}(i)} p_{ij}(u)q_{ij}(u), \quad \text{a.s.} \tag{35}$$

iii) Utility yield. The utility yield of the ergodic averages of sequences $a_i^k(\mathbb{N})$ and $p_i(\mathbb{N})$ converges to a value within $\epsilon\hat{S}^2/2$ of $\tilde{\mathsf{P}}$, i.e.,

$$\tilde{\mathsf{P}} - \left[\sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(\bar{a}_i^k\right) - \sum_{i\in\mathcal{V}} V_i(\bar{p}_i)\right] \leq \frac{\epsilon\hat{S}^2}{2}, \quad \text{a.s.} \tag{36}$$

The feasibility results in (34) for the flow conservation and rate constraints are identical to (1) and (2). As such they imply that the ergodic limits $\bar{a}_i^k$, $\bar{r}_{ij}^k$, $\bar{c}_{ij}$ obtained from recursive application of (17)–(25) satisfy these constraints with probability 1. Notice that these limits may be different for different realizations of the algorithm's run. Nonetheless, constraints (1) and (2) are satisfied for almost all runs. The feasibility result in (33) for the link capacity constraint, however, is not identical to (5). The difference is not only the presence of the $M(\epsilon)$ feasibility gap, but the fact that (5) involves an expectation over channel realizations whereas (33) does not. In fact, asides from the $M(\epsilon)$ constant, (33) is stronger than (5). The feasibility result in (33) states that even though sequences $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ may not be ergodic, the possibly different ergodic limits in the right and left-hand sides of (33) satisfy the stated inequality. This implies that operating the network using variables $\mathbf{x}_i(t)$ and $\mathbf{P}_i(t)$ as generated by (17)–(25) results in long-term feasibility in that all packets are (almost surely) delivered to their corresponding destinations. Further notice that the power feasibility result in (35) is not identical to the corresponding power constraint in (3) because (3) involves an expected value whereas (35) does not. The same comments stated for the comparison of (33) and (5) extend naturally.

The utility yield result in (36) states that the long term performance of the network, as determined by average end-to-end rates $\bar{a}_i^k$ and powers $\bar{p}_i$, is close to the optimal yield $\tilde{\mathsf{P}}$ of the reformulated problem. The gap between $\tilde{\mathsf{P}}$ and the attained yield can be controlled by reducing ϵ. Notice that reducing the step size ϵ also reduces the feasibility gap $M(\epsilon)$ in (33). We also remark that the use of constant step sizes ϵ endows the algorithm with adaptability to time-varying channel distributions. This is important in practice because wireless channels are non-stationary due to user mobility and environmental dynamics.

A. *Proof of Theorem 1*

Hypotheses (h1) and (h2) are sufficient for [21, Theorem 1] to hold. The utility yield result in (36) is a direct consequence of [21, Theorem 1]. It also follows that all constraints in problem (11) are almost surely satisfied in an ergodic sense. Since the flow conservation constraint in (1) and the power constraint in (3) are part of (11) the first inequality in (34) and the inequality in (35) follow from direct application of [21, Theorem 1]. In addition, considering the constraint $\sum_{k\in\mathcal{K}} r_{ij}^k \leq 1+\tilde{c}_{ij}$ Theorem 1 of [21] gives us

$$\sum_{k\in\mathcal{K}} \bar{r}_{ij}^k(u) \leq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} [1+\tilde{c}_{ij}(u)], \quad \text{a.s.} \tag{37}$$

Recall now that at every iteration we set the link capacity to $c_{ij}(u) = 1+\tilde{c}_{ij}(u)$. Substituting this equality into (37) the second inequality in (34) follows from the definition $\bar{c}_{ij} := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} c_{ij}(u)$.

The result that does not follow as a simple application of [21, Theorem 1] is the almost sure near feasibility of the average transmission rate constraint as shown in (33). Since we introduced auxiliary variables x_{ij} and y_i and decomposed the average transmission rate constraint in two separate constraints [21, Theorem 1] does not make a claim on the feasibility of (5). Instead, the claim is for the last three constraints in (11), i.e.,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \tilde{c}_{ij}(u) \leq \log[\bar{x}_{ij}(u)] + \sum_{k\in\mathcal{M}_i(j)} \log[1-\bar{y}_k(u)], \quad \text{a.s.} \tag{38}$$

$$\bar{x}_{ij} \leq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} C_{ij}(h_{ij}(u)p_{ij}(u))q_{ij}(u), \quad \text{a.s.} \tag{39}$$

$$\bar{y}_i \leq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} q_i(u), \quad \text{a.s.} \tag{40}$$

Since link capacity iterates are set to $c_{ij}(u) = 1+\tilde{c}_{ij}(u)$ we use the fact that $1+x \leq e^x$ for all x to write

$$\bar{c}_{ij} := \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} c_{ij}(u) = 1 + \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \tilde{c}_{ij}(u) \leq \exp\left[\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \tilde{c}_{ij}(u)\right]. \tag{41}$$

Substitute now the inequality in (38) into the exponent in (41) to obtain

$$\bar{c}_{ij} \leq \exp\left[\log[\bar{x}_{ij}(u)] + \sum_{l\in\mathcal{M}_i(j)} \log[1-\bar{y}_l(u)]\right] = \bar{x}_{ij} \prod_{l\in\mathcal{M}_i(j)} [1-\bar{y}_l(u)] \tag{42}$$

where in the equality we cancelled out the exponential and logarithm functions. Further substituting (39) and (40) into the right-hand side of (42) yields

$$\bar{c}_{ij} \le \left[\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t} C_{ij}(h_{ij}(u)p_{ij}(u))q_{ij}(u)\right] \times \prod_{l\in\mathcal{M}_i(j)}\left[\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}[1-q_l(u)]\right]. \quad (43)$$

While similar, (43) is substantially different from the statement in (33) that we want to prove. To see the difference exploit ergodicity, possibly restricted to an ergodic component, to replace the ergodic limit in (33) by the corresponding expected value so as to write

$$\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}\left[C_{ij}(h_{ij}(u)p_{ij}(u))q_{ij}(u)\prod_{l\in\mathcal{M}_i(j)}[1-q_l(u)]\right] = \lim_{t\to\infty}\mathbb{E}\left[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)\prod_{l\in\mathcal{M}_i(j)}[1-q_l(t)]\right]. \quad (44)$$

Similarly, consider the product of ergodic limits in (43) and use ergodicity, also possibly restricted to an ergodic component, to write each *individual* limit as an expectation,

$$\bar{c}_{ij} \le \lim_{t\to\infty}\mathbb{E}\left[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)\right]\times\prod_{l\in\mathcal{M}_i(j)}\mathbb{E}\left[[1-q_l(t)]\right]. \quad (45)$$

If schedules of different terminals were independent, the expectation in (44) would coincide with the product of expectations in (45) yielding the result in (33) with $M(\epsilon)=0$ after substituting (44) into (45). However, due to the message passing between neighboring terminals correlation in transmission decisions is introduced, independence is violated, and the expectation in (44) may not coincide with the product of expectations in (45). It follows from this discussion that the key point in establishing (33) is to show that the correlation between schedules introduced by message passing is small so that the expectation in (44) equals the product of expectations in (45) except for the vanishingly small difference $M(\epsilon)$.

To prove so start noting that while $C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)$ and $q_l(t)$ for $l\in\mathcal{M}_i(j)$ correlate through message passing, they are conditionally uncorrelated if multipliers $\mathbf{\Lambda}(t)$ are given. This is true because for given $\mathbf{\Lambda}(t)$ schedules and power allocations depend only on local channel realizations, which are assumed independent for different channels. We can therefore write

$$\mathbb{E}_{\mathbf{h}}\left[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)\prod_{l\in\mathcal{M}_i(j)}[1-q_l(t)]\,\middle|\,\mathbf{\Lambda}(t)\right] = \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)\,|\,\mathbf{\Lambda}(t)\right] \times\prod_{l\in\mathcal{M}_i(j)}\left[1-\mathbb{E}_{\mathbf{h}_l}[q_l(t)\,|\,\mathbf{\Lambda}(t)]\right]. \quad (46)$$

The conditional expectations in (46) and the (unconditional) ones in (44) and (45) can be related through double integration, e.g.,

$$\mathbb{E}[q_i(t)] = \int \mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)]\,d\mathbf{\Lambda}(t). \quad (47)$$

The crucial observation is that since (17)–(25) descends in the dual domain, $\mathbf{\Lambda}(t)$ approaches the optimal multiplier $\mathbf{\Lambda}^*$ as t grows; see e.g., [21, Theorem 2]. This motivates the introduction of a set $\mathcal{A}$ containing all multipliers $\mathbf{\Lambda}$ within a given small distance $\sqrt{\delta}$ of $\mathbf{\Lambda}^*$, i.e., $\mathcal{A}=\{\mathbf{\Lambda}\,|\,\|\mathbf{\Lambda}-\mathbf{\Lambda}^*\|^2\le\delta\}$. We can then separate the integration with respect to $\mathbf{\Lambda}(t)$ in (47) into terms that contain multipliers inside and outside $\mathcal{A}$,

$$\mathbb{E}[q_i(t)] = \int_{\mathbf{\Lambda}(t)\in\mathcal{A}}\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)]\,d\mathbf{\Lambda}(t) + \int_{\mathbf{\Lambda}(t)\in\mathcal{A}^c}\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)]\,d\mathbf{\Lambda}(t). \quad (48)$$

By making δ small enough the first integral in (48) can be made arbitrarily close to $\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*]$. Since $\mathbf{\Lambda}(t)$ gets close to $\mathbf{\Lambda}^*$ as t increases, the second integral can be made small for sufficiently large t.

While we have exemplified the argument for the expectation $\mathbb{E}[q_i(t)]$ the same is true for the other expectations in (44) and (45). The idea to complete the proof is to show that for sufficiently large t all expectations can be written as conditional expectations given $\mathbf{\Lambda}^*$ plus small error terms. Conditional independence is then used to claim (46) from the equivalence of the right-hand sides of (44) and (45). In summary we need to make the following arguments in order to conclude the proof:

A1) For sufficiently large t, the probability of $\mathbf{\Lambda}(t)$ staying within a small distance of $\mathbf{\Lambda}^*$ is close to 1. The distance can be made arbitrarily small and the probability arbitrarily close to 1 by reducing ϵ. This argument is formalized and proved in Lemma 1.

A2) All of the expectations in (44) and (45) can be written as integrals of conditional expectations of the form shown in (48) for $\mathbb{E}[q_i(t)]$. By making the ball $\mathcal{A}$ sufficiently small the (first) integral with respect to multipliers $\mathbf{\Lambda}(t)\in\mathcal{A}$ can be made arbitrarily close to the expectation conditional on $\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*$. From A1) it follows that for any small ball $\mathcal{A}$ the (second) integral with respect to $\mathbf{\Lambda}(t)$ for multipliers $\mathbf{\Lambda}(t)\notin\mathcal{A}$ can be made close to 0 by reducing ϵ. Therefore, it follows that unconditional, e.g., $\mathbb{E}[q_i(t)]$, and conditional, e.g., $\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*]$, expectations get arbitrarily close as $\epsilon\to 0$. This argument is formalized and proved in Lemma 2.

A3) From Argument A2), it follows that the unconditional expectation in (44) can be expressed as an expectation conditioned on $\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*$ plus an arbitrarily small error term. Recalling the fact that given $\mathbf{\Lambda}(t)$ schedules and power allocations for different terminals are uncorrelated we can write the resulting conditional expectation as a product of conditional expectations [cf. (46)].

In turn, Argument A2) implies that each of these expectations is close to the unconditional expectation plus an small error term. The result in (33) follows from ergodicity. This argument is formalized after Lemma 2 to conclude the proof.

Let us start by formalizing argument A1) in the following lemma. The proof of is technical and relegated to Appendix A.

Lemma 1: Consider the stochastic descent algorithm in (17)–(25) with the same hypotheses and definitions of Theorem 1. Let the dual variable $\mathbf{\Lambda}(T_0)$ at given time T_0 be given. Then, there exists time $T_1 > T_0$ such that for all $t > T_1$ it holds

$$\Pr[\|\mathbf{\Lambda}(t) - \mathbf{\Lambda}^*\|^2 \geq L(\epsilon)|\mathbf{\Lambda}(T_0)] \leq L(\epsilon) \tag{49}$$

where $L(\epsilon)$ is a function of the step size ϵ such that $\lim_{\epsilon\to 0} L(\epsilon) = 0$.

Proof: See Appendix A. ∎

Lemma 1 states, as required by argument A1), that the probability of $\mathbf{\Lambda}(t)$ being outside arbitrarily small distance $\sqrt{L(\epsilon)}$ of $\mathbf{\Lambda}^*$ is the arbitrarily small factor $L(\epsilon)$. To formalize A2), we introduce a bounded function $D(\mathbf{h}(t), \mathbf{P}(t))$ to stand in for the functions inside the expectations in (44) and (45). We show that for arbitrary bounded function $D(\mathbf{h}(t), \mathbf{P}(t))$, its unconditional mean is within a small $N(\epsilon)$ constant of its expectation conditional on $\mathbf{\Lambda}(t) = \mathbf{\Lambda}^*$ as long as the conditional expectation is a continuous function of $\mathbf{\Lambda}(t)$.

Lemma 2: Consider the stochastic descent algorithm in (17)–(25) with the same hypotheses and definitions of Theorem 1. Let $0 \leq D(\mathbf{h}(t), \mathbf{P}(t)) \leq D_{\max}$ be a nonnegative continuous function of $\mathbf{h}(t)$, $\mathbf{p}(t)$ and $\mathbf{q}(t)$ upper bounded by $D_{\max}$. Assume the dual variable $\mathbf{\Lambda}(T_0)$ at given time T_0 is given and that the conditional expectation $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]$ is continuous in $\mathbf{\Lambda}(t)$. Then almost surely there exists $T_1 > T_0$ such that for all $t > T_1$ it holds

$$\begin{aligned}|\mathbb{E}\left[D(\mathbf{h}(t), \mathbf{P}(t))|\, \mathbf{\Lambda}(T_0)\right]& \\ -\mathbb{E}_{\mathbf{h}}\left[D(\mathbf{h}(t), \mathbf{P}(t))|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*\right]| &\leq N(\epsilon)\end{aligned} \tag{50}$$

where the first and the second expectations are with respect to $\mathbf{h}(T_0), \ldots, \mathbf{h}(t)$ and $\mathbf{h}(t)$, respectively, and $N(\epsilon)$ is a function of the step size ϵ such that $\lim_{\epsilon\to 0} N(\epsilon) = 0$.

Proof: Start noting that we can write $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)]$ as an integral of conditional expectations [cf. (47)],

$$\begin{aligned}&\mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)] \\ &\quad = \int \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t), \mathbf{\Lambda}(T_0)]\, d\mathbf{\Lambda}(t) \\ &\quad = \int \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t)\end{aligned} \tag{51}$$

where the second equality follows because $\mathbf{\Lambda}(\mathbb{N})$ is a Markov process. Partitioning the integration space into the sets $\mathcal{A}_\epsilon = \{\mathbf{\Lambda} \,|\, \|\mathbf{\Lambda} - \mathbf{\Lambda}^*\|^2 < L(\epsilon)\}$ and $\mathcal{A}_\epsilon^c = \{\mathbf{\Lambda} \,|\, \|\mathbf{\Lambda} - \mathbf{\Lambda}^*\|^2 \geq L(\epsilon)\}$ allows us to rewrite (51) as [cf. (48)]

$$\begin{aligned}&\mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)] \\ &\quad = \int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t) \\ &\quad\quad + \int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon^c} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t).\end{aligned} \tag{52}$$

Since we are assuming that $0 \leq D(\mathbf{h}(t), \mathbf{P}(t)) \leq D_{\max}$ we can bound the second integral on the right-hand side of (52) by

$$\begin{aligned}0 &\leq \int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon^c} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t) \\ &\leq D_{\max} \Pr\left[\mathbf{\Lambda}(t) \in \mathcal{A}_\epsilon^c |\, \mathbf{\Lambda}(T_0)\right].\end{aligned} \tag{53}$$

According to Lemma 1, we know that there exists time $T_1 > T_0$ such that for all $t > T_1$ we have $\Pr[\mathbf{\Lambda}(t) \in \mathcal{A}_\epsilon^c \,|\, \mathbf{\Lambda}(T_0)] = \Pr[\|\mathbf{\Lambda}(t) - \mathbf{\Lambda}^*\|^2 \geq L(\epsilon) \,|\, \mathbf{\Lambda}(T_0)] \leq L(\epsilon)$. Substituting this bound into (53) yields

$$0 \leq \int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon^c} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t) \leq D_{\max} L(\epsilon) \tag{54}$$

for all times $t > T_1$. For the first integral on the right-hand side of (52), observe that since $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]$ is continuous in $\mathbf{\Lambda}(t)$ we can use the mean value theorem to write the integral as

$$\begin{aligned}&\int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t) \\ &= \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0] \Pr[\mathbf{\Lambda}(t) \in \mathcal{A}_\epsilon \,|\, \mathbf{\Lambda}(T_0)],\end{aligned} \tag{55}$$

for a certain $\mathbf{\Lambda}_0 \in \mathcal{A}_\epsilon$. Since for any $t > T_1$ we have $0 \leq \Pr[\mathbf{\Lambda}(t) \in \mathcal{A}_\epsilon^c \,|\, \mathbf{\Lambda}(T_0)] \leq L(\epsilon)$, it follows that $1 - L(\epsilon) \leq \Pr[\mathbf{\Lambda}(t) \in \mathcal{A}_\epsilon \,|\, \mathbf{\Lambda}(T_0)] \leq 1$. Substituting this into (55), we have

$$\begin{aligned}&[1 - L(\epsilon)]\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0] \\ &\quad \leq \int_{\mathbf{\Lambda}(t)\in\mathcal{A}_\epsilon} \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]\, d\mathbf{\Lambda}(t) \\ &\quad \leq \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0].\end{aligned} \tag{56}$$

Substituting (54) and (56) into (52) yields

$$\begin{aligned}&[1 - L(\epsilon)]\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0] \\ &\quad \leq \mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)] \\ &\quad \leq D_{\max} L(\epsilon) + \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0].\end{aligned} \tag{57}$$

To show that (50) is true we find upper bounds for $\mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)] - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*]$ and its opposite $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*] - \mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)]$. Define $L'(\epsilon) := |\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t))|\mathbf{\Lambda}(t) = \mathbf{\Lambda}_0] - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*]|$ and observe that since $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t)]$ is continuous in $\mathbf{\Lambda}(t)$ and $\mathbf{\Lambda}_0 \in \mathcal{A}_\epsilon$, it follows that $\lim_{\epsilon\to 0} L'(\epsilon) = 0$. Using this definition for $L'(\epsilon)$ and the upper bound in (57), we obtain

$$\begin{aligned}&\mathbb{E}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(T_0)] \\ &\quad - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*] \\ &\quad\quad \leq D_{\max} L(\epsilon) + \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}_0] \\ &\quad\quad\quad - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t), \mathbf{P}(t)) \,|\, \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*] \\ &\quad\quad \leq D_{\max} L(\epsilon) + L'(\epsilon).\end{aligned} \tag{58}$$

Similarly, using the definition of $L'(\epsilon)$ and the lower bound in (57) we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*] \\
&\quad - \mathbb{E}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(T_0)] \\
&\qquad \le \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*] \\
&\qquad\quad - [1-L(\epsilon)]\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}_0] \\
&\qquad = L(\epsilon)\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}_0] \\
&\qquad\quad + \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*] \\
&\qquad\quad - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}_0] \\
&\qquad \le D_{\max}L(\epsilon) + L'(\epsilon) \qquad (59)
\end{aligned}$$

where the last inequality follows from the fact that $D(\mathbf{h}(t),\mathbf{P}(t)) \le D_{\max}$. From (58) and (59), we conclude

$$\begin{aligned}
&|\mathbb{E}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(T_0)] \\
&\quad - \mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*]| \le D_{\max}L(\epsilon) + L'(\epsilon).
\end{aligned} \qquad (60)$$

Making $N(\epsilon) := D_{\max}L(\epsilon) + L'(\epsilon)$ in (60) yields (50). Since both $L(\epsilon)$ and $L'(\epsilon)$ approach 0 as ϵ goes to 0, it follows $\lim_{\epsilon\to 0} N(\epsilon) = 0$. ■

In Lemma 2, continuity of $\mathbb{E}_{\mathbf{h}}[D(\mathbf{h}(t),\mathbf{P}(t))\,|\,\mathbf{\Lambda}(t)]$ is assumed. Specifically, we need continuity of $\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)]$ and $\mathbb{E}_{h_{ij}}[q_{ij}(t)C(h_{ij}(t)p_{ij}(t))\,|\,\mathbf{\Lambda}(t)]$. This is indeed true as claimed by the following lemma.

Lemma 3: Consider the calculation of primal variables $p_{ij}(t)$ and $q_{ij}(t)$ as shown in (25), $\mathbb{E}_{\mathbf{h}_i}[q_i(t)\,|\,\mathbf{\Lambda}(t)]$ and $\mathbb{E}_{h_{ij}}[q_{ij}(t)C(h_{ij}(t)p_{ij}(t))\,|\,\mathbf{\Lambda}(t)]$ are continuous functions of $\mathbf{\Lambda}(t)$.

Proof: See Appendix B. ■

Using Lemma 3 we conclude that the hypotheses of Lemma 2 are satisfied. Applying the result in Lemma 2 we then have that for sufficiently large time index t we can rewrite (45) as

$$\begin{aligned}
\bar{c}_{ij} \le\ & \mathbb{E}[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*] \\
& \times \prod_{l\in\mathcal{M}_i(j)} \mathbb{E}[1-q_l(t)\,|\,\mathbf{\Lambda}(t)=\mathbf{\Lambda}^*] + N_1(\epsilon) \qquad (61)
\end{aligned}$$

where $\lim_{\epsilon\to 0} N_1(\epsilon) = 0$. Given $\mathbf{\Lambda}(t)$, $C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t)$ and $q_l(t)$ are uncorrelated [cf. (46)]. This allows us to write the product of expectations on the right-hand side of (61) as an expectation of products, i.e.,

$$\begin{aligned}
\bar{c}_{ij} \le \mathbb{E}_{\mathbf{h}_i}\Bigg[& C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t) \\
& \times \prod_{l\in\mathcal{M}_i(j)} [1-q_l(t)] \,\Bigg|\, \mathbf{\Lambda}(t)=\mathbf{\Lambda}^* \Bigg] + N_1(\epsilon). \qquad (62)
\end{aligned}$$

Using Lemma 2 again, the conditional expectation on the right-hand side of (62) can be expressed as an unconditional expectation plus a small term $N_2(\epsilon)$, leading us to

$$\begin{aligned}
\bar{c}_{ij} \le \mathbb{E}_{\mathbf{h}_i}\Bigg[C_{ij}(h_{ij}(t)p_{ij}(t))q_{ij}(t) \prod_{l\in\mathcal{M}_i(j)} [1-q_l(t)] \Bigg] \\
+ N_2(\epsilon) + N_1(\epsilon) \qquad (63)
\end{aligned}$$

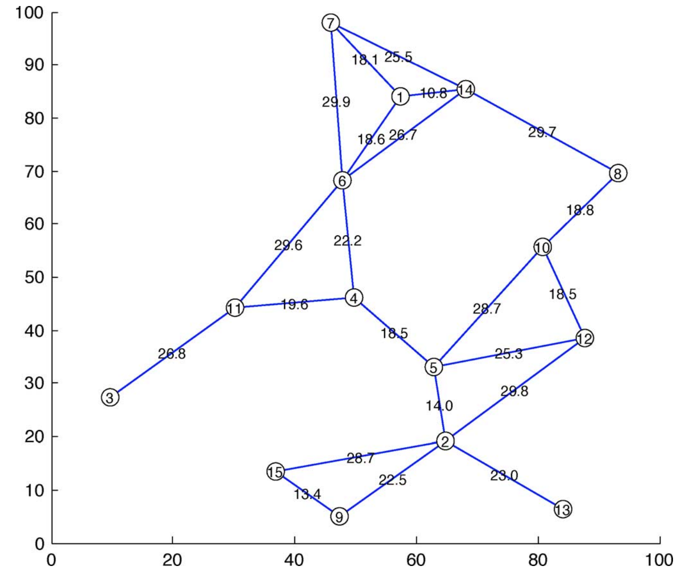

Fig. 4. Connectivity graph of a network with $n = 15$ terminals randomly placed in a square with side $L = 100$ meters. Terminals can communicate with neighbors whose distances are within 30 meters. The numbers on each edge shows the distance (in meters) between two communicating terminals.

where $\lim_{\epsilon\to 0} N_2(\epsilon) = 0$. Define $M(\epsilon) = N_1(\epsilon) + N_2(\epsilon)$ and substitute (44) into (63) to obtain (33). ■

V. Numerical Results

We illustrate performance of the proposed algorithm by implementing and simulating it over a network with $n = 15$ terminals randomly placed in a square with side $L = 100$ meters. Terminals can communicate with neighbors whose distances are within 30 meters. Numerical experiments here utilize the realization of this random placement shown in Fig. 4. Channel gains $h_{ij}(t)$ are Rayleigh distributed with mean $\bar{h}_{ij}$ and are independent across links and time. The average channel gain $\bar{h}_{ij} := \mathbb{E}[h_{ij}]$ follows an exponential pathloss law, $\bar{h}_{ij} = \alpha d_{ij}^{-\beta}$ with d_{ij} denoting the distance in meters between T_i and T_j and constants $\alpha = 10^{-1}\ \mathrm{m}^{-1}$ and $\beta = 2.5$. Assume the use of capacity achieving codes so that the instantaneous transmission rate takes the form

$$C_{ij}(h_{ij}(t)p_{ij}(t)) = \log\left(1 + \frac{h_{ij}(t)p_{ij}(t)}{N_0}\right) \qquad (64)$$

where N_0 is the channel noise set to $N_0 = 10^{-4}$ for all links. Fading channels are generated as i.i.d. There are two flows supported by the network, one from T_1 to T_2 and the other from T_8 to T_{11}. For each flow the minimum and maximum amount of information to be delivered are constrained by $a_i^{\min} = 0.1$ bits/s/Hz and $a_i^{\max} = 1$ bits/s/Hz for all nodes i. The routing and link capacity variables are bounded by $r_{ij}^{\min} = c_{ij}^{\min} = 0$ bits/s/Hz and $r_{ij}^{\max} = c_{ij}^{\max} = 1$ bits/s/Hz. The maximum average power consumption per terminal and maximum instantaneous power consumption per terminal are set to 2, i.e., $p_i^{\max} = p_{ij}^{\max} = 2$. Our objective is to maximize total amount of information delivered by the network, i.e., $U_i^k(a_i^k) = a_i^k$ and $V_i(p_i) = 0$. We set $\epsilon = 0.02$ and the simulation is conducted for 10^4 time slots. Successive convex approximation is used.

Fig. 5 shows feasibility of the proposed algorithm in terms of constraint violations. Specifically, $\bar{V}_{\lambda_i^k}(t)$, $\bar{V}_{\mu_{ij}}(t)$, $\bar{V}_{\nu_{ij}}(t)$ and

$\bar{V}_{\xi_i}(t)$, representing average violations of the flow conservation, link capacity, average rate and average power constraints, respectively, are presented in the figure. At each time t, we compute

$$\bar{V}_{\lambda_i^k}(t) = \frac{1}{t}\sum_{u=1}^{t}\left[\sum_{j\in\mathcal{N}(i)}\left(r_{ij}^k(u) - r_{ji}^k(u)\right) - a_i^k(u)\right] \tag{65}$$

$$\bar{V}_{\mu_{ij}}(t) = \frac{1}{t}\sum_{u=1}^{t}\left[c_{ij}(u) - \sum_{k\in\mathcal{K}} r_{ij}^k(u)\right] \tag{66}$$

$$\bar{V}_{\nu_{ij}}(t) = \frac{1}{t}\sum_{u=1}^{t}\Bigg[C_{ij}\big(h_{ij}(u)p_{ij}(u)\big)q_{ij}(u) \tag{67}$$

$$\times \prod_{k\in\mathcal{M}_i(j)}[1-q_k(u)] - c_{ij}(u)\Bigg] \tag{68}$$

$$\bar{V}_{\xi_i}(t) = \frac{1}{t}\sum_{u=1}^{t}\left[p_i(u) - \sum_{j\in\mathcal{N}(i)} p_{ij}(u)q_{ij}(u)\right]. \tag{69}$$

If the above values are nonnegative, it means the corresponding constraints are satisfied in an average sense. As we can see, after about 500 steps all constraints are satisfied within 10^{-2} tolerance. The average rate constraint takes the longest time to be satisfied (see Fig. 5(c)). This is because the transmission rate on link $T_i \to T_j$ depends not only on schedules and powers of T_i but also on those of T_j and his neighbors. This requires information to be received from, and propagated to, 2-hop networks.

To show optimality of the algorithm we compare ergodic primal and dual objectives. Since we are maximizing total admission control variables, the ergodic primal objective is

$$\mathsf{P}(t) = 1/t\sum_{u=1}^{t}\sum_{k\in\mathcal{K}} a_i^k(u). \tag{70}$$

Furthermore, upon defining average Lagrange multipliers as $\bar{\lambda}_i^k(t) = 1/t\sum_{u=1}^{t}\lambda_i^k(u)$, $\bar{\mu}_{ij}(t) = 1/t\sum_{u=1}^{t}\mu_{ij}(u)$, $\bar{\nu}_{ij}(t) = 1/t\sum_{u=1}^{t}\nu_{ij}(u)$ and $\bar{\xi}_i(t) = 1/t\sum_{u=1}^{t}\xi_i(u)$, we can compute the ergodic dual objective as

$$\mathsf{D}(t) = \mathsf{P}(t) + \sum_{i\in\mathcal{V}}\sum_{k\in\mathcal{K}}\bar{V}_{\lambda_i^k}(t)\bar{\lambda}_i^k(t) + \sum_{(i,j)\in\mathcal{E}}\bar{V}_{\mu_{ij}}(t)\bar{\mu}_{ij}(t) + \sum_{(i,j)\in\mathcal{E}}\bar{V}_{\nu_{ij}}(t)\bar{\nu}_{ij}(t) + \sum_{i\in\mathcal{V}}\bar{V}_{\xi_i}(t)\bar{\xi}_i(t). \tag{71}$$

Fig. 6(a) compares the ergodic primal and dual objectives. As time grows, the convergence of the proposed algorithm is observed as the primal and dual values approach each other. By Theorem 1, the algorithm is almost surely near optimal in the sense that the ergodic average of the utility almost surely converges to a value with optimality gap smaller than $\epsilon\hat{S}^2/2$ with respect to the optimal objective. Indeed, this is true as shown in Fig. 6(a) that the gap between primal and dual values becomes a small constant (about 0.05) as t increases. Moreover, we compute the correlation between $Q_1(t)$ and $Q_6(t)$ using samples from time 1 to t. The result is shown in Fig. 6(b). At the beginning, there is significant correlation between $Q_1(t)$ and $Q_6(t)$. But as time grows, the correlation vanishes and becomes negligible.

Optimal routes for flow 1 and 2 are shown in Fig. 7(a) and (b). In addition to the shortest path from source to destination, other longer paths are used to deliver information for both flows. For example, the shortest path for flow 2 is $T_8 \to T_{14} \to T_6 \to T_{11}$, but a longer path $T_8 \to T_{10} \to T_5 \to T_4 \to T_{11}$ is utilized as well. It is interesting to note that the longer path delivers more information than the shorter path does. This is because the shorter path goes through T_{14} and T_6 which interfere with the source node of flow 1 (T_1). To limit interference with flow 1, some packets in flow 2 are transmitted via other longer paths.

VI. Conclusion

We developed algorithms for optimal design of wireless networks using local channel state information. Due to the time-varying nature of fading states, random access is the natural medium access choice leading to the formulation of an optimization problem for random access networks. To obtain a distributed solution, we approximated the problem so that it can be decomposed in the dual domain and developed a stochastic subgradient descent algorithm. Based on instantaneous local channel conditions, the algorithm finds network operating points that are almost surely feasible and optimal in an ergodic sense. The solution exhibits a layered architecture in which variables in each layer are computed using information from interfaces to adjacent layers. The algorithm is fully distributed in that all operations necessary to achieve optimal operation are based on local information and information exchanges between neighboring terminals. The computational cost per iteration is minimal. In the proposed algorithm, all terminals act independently of each other. Algorithms that consider collaboration among terminals will be a future research direction.

Appendix A
Proof of Lemma 1

Define $g(t) := g(\mathbf{\Lambda}(t))$. According to [21, Theorem 2], for arbitrary $\delta > 0$, $g(t) - \tilde{\mathsf{D}}$ falls below $\epsilon\hat{S}^2/2 + \delta$ at least once almost surely as t grows. If $g(t) - \tilde{\mathsf{D}}$ falls below $\epsilon\hat{S}^2/2+\delta$, it may stay below or jump above $\epsilon\hat{S}^2/2+\delta$. The key idea in this proof is to show that if $g(t)$ exceeds $\tilde{\mathsf{D}} + \epsilon\hat{S}^2/2 + \delta$ the probability that it gets even bigger is very small. Let us then define T_1 as a time at which $g(T_1)$ stays below $\tilde{\mathsf{D}} + \epsilon\hat{S}^2/2 + \delta$ but jumps above it at time T_1+1, i.e., $g(T_1+1) - \tilde{\mathsf{D}} > \epsilon\hat{S}^2/2 + \delta$. The rest of the proof relies on the following chain of arguments:

A1) The expected value of the distance between $\mathbf{\Lambda}(T_1+1)$ and the optimal dual variable $\mathbf{\Lambda}^*$ is bounded by a function $L_0(\epsilon)$ where $\lim_{\epsilon\to 0} L_0(\epsilon) = 0$, i.e.,

$$\mathbb{E}[\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2] \le L_0(\epsilon). \tag{72}$$

A2) Define $g_{\text{best}}(t) = \min_{u\in[0,t]} g(u)$ and $\psi(t) = \|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 \mathbb{I}\{g_{\text{best}}(T_1+t) - \tilde{\mathsf{D}} \ge \epsilon\hat{S}^2/2\}$ for $t = 1, 2, \ldots$ and $\mathbb{I}\{\cdot\}$ denotes the indicator function. Then, $\psi(t)$ is a supermartingale, i.e.,

$$\mathbb{E}[\psi(t+1)\,|\,\psi(1:t)] \le \psi(t). \tag{73}$$

A3) Assume $L_0(\epsilon)$ is small enough such that $L_0(\epsilon) < \sqrt{L_0(\epsilon)}$. Define then a stopping rule $\psi(t) \ge \sqrt{L_0(\epsilon)}$ or $\psi(t) = 0$. Let T be a stopping time, by the optional stopping theorem [23, Theorem 10.10] we have

$$\mathbb{E}[\psi(T)] \le \mathbb{E}[\psi(1)]. \tag{74}$$

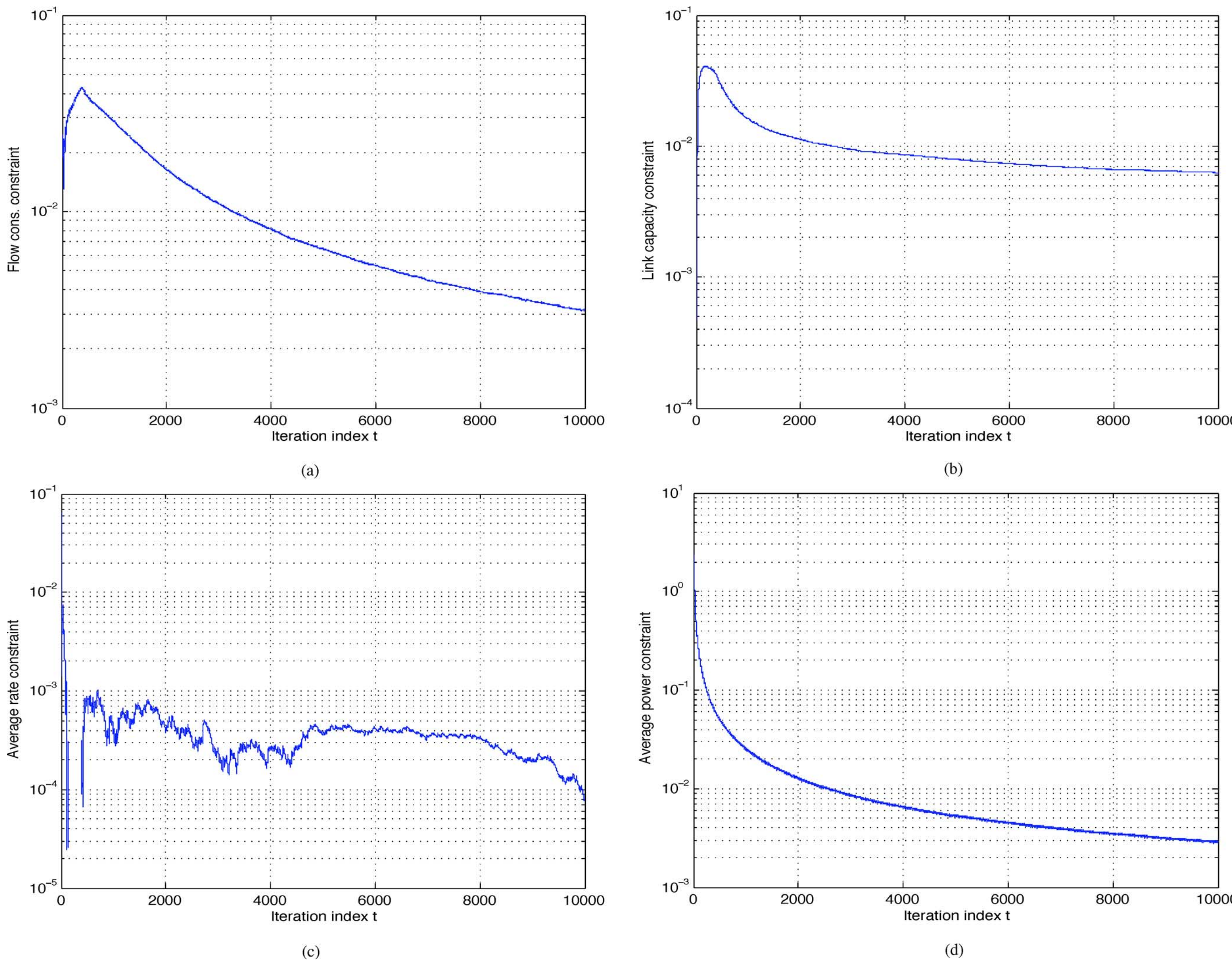

Fig. 5. Feasibility. After about 500 steps, all constraints are satisfied in an ergodic sense within 10^{-2} tolerance. The average rate constraint takes the longest time to be satisfied. This is because the transmission rate on link $T_i \rightarrow T_j$ depends not only on schedules and powers of T_i but also on those of T_j and neighbors of T_j. This requires information to be received from, and propagated to, 2-hop neighbors. (a) Flow conservation constraint, (b) Link capacity constraint, (c) Average rate constraint, (d) Average power constraint.

Using the fact that $\psi(1) = \|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2$ and results in (72) we can further bound (74) by

$$\mathbb{E}[\psi(T)] \leq \mathbb{E}[\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2] \leq L_0(\epsilon). \tag{75}$$

According to the stopping rule, either $\psi(T) \geq \sqrt{L_0(\epsilon)}$ or $\psi(T) = 0$. As a result, we can lower bound $\mathbb{E}[\psi(T)]$ by

$$\mathbb{E}[\psi(T)] \geq \sqrt{L_0(\epsilon)}\Pr[\|\mathbf{\Lambda}(T_1+T) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_0(\epsilon)}|\mathbf{\Lambda}(T_0)]. \tag{76}$$

Substituting (76) into (75) and dividing both sides by $\sqrt{L_0(\epsilon)}$ yields

$$\Pr[\|\mathbf{\Lambda}(T_1+T) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_0(\epsilon)}|\mathbf{\Lambda}(T_0)] \leq \sqrt{L_0(\epsilon)}. \tag{77}$$

A4) For any $t > 0$, the event $\|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_0(\epsilon)}$ happens only when there exists $T \leq t$ such that T is a stopping time and $\|\mathbf{\Lambda}(T_1+T) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_0(\epsilon)}$. Then, we have

$$\begin{aligned}&\Pr[\|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_1(\epsilon,\delta)}|\mathbf{\Lambda}(T_0)] \\ &\leq \Pr[\|\mathbf{\Lambda}(T_1+T) - \mathbf{\Lambda}^*\|^2 \geq \sqrt{L_0(\epsilon)}|\mathbf{\Lambda}(T_0)] \\ &\leq \sqrt{L_0(\epsilon)}\end{aligned} \tag{78}$$

where the second inequality follows from (77). Substituting $L(\epsilon) = \sqrt{L_0(\epsilon)}$ into (78) completes the proof. In the following, we provide detailed proofs for A1) and A2).

First, we show that (72) is true, i.e., $\mathbb{E}[\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2] \leq L_0(\epsilon)$. Start by noting that $g(t)$ is a convex function of $\mathbf{\Lambda}(t)$ with a unique minimizer $\mathbf{\Lambda}^*$, then $g(T_1) - \tilde{\mathsf{D}} \leq \epsilon\hat{S}^2/2 + \delta$ is equivalent to

$$\|\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^*\|^2 \leq L_1(\epsilon\hat{S}^2/2 + \delta) \tag{79}$$

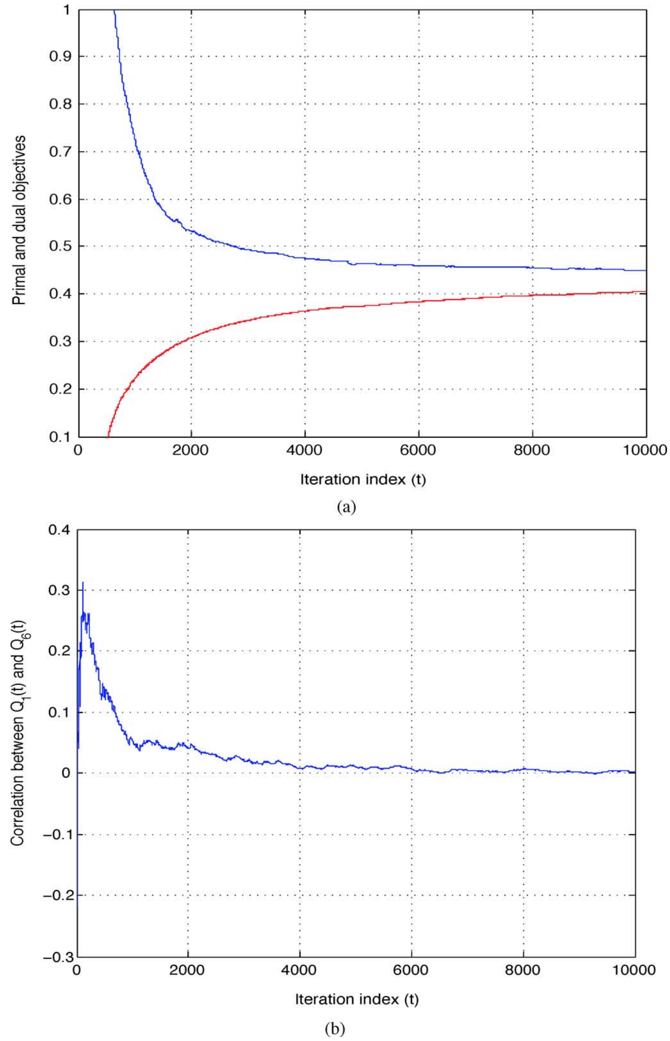

Fig. 6. (a) Optimality. As time grows, primal and dual objectives approach each other. (b) Correlation between $Q_1(t)$ and $Q_6(t)$. At the beginning, there is significant correlation between $Q_1(t)$ and $Q_6(t)$. But as time grows, the correlation vanishes and becomes negligible.

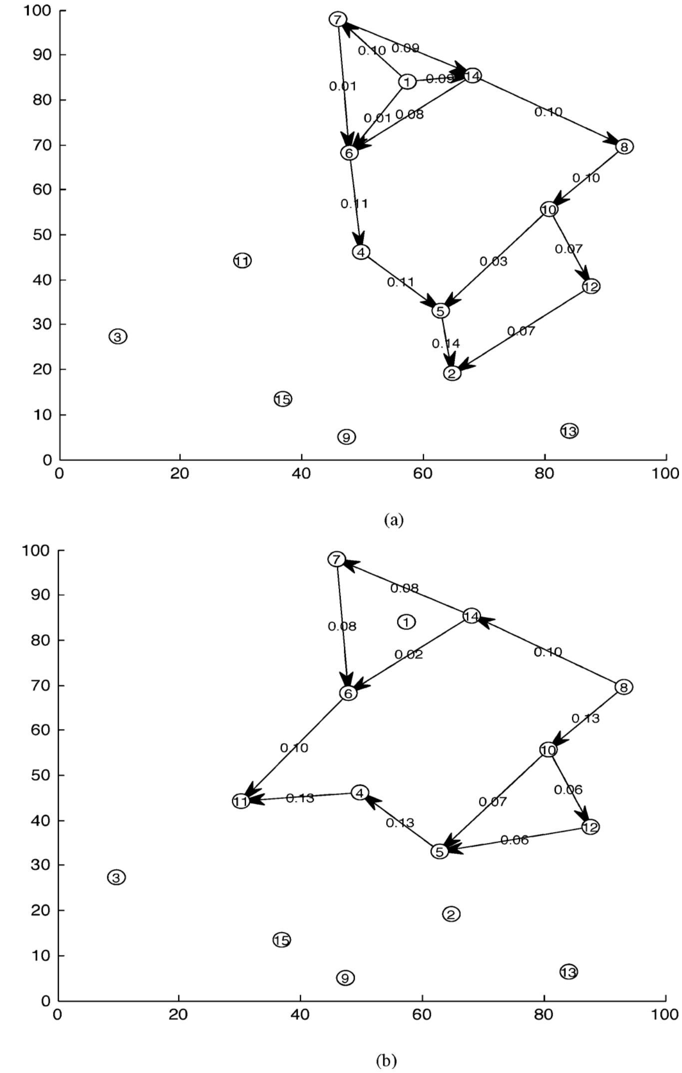

Fig. 7. Optimal routes for flow 1 (from T_1 to T_2) and flow 2 (from T_8 to T_{11}). (a) Flow 1: from T_1 to T_2, (b) Flow 2: from T_8 to T_{11}.

where $L_1(\cdot)$ is a nonnegative function such that $\lim_{x\to 0} L_1(x) = 0$. According to the dual update (17), we can write $\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2$ as

$$\begin{aligned}
&\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2 \\
&\quad = \|[\mathbf{\Lambda}(T_1) - \epsilon\mathbf{s}(T_1)]^+ - \mathbf{\Lambda}^*\|^2 && (80)\\
&\quad \le \|\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^* - \epsilon\mathbf{s}(T_1)\|^2 && (81)\\
&\quad = \|\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^*\|^2 + \epsilon^2\|\mathbf{s}(T_1)\|^2 \\
&\qquad - 2\epsilon\mathbf{s}^T(T_1)[\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^*], && (82)
\end{aligned}$$

where inequality (81) follows because setting negative elements in $\mathbf{\Lambda}(T_1) - \epsilon\mathbf{s}(T_1)$ to zero reduces its distance to $\mathbf{\Lambda}^*$. Expanding (81) yields (82). Taking expectation conditioned on $\mathbf{\Lambda}(T_1)$ for both sides of (82) yields

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2|\mathbf{\Lambda}(T_1)] \\
&\quad \le \|\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^*\|^2 + \epsilon^2\mathbb{E}[\|\mathbf{s}(T_1)\|^2|\mathbf{\Lambda}(T_1)] \\
&\qquad - 2\epsilon\mathbb{E}[\mathbf{s}^T(T_1)\,|\,\mathbf{\Lambda}(T_1)][\mathbf{\Lambda}(T_1) - \mathbf{\Lambda}^*]. && (83)
\end{aligned}$$

Note that the first term on the right-hand side of (83) is upper bounded by $L_1(\epsilon\hat{S}^2/2 + \delta)$ [cf. (79)]. As per the hypothesis, $\mathbb{E}[\|\mathbf{s}(T_1)\|^2\,|\,\mathbf{\Lambda}(T_1)]$ is upper bounded by $\hat{S}^2$. The third term is lower bounded by 0 because $\mathbb{E}[\mathbf{s}(T_1)\,|\,\mathbf{\Lambda}(T_1)]$ is subgradient of $g(\mathbf{\Lambda}(T_1))$ [21, Proposition 1]. Plugging these bounds into (83) yields

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{\Lambda}(T_1+1) - \mathbf{\Lambda}^*\|^2|\mathbf{\Lambda}(T_1)] \\
&\qquad \le L_1(\epsilon\hat{S}^2/2 + \delta) + \epsilon^2\hat{S}^2 := L_2(\epsilon,\delta) && (84)
\end{aligned}$$

where we defined function $L_2(\epsilon,\delta)$. Taking expectation with respect to $\mathbf{\Lambda}(T_1)$ on both sides of (84) and defining $L_0(\epsilon) = \lim_{\delta\to 0} L_2(\epsilon,\delta)$ lead us to (72).

We then show $\psi(t)$ is a supermartingale. We discuss two cases $\psi(t) = 0$ and $\psi(t) > 0$ separately. If $\psi(t) = 0$, it implies either $\mathbf{\Lambda}(T_1+t) = \mathbf{\Lambda}^*$ or $g_{\text{best}}(T_1+t) - \tilde{\mathsf{D}} < \epsilon\hat{S}^2/2$. If $\mathbf{\Lambda}(T_1+t) = \mathbf{\Lambda}^*$, then it must be $g(T_1+t) = \tilde{\mathsf{D}}$. Since the dual function is lower bounded by $\tilde{\mathsf{D}}$, it implies $g_{\text{best}}(T_1+t+1) = g_{\text{best}}(T_1+t) = \tilde{\mathsf{D}}$. If $g_{\text{best}}(T_1+t) - \tilde{\mathsf{D}} < \epsilon\hat{S}^2/2$, it follows that $g_{\text{best}}(T_1+t+1) - \tilde{\mathsf{D}} < \epsilon\hat{S}^2/2$ since $g_{\text{best}}(T_1+t+1) \le g_{\text{best}}(T_0+t)$. In either case, $\psi(t+1) = 0$ and (73) holds for equality. If $\psi(t) \neq 0$, it must be $g_{\rm best}(T_1+t) - \tilde{D} \geq \epsilon \hat{S}^2/2$, which implies $g(T_1+t) - \tilde{\mathsf{D}} \geq \epsilon \hat{S}^2/2$ and $\psi(t) = \|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2$. Since $\psi(t)$ is completely determined by $\mathbf{\Lambda}(t)$, we can write following relationship:

$$\begin{aligned}
&\mathbb{E}[\psi(t+1)\,|\,\psi(1:t)] \\
&\quad = \mathbb{E}[\psi(t+1)\,|\,\mathbf{\Lambda}(T_1+1:T_1+t)] && (85)\\
&\quad \leq \mathbb{E}[\|\mathbf{\Lambda}(T_1+t+1) - \mathbf{\Lambda}^*\|^2\,|\,\mathbf{\Lambda}(T_1+1:T_1+t)] && (86)\\
&\quad = \mathbb{E}[\|\mathbf{\Lambda}(T_1+t+1) - \mathbf{\Lambda}^*\|^2\,|\,\mathbf{\Lambda}(T_1+t)] && (87)
\end{aligned}$$

where inequality (86) follows because $\psi(t+1) = \|\mathbf{\Lambda}(T_1+t+1) - \mathbf{\Lambda}^*\|^2 \mathbb{I}\{g_{\rm best}(T_1+t+1) - \tilde{\mathsf{D}} \geq \epsilon \hat{S}^2/2\} \leq \|\mathbf{\Lambda}(T_1+t+1) - \mathbf{\Lambda}^*\|^2$ and equality (87) is true since $\mathbf{\Lambda}(\mathbb{N})$ is a Markov process. Using the dual update rule (17) we can bound (87) by

$$\begin{aligned}
&\mathbb{E}[\psi(t+1)\,|\,\psi(1:t)] \\
&\quad \leq \|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 + \epsilon^2 \mathbb{E}[\|\mathbf{s}(T_1+t)\|^2\,|\,\mathbf{\Lambda}(T_1+t)] \\
&\qquad - 2\epsilon \mathbb{E}[\mathbf{s}^T(T_1+t)\,|\,\mathbf{\Lambda}(T_1+t)][\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*] && (88)\\
&\quad \leq \|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 + \epsilon \hat{S}^2 - 2\epsilon[g(T_1+t) - \tilde{\mathsf{D}}] && (89)\\
&\quad \leq \|\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*\|^2 = \psi(t). && (90)
\end{aligned}$$

where (89) follows because $\mathbb{E}[\|\mathbf{s}(T_1+t)\|^2\,|\,\mathbf{\Lambda}(T_1+t)] \leq \hat{S}^2$ and $\mathbb{E}[\mathbf{s}^T(T_1+t)\,|\,\mathbf{\Lambda}(T_1+t)][\mathbf{\Lambda}(T_1+t) - \mathbf{\Lambda}^*]$ is lower bounded by $g(T_1+t) - \tilde{\mathsf{D}}$ and (90) follows from the fact that $g(T_1+t) - \tilde{\mathsf{D}} \geq \epsilon \hat{S}^2/2$. Therefore, for both cases $\psi(t) = 0$ and $\psi(t) > 0$ (73) holds true.

Appendix B
Proof of Lemma 3

For notational simplicity, we ignore time index t in this proof. Recall that q_i is uniquely determined by $\mathbf{h}_i$ and $\mathbf{\Lambda}_i$. Thus, we can write q_i as a function of $\mathbf{h}_i$ and $\mathbf{\Lambda}_i$, i.e., $q_i(\mathbf{h}_i, \mathbf{\Lambda}_i)$. To show $\mathbb{E}_{\mathbf{h}_i}[q_i\,|\,\mathbf{\Lambda}_i]$ is continuous in $\mathbf{\Lambda}_i$, we have to establish that for any sequence $\mathbf{\Lambda}_i(n)$ that converges to $\mathbf{\Lambda}_i$ as $n \to \infty$, $\mathbb{E}_{\mathbf{h}_i}[q_i\,|\,\mathbf{\Lambda}_i(n)]$ converges to $\mathbb{E}_{\mathbf{h}_i}[q_i\,|\,\mathbf{\Lambda}_i]$, i.e.,

$$\lim_{n\to\infty} \int q_i(\mathbf{h}_i, \mathbf{\Lambda}_i(n))\, d\mathbf{h}_i = \int q_i(\mathbf{h}_i, \mathbf{\Lambda}_i)\, d\mathbf{h}_i. \tag{91}$$

To show (91) is true, define

$$W_{ij}(\mathbf{h}_i, \mathbf{\Lambda}_i) = \max_{p \in [0, p_{ij}^{\max}]} \{\alpha_{ij} C_{ij}(h_{ij} p) - \beta_i - \xi_i p\} \tag{92}$$

and

$$W_i(\mathbf{h}_i, \mathbf{\Lambda}_i) = \max_{j \in \mathcal{N}(i)} \{W_{ij}(\mathbf{h}_i, \mathbf{\Lambda}_i)\}. \tag{93}$$

Note that the objective on the right-hand side of (92) is a linear function of $\mathbf{\Lambda}_i$. Given $\mathbf{h}_i$, $W_{ij}(\mathbf{h}_i, \mathbf{\Lambda}_i)$ is the maximum of a set of linear functions of $\mathbf{\Lambda}_i$. As a consequence, $W_{ij}(\mathbf{h}_i, \mathbf{\Lambda}_i)$ is a convex function of $\mathbf{\Lambda}_i$ given $\mathbf{h}_i$. Moreover, note that $W_i(\mathbf{h}_i, \mathbf{\Lambda}_i)$ is the maximum of $W_{ij}(\mathbf{h}_i, \mathbf{\Lambda}_i)$ for all $j \in \mathcal{N}(i)$, then given $\mathbf{h}_i$ it is a convex function of $\mathbf{\Lambda}_i$ as well. Since convexity implies continuity, $W_i(\mathbf{h}_i, \mathbf{\Lambda}_i)$ is a continuous function of $\mathbf{\Lambda}_i$ for any given $\mathbf{h}_i$. This implies

$$\lim_{n\to\infty} W_i(\mathbf{h}_i, \mathbf{\Lambda}_i(n)) = W_i(\mathbf{h}_i, \mathbf{\Lambda}_i). \tag{94}$$

Recall that $q_i(\mathbf{h}_i, \mathbf{\Lambda}_i)$ equals to 1 if $W_i(\mathbf{h}_i, \mathbf{\Lambda}_i) > 0$ and 0 otherwise. Therefore, $q_i(\mathbf{h}_i, \mathbf{\Lambda}_i(n))$ converges pointwise to $q_i(\mathbf{h}_i, \mathbf{\Lambda}_i)$ almost everywhere. Furthermore, note that $q_i(\mathbf{h}_i, \mathbf{\Lambda}_i(n))$ is upper bounded by 1. Using dominated convergence theorem [23, Ch. 5.9], (91) follows. The argument for the continuity of the expectation $\mathbb{E}_{\mathbf{h}_i}[q_{ij} C_{ij}(h_{ij} p_{ij})\,|\,\mathbf{\Lambda}_i]$ is analogous.

References

[1] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," in *Proc. IEEE*, Cambridge, U.K., Jan. 2007, no. 1, pp. 255–312.
[2] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, Aug. 2006.
[3] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Netw.*, vol. 1, no. 1, pp. 1–144, 2006.
[4] K. Kar, S. Sarkar, and L. Tassiulas, "Achieving proportional fairness using local information in aloha networks," *IEEE Trans. Autom. Control*, vol. 4, no. 10, pp. 1858–1862, Oct. 2007.
[5] X. Wang and K. Kar, "Cross-layer rate optimization for proportional fairness in multihop wireless networks with random access," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1548–1559, Aug. 2006.
[6] J. Lee, M. Chiang, and A. R. Calderbank, "Utility-optimal random-access control," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2741–2751, July 2007.
[7] Y. Yu and G. B. Giannakis, "Cross-layer congestion and contention control for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 37–42, Jan. 2008.
[8] C. Long, B. Li, Q. Zhang, B. Zhao, B. Yang, and X. Guan, "The end-to-end rate control in multiple-hop wireless networks: Cross-layer formulation and optimal allocation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 719–731, May 2008.
[9] A. H. Mohsenian-Rad, J. Huang, M. Chiang, and V. W. S. Wong, "Utility-optimal random access: Reduced complexity, fast convergence, and robust performance," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 898–911, Feb. 2009.
[10] A. H. Mohsenian-Rad, J. Huang, M. Chiang, and V. W. S. Wong, "Utility-optimal random access without message passing," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1073–1079, Mar. 2009.
[11] Y. Yu and G. B. Giannakis, "Opportunistic medium access for wireless networking adapted to decentralized CSI," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1445–1455, June 2006.
[12] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: An optimal stopping approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 205–222, Jan. 2009.
[13] S. Adireddy and L. Tong, "Expoiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.
[14] X. Qin and R. A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.
[15] K. Bai and J. Zhang, "Opportunistic multichannel aloha: Distributed multiaccess control scheme for OFDMA wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 848–855, May 2006.
[16] Y. Xue, T. Kaiser, and A. B. Gershman, "Channel-aware aloha-based OFDM subcarrier assignment in single-cell wireless communications," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 953–962, May 2007.
[17] M. H. Ngo, V. Krishnamurthy, and L. Tong, "Optimal channel-aware aloha protocol for random access in wlans with multipacket reception and decentralized channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2575–2588, June 2008.
[18] Y. Hu and A. Ribeiro, "Adaptive distributed algorithms for optimal random access channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2703–2715, Aug. 2011.
[19] G. Miao, G. Y. Li, and A. Swami, "Decentralized optimization for multichannel random access," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3012–3023, Oct. 2009.
[20] A. Ribeiro and G. B. Giannakis, "Separation principles of wireless networking," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4488–4505, Sept. 2010.
[21] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6369–6386, Dec. 2010.
[22] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[23] D. Williams, *Probability With Martingales*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

Yichuan Hu (S'10) received the B.Eng. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2004 and 2007, respectively, and the M.S. degree in electrical and computer engineering from the University of Delaware, Newark, in 2009.

Since 2009, he has been working towards the Ph.D. degree in the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia. From June to August 2009, he was a research intern in the Standards Research Laboratory at Samsung Telecommunications America, Dallas, TX. From June to August 2010, he was a summer associate in the Quantitative Trading Group at the Bank of America Merrill Lynch, New York. His research interests include signal processing, optimization, and machine learning.

Alejandro Ribeiro (M'10) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, in 1998 and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, in 2005 and 2007, respectively.

From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. He was with the Department of Electrical and Computer Engineering, University of Minneapolis from 2003 to 2008. Since 2008, he has been with the Department of Electrical and Systems Engineering at the University of Pennsylvania, Philadelphia, where he is currently an Assistant Professor. His research interests lie in the areas of communication, signal processing, and networking. His current research focuses on the study of networked phenomena arising in technological, human, and natural networks.

Dr. Ribeiro received the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and student paper awards at ICASSP 2005 and ICASSP 2006. He is also a Fulbright Scholar.