Separation Principles in Wireless Networking

Alejandro Ribeiro, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—A general wireless networking problem is formulated whereby end-to-end user rates, routes, link capacities, transmit-power, frequency, and power resources are jointly optimized across fading states. Even though the resultant optimization problem is generally nonconvex, it is proved that the gap with its Lagrange dual problem is zero, so long as the underlying fading distribution function is continuous. The major implication is that separating the design of wireless networks in layers and per-fading state subproblems can be optimal. Subgradient descent algorithms are further developed to effect an optimal separation in layers and layer interfaces.**

***Index Terms*—Fading, Lagrangian duality, optimization, wireless networking.**

I. Introduction

THE design goal of communication networks is to transport information from a generating source to an intended destination. Of the many ways this can be accomplished, there are networks designed to maximize properly selected optimality criteria. This paper is concerned with basic principles governing the associated optimization problem for wireless communication networks in the presence of fading.

A. Related Work

Optimization as a mathematical tool to analyze network protocols appeared first in the network utility maximization (NUM) framework independently proposed by [1] and [2]. The gist of these works is that congestion control protocols can be viewed as distributed implementations of algorithms that solve utility maximization problems. Source rates are regarded as primal variables and congestion parameters constitute variables of the corresponding Lagrange dual problems. Recursive schemes updating these variables boil down to subgradient descent iterations on the dual function—the kind of optimization algorithm also known as dual decomposition. The connection between congestion control and NUM has been fruitful in understanding which NUM problems are solvable by heuristic congestion control schemes and also for introducing protocols as solvers of suitably formulated NUM problems; see, e.g., [3], [4], and [5] for recent accounts on NUM.

Extending the NUM framework to wireless networking problems is not a simple pursuit. Different from wireline networks where pairs of nodes are individually connected at fixed capacities, node links in wireless networks are not predetermined and their capacities are not fixed. Rather, connections and link capacities are variables of the optimization problem itself. Nonetheless, wireless physical layer models have been incorporated into NUM formulations, and arguably represent one of the most promising research directions on cross-layer network designs; see, e.g., [6]–[10] and references therein.

Although some recent works advocate alternative decomposition methods [11]–[13], most of the NUM literature relies on dual decomposition. This is because the associated Lagrangian function exhibits a separable structure reminiscent of layered network designs. This has been pointed out in, e.g., [14], [15], and recently popularized by [16]. Indeed, using NUM formulations, it is possible to understand specific layered architectures as decompositions of optimization problems [16]. On this issue, though, it is important to stress that since wireless networking problems are nonconvex, the duality gap is generally nonzero. As a result, the dual optimum is generally different from the primal optimum and for this reason layering is believed to come at the price of optimality loss.

A different approach to the wireless networking problem is the work on stochastic network optimization, originally reported by [17] and subsequently extended in [18] and [19]; see also [20] for a comprehensive treatment and [21] for a related approach. These works build on the backpressure algorithm [22]–[24], and extend it to deal with a finite number of random network states as those generated by fading. While the original backpressure algorithm was introduced to stabilize all queues in the network without further optimality considerations, [17]–[19] developed modified versions that can provably approximate solutions of various wireless networking problems. These modified versions are obtained after observing a similarity between Lagrange multiplier updates in dual decomposition and evolution of (virtual) queue lengths in a communication network.

All of the approaches outlined so far pertain to either wireless networks with deterministic links, or, if random fading effects are accounted for, the channel links are confined to take on a *finite* number of values. This is simply not the case in wireless fading propagation and, given subtleties involved in limits of stochastic processes, it cannot be regarded as practically irrelevant. Further motivation to consider fading channels with *infinite* number of states is provided by the fact that when this has been incorporated in related settings it has actually turned out to yield a simpler problem.

Rate utility maximization of time or frequency division multiple access (TDMA/FDMA) with fading coefficients taking on

Manuscript received February 01, 2008; revised September 21, 2009. Date of current version August 18, 2010. The work in this paper was prepared through collaborative participation in the Communications and Networks Consortium supported by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The material in this paper was presented in part at the CISS 2008, Princeton, NJ, March 2008.

A. Ribeiro is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: aribeiro@seas.upenn.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Communicated by R. A. Berry, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2010.2053897

a continuum of values was considered in [25]. To deal with this non-convex problem, time sharing of carriers is introduced, a strategy that renders optimization tractable but complicates implementation. However, it is observed that with probability 1 time sharing is not needed. This result does not hold when the number of fading states is finite [26]. In this sense, TDMA and FDMA problems become simpler when fading takes on an infinite number of states. Another related problem is that of optimal subcarrier allocation at the physical layer of digital subscriber lines (DSL). Here too the problem is nonconvex, but when formulated in a continuous (as opposed to discrete) frequency-domain the duality gap becomes zero. This was first observed for sum-rate maximization in [27] and extended to general utilities by [28] as an application of Lyapunov's convexity theorem [29].

Different from NUM approaches [6]–[10], [14], [15] and stochastic network optimization schemes [17]–[20], the goal of the present paper is not to put forth or analyze a specific algorithm but to look at a structural property—the duality gap—of the underlying optimization problem. Distinct from e.g., [16], the objective is not to argue whether layered architectures can be formally understood as decompositions of optimization problems. Our goal is to prove that layering can be optimal in wireless networks. Instead of the specific medium access control problem treated by [25] or the DSL spectrum assignment dealt with in [27] and [28], this work considers a general wireless networking problem.

B. Organization and Contributions

The paper starts by modeling the relation among wireless networking variables in Section II. The effects of random fading in determining link capacities and power consumption are specified in Section II-A. Using these relations, the optimal wireless networking problem is formulated in Section III. This problem is non-convex and therefore belongs to the unfortunate side of the watershed division between "easy" and "difficult" optimization problems. The Lagrange dual of this problem is introduced in Section III.A.

The main result of the paper is presented in Section IV. Albeit nonconvex, Theorem 1 asserts that the optimization problem of Section III exhibits zero duality gap. The bulk of Section IV is dedicated to prove this claim. The result has far reaching implications that are yet to be explored fully. These stem from the fact that it is easier to work with the dual problem than with the original primal problem. The zero duality gap ensures that the dual problem incurs no loss of optimality.

In Section V, the separable structure of the Lagrangian is used to establish two separation principles:

Separation per layer. This result proved in Theorem 3 establishes optimal separation of wireless networking problems into layers. Separate optimization subproblems can yield optimal routes, optimal link capacity allocations as well as optimal power and frequency assignments.

Separation per fading state. This result proved in Theorem 4 asserts that the network optimization problem is further separable in per-fading-state subproblems.

The final section of the paper deals with subgradient descent algorithms enabling optimal designs of wireless networks (Section VI-A) and demonstrates how these algorithms effect separability in layers and layer interfaces (Section VI-B).

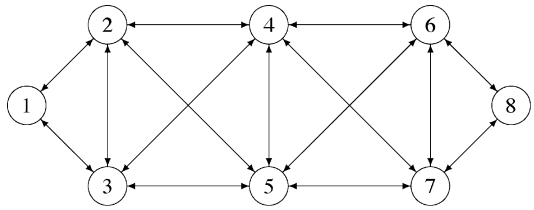

Fig. 1. Connectivity graph of an example wireless network.

II. Problem Formulation

Consider an *ad-hoc* wireless network comprising J user terminals $\{T_i\}_{i=1}^J$. Terminal T_i wishes to deliver packets for different application-level flows generically denoted by k. The flow k specifies the destination of the flow's packets, but the same destination may be associated with different flows to, e.g., accommodate different types of traffic (video, voice, or data). The destination of flow k is denoted as T^k to emphasize that flow indexing is different from terminal indexing. For every flow k, packet arrivals at T_i form a stationary stochastic process with mean a_i^k.

Network connectivity is modeled as a graph $\mathcal{G}(v,e)$ with vertices $v \in \mathcal{V} := \{1,\ldots,J\}$ and edges $e \in \mathcal{E}$ connecting pairs of vertices (i,j) when and only when T_i and T_j can communicate with each other; see Fig. 1. The neighborhood of i is denoted by $n(i) := \{j : (i,j) \in \mathcal{E}\}$. Each terminal $\{T_j\}_{j\in n(i)}$ that can communicate with T_i will be referred to as a neighbor. Given this model, terminals rely on multihop transmissions to deliver packets to the intended destination T^k of the flow k. For that matter, T_i selects an average rate r_{ij}^k for transmitting kth flow packets to T_j. Assuming that packets are not discarded and queues are stable throughout the network, average rates a_i^k of exogenous packet arrivals (from the application layer) are related with endogenous (to the network layer) average rates r_{ij}^k transmitted to and from neighboring nodes. Endogenous and exogenous rates are related through the flow conservation equation per flow k as (see, e.g., [20])

$$a_i^k = \sum_{j\in n(i)} \left(r_{ij}^k - r_{ji}^k\right) \tag{1}$$

which we interpret as stating that the rate a_i^k offered to the application is the difference between the outgoing rates r_{ij}^k and the incoming rates r_{ji}^k. Consider now the average rates r_{ij}^k of all flows k traversing the link $T_i \to T_j$. Letting c_{ij} denote the information capacity of this link, queue stability is ensured by requiring [20]

$$\sum_k r_{ij}^k \leq c_{ij}. \tag{2}$$

The constraints in (1) and (2) are basic in describing traffic flow over a wireline network with fixed capacities c_{ij}. In this setting, T_i needs to determine exogenous arrival rates a_i^k and transmission rate variables r_{ij}^k to satisfy certain optimality criteria. In a wireless network however, c_{ij} is not a fixed resource given to the terminals. In fact, operating conditions are determined by

a set of available frequencies (tones) $\mathcal{F}$ and prescribed powers $p_{\max i}$. Thus, in addition to a_i^k and r_{ij}^k, terminal T_i has to decide how to split its power budget $p_{\max i}$ among tones $f \in \mathcal{F}$ and neighbors T_j, $j \in n(i)$. Matters are further complicated by fading propagation effects as described in the next section.

A. Link Capacities and Transmit-Powers

For every frequency $f \in \mathcal{F}$ and $(i,j) \in \mathcal{E}$, let h_{ij}^f denote the channel gain from T_i to T_j. As is customary practice in wireless communications h_{ij}^f is modeled as the realization of a random variable H_{ij}^f. Channel realizations of all network links are collected in the vector $\boldsymbol{h}$ with the corresponding random variable denoted as $\mathbf{H}$. The range of $\mathbf{H}$ is the set $\boldsymbol{\mathcal{H}}$.

Let $p_{ij}^f(\boldsymbol{h})$ denote the power used by T_i for sending packets to T_j on the tone f when the channel vector realization is $\boldsymbol{h}$. Consequently, the instantaneous total power $p_i(\boldsymbol{h})$ used by T_i is the sum of the power used to transmit to all selected neighbors on all selected tones, i.e.

$$p_i(\boldsymbol{h}) := \sum_{j \in n(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\boldsymbol{h}). \tag{3}$$

Averaging over all possible channel realizations yields the average power used by T_i as

$$p_i := \mathbb{E}_{\mathbf{H}}[p_i(\boldsymbol{h})] = \mathbb{E}_{\mathbf{H}}\left[\sum_{j \in n(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\boldsymbol{h})\right] \tag{4}$$

where $\mathbb{E}_{\mathbf{H}}[\cdot]$ denotes expectation over the channel probability distribution function.

The rate of information transmission over the $T_i \to T_j$ link is a function of the power distribution $p_{ij}^f(\boldsymbol{h})$ and the channel realization $\boldsymbol{h}$. To maintain generality of the model, define a function $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$ to map channels and powers to link capacities so that the capacity $c_{ij}^f(\boldsymbol{h})$ of the link $T_i \to T_j$ on the tone f is

$$c_{ij}^f(\boldsymbol{h}) := C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})). \tag{5}$$

Function $C(\cdot)$ is determined by the capabilities and operating conditions of the terminals. If, e.g., terminals perform single user detection, link capacity is determined by the signal-to-interference-plus-noise ratio (SINR). Interference $\iota_{ij}^f(\boldsymbol{h})$ to the $T_i \to T_j$ link comes from: i) terminals T_k in T_j's neighborhood $n(j)$ transmitting to any of its neighbors $T_l \in n(k)$; and ii) transmissions of T_i itself to terminals other than T_j. Therefore

$$\begin{aligned}\iota_{ij}^f(\boldsymbol{h}) &:= \sum_{\substack{k \in n(j) \\ k \neq i}} \sum_{l \in n(k)} h_{kj}^f p_{kl}^f(\boldsymbol{h}) + \sum_{\substack{l \in n(i) \\ l \neq j}} h_{il}^f p_{il}^f(\boldsymbol{h}) \\ &= \sum_{(k,l) \neq (i,j)} h_{kj}^f p_{kl}^f(\boldsymbol{h})\end{aligned} \tag{6}$$

where $(k,l) \neq (i,j)$ in the last sum signifies all pairs (k,l) different from (i,j) with proper neighborhood restrictions. Note that the first sum includes terms of the form $h_{jj}^f p_{jl}^f(\boldsymbol{h})$ to account for the interference of T_j's transmissions to packets received at T_j. Typically h_{jj}^f is very large discouraging transmission and reception of packets over the same tone f. This is not prevented a fortiori but will likely emerge from the network optimization to be described later. With interference $\iota_{ij}^f(\boldsymbol{h})$ as in (6) and σ_j^f denoting the noise power at T_j's receiver end, the SINR of the $T_i \to T_j$ link on tone f is

$$\gamma_{ij}^f(\boldsymbol{h}) := \frac{h_{ij}^f p_{ij}^f(\boldsymbol{h})}{\sigma_j^f + \iota_{ij}^f(\boldsymbol{h})} = \frac{h_{ij}^f p_{ij}^f(\boldsymbol{h})}{\sigma_j^f + \sum_{(k,l) \neq (i,j)} h_{kj}^f p_{kl}^f(\boldsymbol{h})}. \tag{7}$$

With, e.g., a capacity-achieving channel code, it follows that $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) = \log[1 + \gamma_{ij}^f(\boldsymbol{h})]$. Another example entails the use of a finite number of adaptive modulation and coding (AMC) modes. In this case, $C(\cdot)$ is a staircase function defined by the rate of the AMC modes considered.

In any event, the ergodic capacity c_{ij} of the wireless link $T_i \to T_j$ is obtained after averaging over all possible channel realizations to obtain

$$c_{ij} := \mathbb{E}_{\mathbf{H}}\left[\sum_{f \in \mathcal{F}} c_{ij}^f(\boldsymbol{h})\right] = \mathbb{E}_{\mathbf{H}}\left[\sum_{f \in \mathcal{F}} C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))\right]. \tag{8}$$

The average power and link capacity expressions in (4) and (8) along with the flow and rate constraints in (1) and (2) describe information flow in a generic wireless network. They can be used to define the wireless network optimization problem described in the next section.

III. Optimal Wireless Networking

Average power p_i and link capacities c_{ij} depend on the chosen power profiles $p_{ij}^f(\boldsymbol{h})$ as per (4) and (8). Average link rates r_{ij}^k are then constrained by (2) and end-to-end flow rates a_i^k by (1). Problem variables $p_{ij}^f(\boldsymbol{h})$, c_{ij}, p_i, r_{ij}^k, and a_i^k that satisfy these equations can be supported by the network. As network designers, we want to select out of these set of feasible variables those that are optimal in some sense. To this end, consider concave $U_i^k(a_i^k)$ and convex $V_i(p_i)$ utility functions, respectively, representing the reward of rate a_i^k and the cost of power p_i. Though not required, it is intuitively reasonable to consider $U_i^k(a_i^k)$ and $V_i(p_i)$ as increasing functions of their arguments. Based on these utility functions, the optimal networking problem is defined as [cf. (1), (2), (4), and (8)]

$$\mathsf{P} = \max \sum\nolimits_{i,k} U_i^k\left(a_i^k\right) - \sum_i V_i(p_i) \tag{9}$$

$$c_{ij} \leq \mathbb{E}_{\mathbf{H}}\left[\sum_{f \in \mathcal{F}} C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))\right] \tag{10}$$

$$p_i \geq \mathbb{E}_{\mathbf{H}}\left[\sum_{j \in \mathcal{N}(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\boldsymbol{h})\right] \tag{11}$$

$$a_i^k \leq \sum_{j \in n(i)} \left(r_{ij}^k - r_{ji}^k\right), \sum_k r_{ij}^k \leq c_{ij} \tag{12}$$

where the constraints (1), (4) and (8) have been relaxed from equalities to inequalities—something that can be done without loss of optimality. All problem variables are nonnegative, but this is left implicit in (9). Also left implicit are power constraints $p_i \leq p_{\max i}$, arrival rate requirements $a_{\min i}^k \leq a_i^k \leq a_{\max i}^k$ as well as upper bound constraints $c_{ij} \leq c_{\max}$ and $r_{ij}^k \leq r_{\max}$ on link capacities and link flow rates. For reference, let $\boldsymbol{X}$ denote the set of primal variables c_{ij}, p_i, r_{ij}^k and a_i^k for all possible

subindices, i.e., all i and all $j \in n(i)$ for c_{ij}, all i for p_i and so on. The aforementioned implicit constraints specify a set of feasible variables

$$\boldsymbol{X} \in \mathcal{B} := \{\boldsymbol{X} : 0 \leq p_i \leq p_{\max i}, a^k_{\min i} \leq a^k_i \leq a^k_{\max i}, \\ 0 \leq c_{ij} \leq c_{\max}, 0 \leq r^k_{ij} \leq r_{\max}\}. \quad (13)$$

Likewise, define the vector-valued power function $\boldsymbol{p}(\boldsymbol{h})$ with entries $p^f_{ij}(\boldsymbol{h})$. Possible instantaneous powers $\boldsymbol{p}(\boldsymbol{h})$ used when the fading realization is $\boldsymbol{h}$ are constrained to the set

$$\boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h}) \subseteq \left\{\boldsymbol{p}(\boldsymbol{h}) : 0 \leq p^f_{ij}(\boldsymbol{h}) \leq p_{\max}, \right\}. \quad (14)$$

Powers $p^f_{ij}(\boldsymbol{h})$ are positive and constrained by a maximum instantaneous power $p_{\max}$ but the set $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$ might impose further restrictions. The set $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$ is not required to be convex or even connected so that it can account for practical situations in which allowable transmit powers are a discrete set, e.g., $\boldsymbol{\mathcal{P}}(\boldsymbol{h}) = \{\boldsymbol{p}(\boldsymbol{h}) : p^f_{ij}(\boldsymbol{h}) \in \{0, p_{\max}\}\}$. Further note that it is not the power allocation function $\boldsymbol{p}(\boldsymbol{h})$ that is restricted to the set $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$. The relation $\boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ denotes that individual values $\boldsymbol{p}(\boldsymbol{h})$ of the power allocation function for channel coefficients $\boldsymbol{h}$ are restricted to the set $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$.

The constraints in (13) and (14) are henceforth referred to as box constraints, recall though, that $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$ is not necessarily a box but a subset of one. They will be kept implicit for the most part but when required by clarity they will be stated explicitly.

Because the function $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$ is not necessarily concave with respect to $\boldsymbol{p}^f(\boldsymbol{h})$, and the sets $\mathcal{P}(\boldsymbol{h})$ need not be convex, problem (9) is a difficult optimization problem. In fact, if channels are deterministic (i.e., there is only one possible channel realization $\boldsymbol{h}$), it has been proved that problem (9) is NP-hard [30]. This difficulty notwithstanding, there are fundamental principles of wireless networking problems to be derived from properties of (9). Revealing these principles can be facilitated by looking at the Lagrange dual problem.

A. Lagrange Dual Problem

To define the dual problem, associate Lagrange multipliers λ_{ij} with the capacity constraints in (10), μ_i with the power in (11), and ν^k_i and ξ_{ij} with the flow and rate constraints in (12). For notational brevity, call $\boldsymbol{\Lambda}$ the set of all dual variables, i.e., $\lambda_{ij}, \mu_i, \nu^k_i, \xi_{ij}$, and write the Lagrangian as

$$\begin{aligned}\boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}] &= \sum_{i,k} U^k_i\left(a^k_i\right) - \sum_i V_i(p_i) \\ &+ \sum_{i,j} \lambda_{ij}\left[\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))\right] - c_{ij}\right] \\ &+ \sum_i \mu_i\left[p_i - \mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} p^f_{ij}(\boldsymbol{h})\right]\right] \\ &+ \sum_{i,k} \nu^k_i\left[\sum_{j\in n(i)}\left(r^k_{ij} - r^k_{ji}\right) - a^k_i\right] \\ &+ \sum_{i,j} \xi_{ij}\left[c_{ij} - \sum_k r^k_{ij}\right]. \end{aligned} \quad (15)$$

The dual function is obtained by maximizing the Lagrangian over the primal variables satisfying the box constraints in (13); i.e.

$$g[\boldsymbol{\Lambda}] = \max_{(\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}))\in\mathcal{B}} \boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}]. \quad (16)$$

And finally the dual problem is defined as

$$\mathsf{D} = \min_{\boldsymbol{\Lambda}\geq\mathbf{0}} g[\boldsymbol{\Lambda}]. \quad (17)$$

Since (9) is nonconvex the duality gap is, in principle, nonzero which implies that $\mathsf{D} \geq \mathsf{P}$. Solving (17) is thus a relaxation in the sense that it yields an upper bound D on the maximum achievable utility P. Consequently, the usefulness of (17) depends on the proximity of D to P. This is explored in the next section.

IV. Optimality of Dual Relaxation

The challenges in solving (9) are now clear. For deterministic channels, the problem is known to be NP-hard. The Lagrange dual problem in (17) is certainly useful in establishing upper bounds on the achievable utility, but may or may not be close to the actual utility P. One expects that introducing fading will complicate matters further. Remarkably, it will turn out that in the presence of fading the duality gap vanishes, i.e., $\mathsf{P} = \mathsf{D}$. We state this result in the following theorem.

Theorem 1: Let P denote the optimum value of the primal problem (9) and D that of its dual in (17). If the channel cumulative distribution function (cdf) is continuous, then

$$\mathsf{P} = \mathsf{D}. \quad (18)$$

To appreciate this result recall that the link capacity function $C(\cdot)$ is not necessarily concave in Theorem 1; hence, the optimization problem is generally nonconvex. The duality gap, however, is null. Continuity of the channel cdf ensures that no channel realization has strictly positive probability. This is satisfied by practical fading channel models including those adhering to Rayleigh, Rice, and Nakagami distributions.

Before proceeding to the proof of Theorem 1, let us recall Lyapunov's convexity theorem and introduce pertinent definitions starting with the concept of nonatomic measure [29].

Definition 1 (Nonatomic Measure): Let w be a measure defined on the Borel field $\mathbb{B}$ of subsets of a space $\boldsymbol{\mathcal{X}}$. Measure w is nonatomic if for any measurable set $E_0 \in \mathbb{B}$ with $w(E_0) > 0$, there exist a subset E of E_0; i.e., $E \subset E_0$, such that $w(E_0) > w(E) > 0$.

Familiar measures are probability related, e.g., the probability of a set for a given channel distribution. To build intuition on the notion of nonatomic measure consider a random variable X taking values in $[0, 1]$ and $[2, 3]$. The probability of landing in each of these intervals is $1/2$ and X is uniformly distributed inside each of them; see Fig. 2. The space $\boldsymbol{\mathcal{X}}$ is the real line, and the Borel field $\mathbb{B}$ comprises all subsets of real numbers. For every subset $E \in \mathbb{B}$ define the measure of E as twice the integral of x, weighted by the probability distribution of X on the set E, i.e.

$$w_X(E) := 2\int_E x dX. \quad (19)$$

Note that, except for the factor 2, the value of $w_X(E)$ represents the contribution of the set E to the expected value of X and that

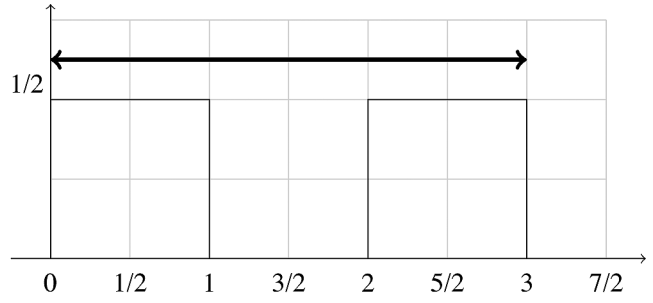

Fig. 2. Consider a random variable X with probability distribution as shown in the figure. An example of non-atomic measure [cf. Definition 1] is twice the integral of x, weighted by the probability distribution of X on the set E, i.e., $w_X(E) := 2\int_E x dX$. Lyapunov's convexity theorem [cf. Theorem 2] states that the range of all possible integrals $\boldsymbol{\mathcal{W}} := \{w_X(E) : \mathrm{E} \in \boldsymbol{\mathcal{B}}\}$ is convex. It is not difficult to see that in this case the range of w_X is the (convex) interval $[0, 3]$.

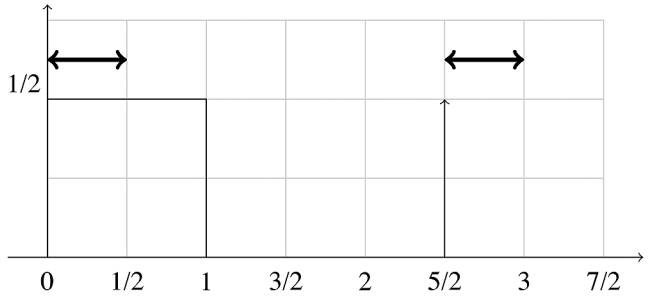

Fig. 3. As in Fig. 2 define the measure $w_Y(E) := 2\int_E y dY$. Because $Y = 1$ has strictly positive probability the measure $w_Y(E)$ is atomic. Theorem 2 does not apply in this case. In fact it can be easily seen that the range of $w_Y(E)$ is the (nonconvex) union of the intervals $[0, 1/2]$ and $[5/2, 3]$.

when E is the whole space $\boldsymbol{\mathcal{X}}$, it holds $w_X(\boldsymbol{\mathcal{X}}) = 2\mathrm{E}_{\boldsymbol{X}}(x)$. According to Definition 1, $w_X(E)$ is a non-atomic measure of elements of $\mathbb{B}$. Indeed, every subset E_0 with $w_X(E_0) > 0$ includes at least an interval (a, b). The measure of the set $E := E_0 - ((a+b)/2, b)$ formed by removing the upper half of (a, b) from E_0 is $w_X(E) = w_X(E_0) - (b-a)/2$. The measure of E satisfies $w_X(E) > 0$ as required for $w_X(E)$ to be nonatomic.

To contrast this with an example of an atomic measure consider a random variable Y landing equiprobably in $[0,1]$ or $5/2$; see Fig. 3. In this case, the measure $w_Y(E) := 2\int_E y dY$ is atomic because the set $E_0 = \{5/2\}$ has positive measure $w_Y(E) = 1$. The only set $E \subset E_0$ is the empty set whose measure is null.

The difference between the distributions of X and Y is that Y contains a point of strictly positive probability, i.e., an atom. This implies presence of delta functions in the probability density function of Y. Or, in a rather cleaner statement the cdf of X is continuous whereas the cdf of Y is not.

Lyapunov's convexity theorem introduced next refers to the range of values taken by (vector) nonatomic measures.

Theorem 2 (Lyapunov's Convexity Theorem [29]): Let $w_1, \ldots, w_n$ be nonatomic measures on the Borel field $\mathbb{B}$ of subsets of a space $\mathcal{X}$. With ^{T} denoting trasposition, consider the vector measure $\mathbf{w}(E) := [w_1(E), \ldots, w_n(E)]^T$. The range $\boldsymbol{\mathcal{W}} := \{\mathbf{w}(E) : E \in \mathbb{B}\}$ of the vector measure is convex; i.e., if $\mathbf{w}(E_1) = \mathbf{w}_1$ and $\mathbf{w}(E_2) = \mathbf{w}_2$, then for any $\alpha \in [0,1]$ there exists $E_0 \in \mathbb{B}$ such that $\mathbf{w}(E_0) = \alpha\mathbf{w}_1 + (1-\alpha)\mathbf{w}_2$.

Returning to the probability measures defined in terms of the probability distributions of the random variables X and Y, Theorem 2 asserts that the range of $w_X(E)$, i.e., the set of all possible values taken by w_X is convex. In fact, it is not difficult to verify that the range of w_X is the convex interval $[0, 3]$ as shown in Fig. 2. Theorem 2 does not claim anything about w_Y. In this case, it is easy to see that the range of w_Y is the (nonconvex) union of the intervals $[0, 1/2]$ and $[5/2, 3]$; see Fig. 3.

Having introduced Theorem 2, we move on to the proof of Theorem 1.

Proof: (of Theorem 1): To establish zero duality gap we will consider a perturbed version of (9) obtained by perturbing the constraints used to define the Lagrangian in (15). The perturbation function $\mathsf{P}(\boldsymbol{\Delta})$ assigns to each (perturbation) parameter set $\boldsymbol{\Delta} := (\Delta c_{ij}, \Delta p_i, \Delta a_i^k, \Delta r_{ij})$ the solution of the (perturbed) optimization problem

$$\mathsf{P}(\boldsymbol{\Delta}) = \max \sum_{i,k} U_i^k\left(a_i^k\right) - \sum_i V_i(p_i) \tag{20}$$

$$c_{ij} - \mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))\right] \le \Delta c_{ij} \tag{21}$$

$$\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} p_{ij}^f(\boldsymbol{h})\right] - p_i \le \Delta p_i \tag{22}$$

$$a_i^k - \sum_{j\in n(i)}\left(r_{ij}^k - r_{ji}^k\right) \le \Delta a_i^k \tag{23}$$

$$\sum_k r_{ij}^k - c_{ij} \le \Delta r_{ij} \tag{24}$$

where the box constraints (13) and (14) are implicit as in (9). The perturbed problem (20) can be interpreted as a modified version of (9), where we allow the constraints to be violated by $\boldsymbol{\Delta}$ amounts. To prove that the duality gap is zero, it suffices to show that $\mathsf{P}(\boldsymbol{\Delta})$ is a concave function of $\boldsymbol{\Delta}$; see, e.g., [31].

Let $\underline{\boldsymbol{\Delta}} := (\Delta\underline{c}_{ij}, \Delta\underline{p}_i, \Delta\underline{a}_i^k, \Delta\underline{r}_{ij})$ and $\bar{\boldsymbol{\Delta}} := (\Delta\bar{c}_{ij}, \Delta\bar{p}_i, \Delta\bar{a}_i^k, \Delta\bar{r}_{ij})$ denote arbitrary sets of perturbations with respective optimal values $\underline{\mathsf{P}} := \mathsf{P}(\underline{\boldsymbol{\Delta}})$ and $\bar{\mathsf{P}} := \mathsf{P}(\bar{\boldsymbol{\Delta}})$. Further, let $\underline{\boldsymbol{X}}$ denote (likewise, $\underline{p}_{ij}^f(\boldsymbol{h})$, $\underline{c}_{ij}$, $\underline{p}_i$, $\underline{r}_{ij}^k$, $\underline{a}_i^k$) the variables achieving the optimum under perturbation $\underline{\boldsymbol{\Delta}}$; and $\bar{\boldsymbol{X}}$ (likewise, $\bar{p}_{ij}^f(\boldsymbol{h})$, $\bar{c}_{ij}$, $\bar{p}_i$, $\bar{r}_{ij}^k$, $\bar{a}_i^k$) the ones achieving optimality under $\bar{\boldsymbol{\Delta}}$. For arbitrary $\alpha \in [0,1]$, we are interested in the solution of (20) under perturbation $\underline{\bar{\boldsymbol{\Delta}}} := \alpha\underline{\boldsymbol{\Delta}} + (1-\alpha)\bar{\boldsymbol{\Delta}}$. In particular, to show that $\mathsf{P}(\underline{\bar{\boldsymbol{\Delta}}})$ is concave we need to establish

$$\begin{aligned}\mathsf{P}(\underline{\bar{\boldsymbol{\Delta}}}) &= \mathsf{P}[\alpha\underline{\boldsymbol{\Delta}} + (1-\alpha)\bar{\boldsymbol{\Delta}}] \\ &\ge \alpha\mathsf{P}(\underline{\boldsymbol{\Delta}}) + (1-\alpha)\mathsf{P}(\bar{\boldsymbol{\Delta}}).\end{aligned} \tag{25}$$

Key in establishing (25) is the following lemma.

Lemma 1: Consider perturbations $\underline{\boldsymbol{\Delta}}$ and $\bar{\boldsymbol{\Delta}}$ and let $\underline{\boldsymbol{X}}$ and $\bar{\boldsymbol{X}}$ be the corresponding optimal arguments that solve (20). Define the perturbation $\underline{\bar{\boldsymbol{\Delta}}} := \alpha\underline{\boldsymbol{\Delta}} + (1-\alpha)\bar{\boldsymbol{\Delta}}$ with $\alpha \in [0,1]$. The average link capacity and power variables

$$\begin{aligned}\underline{\bar{c}}_{ij} &:= \alpha\underline{c}_{ij} + (1-\alpha)\bar{c}_{ij} \\ \underline{\bar{p}}_i &:= \alpha\underline{p}_i + (1-\alpha)\bar{p}_i\end{aligned} \tag{26}$$

are feasible for (20)–(24) with perturbation $\underline{\bar{\mathbf{\Delta}}}$ in the sense that they satisfy the box constraints in (13) and the constraints (21) and (22) for some power allocation function $\underline{\bar{\boldsymbol{p}}}(\boldsymbol{h})$.

Before proceeding to the proof of Lemma 1, let us see first how it can be used to complete the proof of Theorem 1.

We will first show that $\underline{\bar{c}}_{ij}$ and $\underline{\bar{r}}^k_{ij} := \alpha \underline{r}^k_{ij} + (1-\alpha)\underline{r}^k_{ij}$ satisfy the constraint (24) for perturbation $\underline{\bar{\mathbf{\Delta}}}$. Indeed, since $(\underline{r}^k_{ij}, \underline{c}_{ij})$ and $(\bar{r}^k_{ij}, \bar{c}_{ij})$ are feasible for the problem (20)–(24) under respective perturbations $\underline{\mathbf{\Delta}}$ and $\bar{\mathbf{\Delta}}$, we have [cf. (23)]

$$\sum_k \underline{r}^k_{ij} - \underline{c}_{ij} \leq \Delta \underline{r}_{ij} \tag{27}$$

$$\sum_k \bar{r}^k_{ij} - \bar{c}_{ij} \leq \Delta \bar{r}_{ij}. \tag{28}$$

Multiplying (27) by α, (28) by $(1-\alpha)$, summing up and rearranging terms, yields

$$\sum_k \left[\alpha \underline{r}^k_{ij} + (1-\alpha)\bar{r}^k_{ij}\right] - \left[\alpha \underline{c}_{ij} + (1-\alpha)\bar{c}_{ij}\right] \leq \left[\Delta \underline{r}_{ij} + (1-\alpha)\Delta \bar{r}_{ij}\right]. \tag{29}$$

Using the definitions $\underline{\bar{r}}^k_{ij} := \alpha \underline{r}^k_{ij} + (1-\alpha)\underline{r}^k_{ij}$, $\underline{\bar{c}}_{ij} := \alpha \underline{c}_{ij} + (1-\alpha)\bar{c}_{ij}$ and $\Delta \underline{\bar{r}}_{ij} := \alpha \Delta \underline{r}_{ij} + (1-\alpha)\Delta \bar{r}_{ij}$, the inequality in (29) is equivalent to

$$\sum_k \underline{\bar{r}}^k_{ij} - \underline{\bar{c}}_{ij} \leq \Delta \underline{\bar{r}}_{ij}. \tag{30}$$

Likewise, we can prove that $\underline{\bar{r}}^k_{ij}$ and $\underline{\bar{a}}^k_i := \alpha \underline{a}^k_i + (1-\alpha)\underline{a}^k_i$ satisfy (23) for perturbation $\underline{\bar{\mathbf{\Delta}}}$ [simply repeat steps (27)–(30) for the flow conservation constraint in (23)].

In summary, we have that $\underline{\bar{a}}^k_i$, $\underline{\bar{r}}^k_{ij}$, $\underline{\bar{c}}_{ij}$, and $\underline{\bar{p}}_i$ are feasible points of (20)–(24) under perturbation $\underline{\bar{\mathbf{\Delta}}}$ because they satisfy all the constraints in (21)–(24) and the box constraints in (13). Since $\mathsf{P}(\underline{\bar{\mathbf{\Delta}}})$ is the maximum of (20) among all feasible (a^k_i, p_i), it must satisfy

$$\begin{aligned}\mathsf{P}(\underline{\bar{\mathbf{\Delta}}}) &\geq \sum_{i,k} U^k_i\left(\underline{\bar{a}}^k_i\right) - \sum_i V_i(\underline{\bar{p}}_i) \qquad (31)\\ &= \sum_{i,k} U^k_i\left[\alpha \underline{a}^k_i + (1-\alpha)\bar{a}^k_i\right] - \sum_i V_i[\alpha \underline{p}_i + (1-\alpha)\bar{p}_i]\end{aligned}$$

where in the second equality we used the definitions of $\underline{\bar{a}}^k_i := \alpha \underline{a}^k_i + (1-\alpha)\bar{a}^k_i$ and $\underline{\bar{p}}_i := \alpha \underline{p}_i + (1-\alpha)\bar{p}_i$. Finally, since by assumption the $U^k_i(\cdot)$ functions are concave and the $V_i(\cdot)$ functions are convex, it follows that

$$\begin{aligned}\mathsf{P}(\underline{\bar{\mathbf{\Delta}}}) &\geq \sum_{i,k}\left[\alpha U^k_i\left(\underline{a}^k_i\right) + (1-\alpha)U^k_i\left(\bar{a}^k_i\right)\right]\\ &\quad - \sum_i [\alpha V_i(\underline{p}_i) + (1-\alpha)V_i(\bar{p}_i)]\\ &= \alpha\left[\sum_{i,k} U^k_i\left(\underline{a}^k_i\right) - \sum_i V_i(\underline{p}_i)\right]\\ &\quad + (1-\alpha)\left[\sum_{i,k} U^k_i\left(\bar{a}^k_i\right) - \sum_i V_i(\bar{p}_i)\right]\\ &= \alpha \mathsf{P}(\underline{\mathbf{\Delta}}) + (1-\alpha)\mathsf{P}(\bar{\mathbf{\Delta}}) \qquad (32)\end{aligned}$$

where the second equality holds true because $(\underline{a}^k_i, \underline{p}_i)$ and $(\bar{a}^k_i, \bar{p}_i)$ are optimal arguments of (20) under perturbations $\underline{\mathbf{\Delta}}$ and $\bar{\mathbf{\Delta}}$, respectively. Comparing (25) to (32), we deduce that the perturbation function $\mathsf{P}(\mathbf{\Delta})$ is concave. Therefore, the duality gap is null, i.e., $\mathsf{D} = \mathsf{P}$. ■

We proceed now to the proof of Lemma 1.

Proof: (of Lemma 1): To prove that $\underline{\bar{c}}_{ij}$ and $\underline{\bar{p}}_i$ as defined in (26) satisfy (21) and (22) for perturbation $\underline{\bar{\mathbf{\Delta}}}$, we have to show that there exist feasible power distributions $\underline{\bar{\boldsymbol{p}}}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ with elements $\underline{\bar{p}}^f_{ij}(\boldsymbol{h})$ for which the following holds true:

$$\underline{\bar{c}}_{ij} - \mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\bar{\boldsymbol{p}}}^f(\boldsymbol{h}))\right] \leq \Delta \underline{\bar{c}}_{ij} \tag{33}$$

$$\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{\bar{p}}^f_{ij}(\boldsymbol{h})\right] - \underline{\bar{p}}_i \leq \Delta \underline{\bar{p}}_i. \tag{34}$$

For this we will use Theorem 2 (Lyapunov's convexity theorem). Consider the space of all possible channel realizations $\boldsymbol{\mathcal{H}}$, and the Borel field $\mathbb{B}$ of all possible subsets of $\boldsymbol{\mathcal{H}}$. For every set $E \in \mathbb{B}$ define the measures

$$w_{ij}(E) := \left[\int_E \left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right] d\mathbf{H}, \int_E \left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right] d\mathbf{H}\right]^T \tag{35}$$

$$v_i(E) := \left[\int_E \left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}^f_{ij}(\boldsymbol{h})\right] d\mathbf{H}, \int_E \left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \bar{p}^f_{ij}(\boldsymbol{h})\right] d\mathbf{H}\right]^T \tag{36}$$

where the integrals are over the set E with respect to the distribution of the channels' random variable $\mathbf{H}$. A vector of channel realizations $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}$ is a point in the space $\boldsymbol{\mathcal{H}}$. The set E is a collection of vectors $\boldsymbol{h}$. Each of these sets is assigned vector measures $w_{ij}(E)$ and $v_i(E)$ defined in terms of the power distributions $\underline{p}^f_{ij}(\boldsymbol{h})$ and $\bar{p}^f_{ij}(\boldsymbol{h})$. The entries of $w_{ij}(E)$ represent the contribution of realizations $\boldsymbol{h} \in E$ to the capacity of the link $T_i \to T_j$. The first entry measures such a contribution when the distribution is $\underline{p}^f_{ij}(\boldsymbol{h})$, i.e., the optimal power distribution under perturbation $\underline{\mathbf{\Delta}}$. Likewise, the second entry of $w_{ij}(E)$ denotes the contribution to the $T_i \to T_j$ link capacity of the optimal power distribution $\bar{p}^f_{ij}(\boldsymbol{h})$ associated with perturbation $\bar{\mathbf{\Delta}}$. Correspondingly, the entries of $v_i(E)$ denote the power consumed by T_i for channel realization $\boldsymbol{h} \in E$ expressed in terms of $\underline{p}^f_{ij}(\boldsymbol{h})$ and $\bar{p}^f_{ij}(\boldsymbol{h})$.

The measures $w_{ij}(E)$ and $v_i(E)$ are nonatomic. This follows from the fact that the channel cdf is continuous and the power distributions $p^f_{ij}(\boldsymbol{h})$ are upper bounded by $p_{\max}$, implying that

there are no channel realizations with positive measure, i.e., $w_{ij}(\boldsymbol{h}) = 0$ and $v_i(\boldsymbol{h}) = 0$ for all $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}$.

Two particular sets that are important for this proof are the empty set $E = \emptyset$ and the entire space $E = \boldsymbol{\mathcal{H}}$. For $E = \boldsymbol{\mathcal{H}}$, the integrals $\int_E(\cdot)d\mathbf{H} = \int_{\boldsymbol{\mathcal{H}}}(\cdot)d\mathbf{H}$ in (35) and (36) coincide with the expected value operators $\mathbb{E}_{\mathbf{H}}(\cdot)$ in (21) and (22). We write this explicitly as

$$w_{ij}(\boldsymbol{\mathcal{H}}) := \left[\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right], \mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right]\right]^T \tag{37}$$

$$v_i(\boldsymbol{\mathcal{H}}) := \left[\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\boldsymbol{h})\right], \mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \bar{p}_{ij}^f(\boldsymbol{h})\right]\right]^T. \tag{38}$$

For $E = \emptyset$, or any other zero-measure set for that matter, we have $w_{ij}(\emptyset) = 0$ and $v_i(\emptyset) = 0$.

Collect now the scalar measures $w_{ij}(E)$ of all network links and $v_i(E)$ of all terminals into the vector measure $\mathbf{w}(E) := [\{w_{ij}(E)\}_{i\in[1,J],j\in n(i)}, \{v_i(\emptyset)\}_{i\in[1,J]}]$. Since the scalar measures are nonatomic, $\mathbf{w}(E)$ is also nonatomic. Theorem 2 then implies that the range of $\mathbf{w}$ is convex; hence

$$\mathbf{w}_0 := \alpha\mathbf{w}(\boldsymbol{\mathcal{H}}) + (1-\alpha)\mathbf{w}(\emptyset) = \alpha\mathbf{w}(\boldsymbol{\mathcal{H}}) \tag{39}$$

belongs to the range of possible measures. Therefore, there must exist a set E_0 such that $\mathbf{w}(E_0) = \mathbf{w}_0 = \alpha\mathbf{w}(\boldsymbol{\mathcal{H}})$. Equivalently, this implies that $w_{ij}(E_0) = \alpha w_{ij}(\boldsymbol{\mathcal{H}})$ and $v_i(E_0) = \alpha w_{ij}(\boldsymbol{\mathcal{H}})$ $\forall(i,j)$. Focusing on the first entries of w_{ij} and v_i, it follows that

$$\int_{E_0}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right]d\mathbf{H} = \alpha\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right] \tag{40}$$

$$\int_{E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H} = \alpha\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\boldsymbol{h})\right] \tag{41}$$

and this holds true for all (i,j). The same relation holds for the second entries of w_{ij} and v_i, i.e., (40) and (41) are valid if $\underline{a}_{ij}^f(\boldsymbol{h})$ and $\underline{p}_{ij}^f(\boldsymbol{h})$ are replaced by $\bar{a}_{ij}^f(\boldsymbol{h})$ and $\bar{p}_{ij}^f(\boldsymbol{h})$, but this is not important for the proof.

Consider now the complement set E_0^c defined as the set for which $E_0 \cup E_0^c = \boldsymbol{\mathcal{H}}$ and $E_0 \cap E_0^c = \emptyset$. Given this definition and the additivity property of measures, we arrive at $\mathbf{w}(E_0) + \mathbf{w}(E_0^c) = \mathbf{w}(\boldsymbol{\mathcal{H}})$. Combining the latter with (39), yields

$$\mathbf{w}(E_0^c) = \mathbf{w}(\boldsymbol{\mathcal{H}}) - \mathbf{w}(E_0) = (1-\alpha)\mathbf{w}(\boldsymbol{\mathcal{H}}). \tag{42}$$

Mimicking the reasoning leading from (39) to (40)–(41), we further deduce that (42) implies that for all (i,j), $w_{ij}(E_0^c) = (1-\alpha)w_{ij}(\boldsymbol{\mathcal{H}})$ and $v_i(E_0^c) = (1-\alpha)w_{ij}(\boldsymbol{\mathcal{H}})$. Restricting focus to the *second* entries of w_{ij} and v_i, it follows that

$$\int_{E_0^c}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right]d\mathbf{H} = (1-\alpha)\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right] \tag{43}$$

$$\int_{E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \bar{p}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H} = (1-\alpha)\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \bar{p}_{ij}^f(\boldsymbol{h})\right] \tag{44}$$

and this, again, holds true for all (i,j).

Define now power distributions $\bar{\underline{p}}_{ij}^f(\boldsymbol{h})$ coinciding with $\underline{p}_{ij}^f(\boldsymbol{h})$ for channel realization $\boldsymbol{h} \in E_0$ and with $\bar{p}_{ij}^f(\boldsymbol{h})$ when $\boldsymbol{h} \in E_0^c$, i.e.

$$\bar{\underline{p}}_{ij}^f(\boldsymbol{h}) = \begin{cases} \underline{p}_{ij}^f(\boldsymbol{h}) & \boldsymbol{h} \in E_0 \\ \bar{p}_{ij}^f(\boldsymbol{h}) & \boldsymbol{h} \in E_0^c \end{cases}. \tag{45}$$

The distribution $\bar{\underline{p}}_{ij}^f(\boldsymbol{h})$ satisfies the power box constraint in (14). Indeed, to see that $\bar{\underline{\boldsymbol{p}}}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ for all $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}$ note that $\underline{\boldsymbol{p}}(\boldsymbol{h})$ and $\bar{\boldsymbol{p}}(\boldsymbol{h})$ are feasible in their respective problems and as such $\underline{\boldsymbol{p}}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ and $\bar{\boldsymbol{p}}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ for all channels $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}$. Because for given channel realization $\boldsymbol{h}$ it holds that either $\bar{\underline{\boldsymbol{p}}}(\boldsymbol{h}) = \underline{\boldsymbol{p}}(\boldsymbol{h})$ when $\boldsymbol{h} \in E_0$ or $\bar{\underline{\boldsymbol{p}}}(\boldsymbol{h}) = \bar{\boldsymbol{p}}(\boldsymbol{h})$ when $\boldsymbol{h} \in E_0^c$ it follows that $\bar{\underline{\boldsymbol{p}}}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})$ for all channel realizations $\boldsymbol{h} \in E_0 \cup E_0^c = \boldsymbol{\mathcal{H}}$.

Using (40) and (43), the average link capacities for power allocation $\bar{\underline{\boldsymbol{p}}}(\boldsymbol{h})$can be expressed in terms of $\underline{\boldsymbol{p}}(\boldsymbol{h})$, $\bar{\boldsymbol{p}}(\boldsymbol{h})$ as

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\underline{\boldsymbol{p}}}^f(\boldsymbol{h}))\right] \\
&= \int_{E_0}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\underline{\boldsymbol{p}}}^f(\boldsymbol{h}))\right]d\mathbf{H} + \int_{E_0^c}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\underline{\boldsymbol{p}}}^f(\boldsymbol{h}))\right]d\mathbf{H} \\
&= \int_{E_0}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right]d\mathbf{H} + \int_{E_0^c}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right]d\mathbf{H} \\
&= \alpha\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right] + (1-\alpha)\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}} C_{ij}(\boldsymbol{h}^f, \bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right].
\end{aligned} \tag{46}$$

The first equality in (46) holds because the space $\boldsymbol{\mathcal{H}}$ is divided into E_0 and its complement E_0^c. The second equality is true because when restricted to E_0, $\underline{\bar{p}}_{ij}^f(\boldsymbol{h}) = \underline{p}_{ij}^f(\boldsymbol{h})$; and when restricted to E_0^c, $\underline{\bar{p}}_{ij}^f(\boldsymbol{h}) = \bar{p}_{ij}^f(\boldsymbol{h})$. The third equality follows from (40) and (43).

An analogous relation also holds for the average power consumptions. Indeed, separating $\boldsymbol{\mathcal{H}}$ into E_0 and E_0^c and using the definition of $\underline{\bar{p}}_{ij}^f(\boldsymbol{h})$ in (45) along with (41) and (44), yields

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{\bar{p}}_{ij}^f(\boldsymbol{h})\right] \qquad (47)\\
&=\int_{E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{\bar{p}}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H}+\int_{E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{\bar{p}}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H}\\
&=\int_{E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H}+\int_{E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\bar{p}_{ij}^f(\boldsymbol{h})\right]d\mathbf{H}\\
&=\alpha\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\boldsymbol{h})\right]+(1-\alpha)\mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\bar{p}_{ij}^f(\boldsymbol{h})\right].
\end{aligned}$$

Based on (46) and (47) we are ready to complete the proof. Considering the link capacity constraints (21) under perturbations $\underline{\boldsymbol{\Delta}}$ and $\bar{\boldsymbol{\Delta}}$, we have that

$$\underline{c}_{ij}-\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f,\underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right]\leq\Delta\underline{c}_{ij} \qquad (48)$$

$$\bar{c}_{ij}-\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f,\bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right]\leq\Delta\bar{c}_{ij}. \qquad (49)$$

As with the link transmission rate constraints in (27) and (28), multiply (48) by α and (49) by $(1-\alpha)$. Summing up, rearranging terms and using the definitions $\underline{\bar{c}}_{ij}:=\alpha\underline{c}_{ij}+(1-\alpha)\bar{c}_{ij}$ and $\Delta\underline{\bar{c}}_{ij}:=\alpha\Delta\underline{c}_{ij}+(1-\alpha)\Delta\bar{c}_{ij}$, we obtain

$$\begin{aligned}
&\underline{\bar{c}}_{ij}-\alpha\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f,\underline{\boldsymbol{p}}^f(\boldsymbol{h}))\right]\\
&\qquad-(1-\alpha)\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f,\bar{\boldsymbol{p}}^f(\boldsymbol{h}))\right]\leq\Delta\underline{\bar{c}}_{ij}. \qquad (50)
\end{aligned}$$

But the terms in (50) involving expectations are precisely the terms in the right-hand side (RHS) of (46). They can be therefore replaced by the expected link capacity for power distribution $\underline{\bar{\boldsymbol{p}}}(\boldsymbol{h})$, to obtain

$$\underline{\bar{c}}_{ij}-\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f,\underline{\bar{\boldsymbol{p}}}^f(\boldsymbol{h}))\right]\leq\Delta\underline{\bar{c}}_{ij} \qquad (51)$$

which coincides with (33).

Likewise, considering the power constraints in (22) under $\underline{\boldsymbol{\Delta}}$ and $\bar{\boldsymbol{\Delta}}$ and mimicking the steps in (48)–(50) yields

$$\begin{aligned}
&\alpha\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\boldsymbol{h})\right]\\
&\qquad+(1-\alpha)\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\bar{p}_{ij}^f(\boldsymbol{h})\right]-\underline{\bar{p}}_i\leq\Delta\underline{\bar{p}}_{ij}. \qquad (52)
\end{aligned}$$

Since the summation of expectations in (52) coincides with the RHS of (47), it can be replaced by the left-hand side (LHS) of (47) to obtain (34). From (51)—equivalently (33)—and (34) it follows that $\underline{\bar{c}}_{ij}:=\alpha\underline{c}_{ij}+(1-\alpha)\bar{c}_{ij}$ and $\underline{\bar{p}}_i:=\alpha\underline{p}_i+(1-\alpha)\bar{p}_i$ and power allocation $\underline{\bar{\boldsymbol{p}}}(\boldsymbol{h})$ as in (45) satisfy the constraints (21) and (22) for perturbation $\underline{\bar{\boldsymbol{\Delta}}}$. Since $\underline{\bar{c}}_{ij}$ and $\underline{\bar{p}}_i$ also satisfy the box constraints they are feasible for (20)-(24) thus establishing Lemma 1. ■

Given that the duality gap of the generic optimal wireless networking problem (9) is zero, the dual problem (17) can be solved instead. Because the dual function $g(\boldsymbol{\Lambda})$ in (16) is convex, descent methods can, in principle, be used to find the optimal multipliers $\boldsymbol{\Lambda}^*$. It is also worth remarking that the primal problem (9) is a variational problem requiring determination of the *function* $\boldsymbol{p}(\boldsymbol{h})$ that maps channel realizations to transmission powers. In that sense, it is an infinite-dimensional optimization problem. The dual problem however, involves a finite number of variables.

An important caveat is that zero duality gap does not necessarily mean it is easy to find the minimum of $g(\boldsymbol{\Lambda})$ in (16). Evaluating $g(\boldsymbol{\Lambda})$ as per (16) requires maximizing the Lagrangian $\boldsymbol{\mathcal{L}}(\mathbf{x},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda})$ in (15). This maximization *may be* difficult to perform depending on the link capacity function $C_{ij}(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h}))$. Nevertheless, Theorem 1 can be exploited to solve (9) at least in certain cases of practical importance. This will be addressed in subsequent submissions. Perhaps more important, Theorem 1 reveals fundamental separation principles of wireless networks.

V. Separation Theorems

A major implication of Theorem 1 is the optimality of conventional layering in wireless networking problems. As is usually the case, the Lagrangian exhibits a separable structure in the sense that it can be written as a sum of terms that depend on a few primal variables. Rearranging terms in (15) and assuming that the optimal set of dual variables $\boldsymbol{\Lambda}^*$ is available, we can write

$$\begin{aligned}
&\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}^*] \qquad (53)\\
&\quad=\sum_{i,k}\left(U_i^k\left(a_i^k\right)-\nu_j^{k^*}a_j^k\right)+\sum_i(\mu_i^*p_i-V_i(p_i))\\
&\qquad+\sum_{i,j}(\xi_{ij}^*-\lambda_{ij}^*)c_{ij}+\sum_{i,j,k}\left(\nu_i^{k^*}-\nu_j^{k^*}-\xi_{ij}^*\right)r_{ij}^k\\
&\qquad+\mathbb{E}_{\mathbf{H}}\left[\sum_{i,j,f}\lambda_{ij}^*C_{ij}(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h}))-\mu_i^*p_{ij}^f(\boldsymbol{h})\right].
\end{aligned}$$

The zero duality gap implies that if $\boldsymbol{\Lambda}^*$ is available, then instead of solving (9), it is possible to solve the (separable) problem

$$\mathsf{P}=\mathsf{D}=g[\boldsymbol{\Lambda}^*]=\max_{\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h})}\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}^*] \qquad (54)$$

where $\boldsymbol{X}$ and $\boldsymbol{p}(\boldsymbol{h})$ satisfy the box constraints.

Because primal variables are decoupled in $\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}^*]$, the maximization in (54) can be split into smaller maximization problems each involving less variables. This separability can be used to prove the next two theorems.

Theorem 3 (Layer Separability): Let λ_{ij}^*, μ_i^*, ν_i^{k*}, and ξ_{ij}^* denote the optimal dual variables that solve (17). Consider the subproblems

$$\mathsf{P}\left(a_i^k\right) = \max_{a_{\min\, i}^k \le a_i^k \le a_{\max\, i}^k} \left[U_i^k\left(a_i^k\right) - \nu_i^{k*} a_i^k\right] \quad (55)$$

$$\mathsf{P}\left(r_{ij}^k\right) = \max_{0\le r_{ij}^k \le r_{\max}} \left[\left(\nu_i^{k*} - \nu_j^{k*} - \xi_{ij}^*\right) r_{ij}^k\right] \quad (56)$$

$$\mathsf{P}(c_{ij}) = \max_{0\le c_{ij} \le c_{\max}} \left[(\xi_{ij}^* - \lambda_{ij}^*) c_{ij}\right] \quad (57)$$

$$\mathsf{P}(p_i) = \max_{0\le p_i \le p_{\max\, i}} \left[\mu_i^* p_i - V_i(p_i)\right]. \quad (58)$$

Define further the optimal power allocation problem

$$\mathsf{P}[\boldsymbol{p}(\boldsymbol{h})] = \max_{0\le p_{ij}^f(\boldsymbol{h}) \le p_{\max}} \mathbb{E}_{\mathbf{H}}\left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i^* p_{ij}^f(\boldsymbol{h})\right]. \quad (59)$$

Then, the optimal utility yield P in (9) is given by

$$\mathsf{P} = \sum_{i,k} \mathsf{P}\left(a_i^k\right) + \sum_{i,j,k} \mathsf{P}\left(r_{ij}^k\right) + \sum_{i,j} \mathsf{P}(c_{ij}) + \sum_i \mathsf{P}(p_i) + \mathsf{P}[\boldsymbol{p}(\boldsymbol{h})] \quad (60)$$

i.e., the primal problem (9) can be separated into the (sub-) problems (55)–(59) without loss of optimality.

Proof: Simply note that the maximization in (54) is subject to box constraints only. Because these do not involve more than one variable, the overall maximum can be obtained as the sum of the maxima of individual terms; i.e.

$$\begin{aligned}
&\max \boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}^*] \\
&= \sum_{i,k} \max\left[U_i^k\left(a_i^k\right) - \nu_j^{k*} a_j^k\right] + \sum_i \max[\mu_i^* p_i - V_i(p_i)] \\
&\quad + \sum_{i,j} \max[(\xi_{ij}^* - \lambda_{ij}^*) c_{ij}] + \sum_{i,j,k} \max\left[\left(\nu_i^{k*} - \nu_j^{k*} - \xi_{ij}^*\right) r_{ij}^k\right] \\
&\quad + \max \mathbb{E}_{\mathbf{H}}\left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i^* p_{ij}^f(\boldsymbol{h})\right]. \quad (61)
\end{aligned}$$

To complete the proof recall that $\mathsf{P} = \mathsf{D} = \max \boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}^*]$ and substitute definitions (55)–(59) into (61). ∎

The rate problem in (55) dictates the amount of traffic allowed into the network. It therefore solves the flow control problem at the transport layer; see Fig. 4. Likewise, (56) represents the network layer routing problem, (57) determines link-level capacities at the data link layer and (58) is the (link layer) power control problem. Finally, (59) represents the resource allocation problem at the physical layer. Thus, it is a consequence of Theorem 3 that layering, in the sense of problem separability as per (60), is optimal in fading wireless networks.

In addition to layer separability, it is not difficult to realize that expectation and maximization can be interchanged in the power allocation subproblem (59). This establishes separability across fading states, a result summarized as follows.

Theorem 4 (Per-Fading-State Separability): Let λ_{ij}^*, denote the optimal dual argument of (17) and define the per-fading-state power allocation problem:

$$\mathsf{P}(\boldsymbol{h}) = \max_{0\le p_{ij}^f(\boldsymbol{h}) \le p_{\max}} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i^* p_{ij}^f(\boldsymbol{h})\right] \quad (62)$$

Then, the optimal power allocation utility $\mathsf{P}[\boldsymbol{p}(\boldsymbol{h})]$ in (59) is given by

$$\mathsf{P}[\boldsymbol{p}(\boldsymbol{h})] = \mathbb{E}_{\mathbf{H}}[\mathsf{P}(\boldsymbol{h})]. \quad (63)$$

Proof: As in (61) note that constraints in (59) involve one variable only. Thus, it is possible to interchange maximization with expectation operators in (59) to obtain

$$\begin{aligned}
&\mathsf{P}[\boldsymbol{p}(\boldsymbol{h})] \\
&= \max_{0\le p_{ij}^f(\boldsymbol{h}) \le p_{\max}} \mathbb{E}_{\mathbf{H}}\left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i^* p_{ij}^f(\boldsymbol{h})\right] \\
&= \mathbb{E}_{\mathbf{H}} \max_{0\le p_{ij}^f(\boldsymbol{h}) \le p_{\max}} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i^* p_{ij}^f(\boldsymbol{h})\right] \\
&= \mathbb{E}_{\mathbf{H}}[\mathsf{P}(\boldsymbol{h})]. \quad (64)
\end{aligned}$$

First and last equality follow from the definitions of $\mathsf{P}[\boldsymbol{p}(\boldsymbol{h})]$ in (59) and $\mathsf{P}(\boldsymbol{h})$ in (63). ∎

With $\boldsymbol{\Lambda}^*$ known, Theorem 4 establishes that the variational problem of finding optimal power functions in (59) reduces to finding optimal power values per fading realization as in (63).

At this point it is worth recalling that Theorems 3 and 4 assume availability of the optimal Lagrange multipliers λ_{ij}^*, μ_i^*, ν_i^{k*} and ξ_{ij}^*. Finding them, while possible, is a nontrivial problem that will be addressed in forthcoming sections. However, it has to be appreciated that Theorems 3 and 4 establish two fundamental properties of wireless networks in the presence of fading: i) the decomposition of the problem into the traditional networking layers can be rendered optimal; and ii) the separability of the resource allocation problem into per-fading-state subproblems is possible. Wireless networks with deterministic channels do not possess either of these two properties.

Of all the per-layer problems (55)–(59), the physical layer optimization in (59) presents the biggest challenge. While some degree of simplification is offered by (62), this still necessitates joint optimization across terminals. The challenge in wireless networks is not as much in cross-layer optimization as in cross-terminal optimization of the physical layer.

VI. Finding Optimal Lagrange Multipliers

Solving the optimal wireless networking problem in (9) can be reduced to finding the optimal dual variables $\boldsymbol{\Lambda}^*$ of (53). This approach not only offers a means of solving (9) but also induces

Layers at terminal T_i

$a_i^{k\,*} \leftarrow \max_{a^k_{\min i} \le a_i^k \le a^k_{\max i}} \left[U_i^k(a_i^k) - \nu_i^{k\,*} a_i^k \right]$	$a_i^{k\,*}$ for all k
$r_{ij}^{k\,*} \leftarrow \max_{0 \le r_{ij}^k \le r_{\max}} \left[\left(\nu_i^{k\,*} - \nu_j^{k\,*} - \xi_{ij}^* \right) r_{ij}^k \right]$	$r_{ij}^{k\,*}$ for $j \in \mathcal{N}(i)$ and all k
$c_{ij}^* \leftarrow \max_{0 \le c_{ij} \le c_{\max}} \left[\left(\xi_{ij}^* - \lambda_{ij}^* \right) c_{ij} \right]$	c_{ij}^* for $j \in \mathcal{N}(i)$
$p_i^* \leftarrow \max_{0 \le p_i \le p_{\max i}} \left[\mu_i^* p_i - V_i(p_i) \right]$	p_i^*

$$\boldsymbol{p}^*(\boldsymbol{h}) \leftarrow \max_{\boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij} \left(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}) \right) - \mu_i^* p_{ij}^f(\boldsymbol{h}) \right]$$

Fig. 4. Having zero duality gap the wireless networking problem can be separated in layers without loss of optimality. Therefore, we can consider separate optimization problems to determine arrival rates $a_i^{k\,*}$ [cf. (55)], link rates $r_{ij}^{k\,*}$ [cf. (56)] link capacities c_{ij}^* [cf. (57)] and average transmitted power p_i^* [cf.(58)]. The physical layer problem can be further separated in per-fading-state subproblems [cf. (59)]. It cannot, alas, be separated in per-terminal problems for general link capacity functions $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$. Thus, the challenge in wireless networking is not as much in cross-layer optimization as in cross-terminal optimization of the physical layer.

separation across layers [cf. Theorem 3] and fading states [cf. Theorem 4]. Because the dual function $g(\boldsymbol{\Lambda})$ is convex, descent algorithms can be used to find $\boldsymbol{\Lambda}^*$. However, $g(\boldsymbol{\Lambda})$ need not be differentiable, and certainly will not be in certain cases. An alternative descent direction is provided by the dual function's subgradient; see, e.g., [32, p. 731].

To find a subgradient of $g(\boldsymbol{\Lambda})$, write (9) in generic form as

$$\mathsf{P} = \max_{\boldsymbol{X} \in \boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})} f_0(\boldsymbol{X}); \quad \text{s. to } \mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})] \ge \mathbf{0} \tag{65}$$

with $f_0(\boldsymbol{X})$ denoting the utility function in (9) and $\mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})] \ge \mathbf{0}$ the constraints (10)–(12). Correspondingly, the Lagrangian (15) in this simplified notation takes the form

$$\boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}] = f_0(\boldsymbol{X}) + \boldsymbol{\Lambda}^T \mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})]. \tag{66}$$

Given a set of multipliers $\boldsymbol{\Lambda}$, a subgradient of the dual function can be obtained from the arguments maximizing the Lagrangian, as detailed in the following theorem. This as well as subsequent results in Theorems 6–8 are known for finite-dimensional optimization problems, [33]. We present them here for the (infinite-dimensional) variational problem (65). The proofs here are patterned after those in [33] and are included to make the presentation self-contained. The reader familiar with subgradient descent algorithms may skip to Section VI-B.

Theorem 5: Let $\boldsymbol{\Lambda}_0 \ge \mathbf{0}$ be an arbitrary dual variable and $\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0)$ primal variables that maximize the Lagrangian in (66) for $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0$; i.e.

$$(\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0)) \in \operatorname*{argmax}_{\boldsymbol{X} \in \boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})} \boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}_0]. \tag{67}$$

Define a subgradient of the dual function at $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0$ as

$$\check{\boldsymbol{g}}(\boldsymbol{\Lambda}_0) = \mathbf{f}[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0)]. \tag{68}$$

The subgradient $\check{\boldsymbol{g}}(\boldsymbol{\Lambda}_0)$ is a descent direction for the dual function, i.e.

$$\check{\boldsymbol{g}}^T(\boldsymbol{\Lambda}_0)(\boldsymbol{\Lambda}_0 - \boldsymbol{\Lambda}^*) \ge g(\boldsymbol{\Lambda}_0) - \mathsf{D} \ge 0 \tag{69}$$

Proof: Since $\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0)$ are optimal primal arguments of the Lagrangian when $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0$, the dual function at $\boldsymbol{\Lambda}_0$ is

$$\begin{aligned} g(\boldsymbol{\Lambda}_0) &:= \max_{\boldsymbol{X} \in \boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})} f_0(\boldsymbol{X}) + \boldsymbol{\Lambda}_0^T \mathbf{f}(\boldsymbol{X}) \\ &= f_0[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0)] + \boldsymbol{\Lambda}_0^T \mathbf{f}[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0)] \end{aligned} \tag{70}$$

where we have used the Lagrangian expression in (66). For any optimal $\boldsymbol{\Lambda}^*$ the Lagrangian value $\boldsymbol{\mathcal{L}}[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0), \boldsymbol{\Lambda}^*]$ cannot exceed the maximum Lagrangian value over all possible $(\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}))$; hence

$$\begin{aligned} g(\boldsymbol{\Lambda}^*) &= \max_{\boldsymbol{X} \in \boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{h})} f_0(\boldsymbol{X}) + \boldsymbol{\Lambda}^{*T} \mathbf{f}(\boldsymbol{X}) \\ &\ge f_0[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0)] + \boldsymbol{\Lambda}^{*T} \mathbf{f}[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0)]. \end{aligned} \tag{71}$$

Note that the RHS of (70) and (71) differ only in the dual variable multiplying the constraint differences. Thus, subtracting (70) from (71) yields

$$g(\boldsymbol{\Lambda}^*) - g(\boldsymbol{\Lambda}_0) \ge (\boldsymbol{\Lambda}^* - \boldsymbol{\Lambda}_0)^T \mathbf{f}[\boldsymbol{X}^\dagger(\boldsymbol{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h}, \boldsymbol{\Lambda}_0)]. \tag{72}$$

Substituting the definition of $\check{\boldsymbol{g}}(\boldsymbol{\Lambda}_0)$ in (68) into the inequality in (72), reordering terms and substituting $\mathsf{D} = g(\boldsymbol{\Lambda}^*)$ the result in (69) follows. ∎

Since the inner product $\check{\boldsymbol{g}}^T(\boldsymbol{\Lambda}_0)(\boldsymbol{\Lambda}_0 - \boldsymbol{\Lambda}^*) = (-\check{\boldsymbol{g}}^T(\boldsymbol{\Lambda}_0))(\boldsymbol{\Lambda}^* - \boldsymbol{\Lambda}_0)$ is positive, (69) proves that the angle between $-\check{\boldsymbol{g}}^T(\boldsymbol{\Lambda}_0)$ and $\boldsymbol{\Lambda}^* - \boldsymbol{\Lambda}_0$ is less than $\pi/2$ radians. Therefore, standing at $\boldsymbol{\Lambda}_0$ the negative of the subgradient

points "towards," i.e., with an angle smaller than $\pi/2$ radians, the optimal argument.

Because there might be more than one argument maximizing (67), the $\arg\max$ operator does not always specify a value but a set, as signified by the symbol $\in$ in (67). One should thus interpret $\boldsymbol{X}^{\dagger}(\boldsymbol{\Lambda})$ as any element of this set.

A. Subgradient Descent Algorithm

Given the subgradient of the dual function specified in Theorem 5, it is possible to devise a descent algorithm for obtaining the optimal multipliers $\boldsymbol{\Lambda}^*$ and the minimum dual value $\mathsf{D} = \mathsf{P}$. With iterations indexed by t, start with given dual variables $\boldsymbol{\Lambda}(t)$ and compute arguments $[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]$ that maximize the Lagrangian in (66), i.e.

$$\begin{aligned}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)] &\in \underset{\boldsymbol{X}\in\boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h})\in\boldsymbol{\mathcal{P}}(\boldsymbol{h})}{\operatorname{argmax}} \boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}(t)] \\ &= \underset{\boldsymbol{X}\in\boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h})\in\boldsymbol{\mathcal{P}}(\boldsymbol{h})}{\operatorname{argmax}} f_0(\boldsymbol{X}) + \boldsymbol{\Lambda}^T(t)\mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})]. \quad (73)\end{aligned}$$

Using (68), it follows that a subgradient of the dual function at $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(t)$ is given by $\breve{\boldsymbol{g}}(t) := \breve{\boldsymbol{g}}[\boldsymbol{\Lambda}(t)] = \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]$. Therefore, the set of dual variables is updated as

$$\begin{aligned}\boldsymbol{\Lambda}(t+1) &= [\boldsymbol{\Lambda}(t) - \epsilon_t \breve{\boldsymbol{g}}(t)]^+ \\ &= [\boldsymbol{\Lambda}(t) - \epsilon_t \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]]^+ \quad (74)\end{aligned}$$

where $[\cdot]^+$ denotes componentwise maximum between 0 and the value inside the square brackets; while ϵ_t is a properly selected stepsize. Because $-\breve{\boldsymbol{g}}(t)$ points towards $\boldsymbol{\Lambda}^*$ the iterates in (74) can approach $\boldsymbol{\Lambda}^*$. As the following result shows, this is indeed true in a well-defined sense.

Theorem 6: Consider the subgradient descent iteration in (74) and define the dual value at iteration t as $g(t) := g[\boldsymbol{\Lambda}(t)]$. Let $G := \max_{\boldsymbol{X}\in\boldsymbol{\mathcal{B}}, \boldsymbol{p}(\boldsymbol{h})\in\boldsymbol{\mathcal{P}}(\boldsymbol{h})} ||\mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})]||$ be a bound on the norm of the subgradient of the dual function. The 2-norm distances $||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||$ of iterates $\boldsymbol{\Lambda}(t)$ from the optimal argument $\boldsymbol{\Lambda}^*$ at iterations t and $t+1$ satisfy the relation

$$||\boldsymbol{\Lambda}(t+1) - \boldsymbol{\Lambda}^*||^2 \leq ||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||^2 + \epsilon_t^2 G^2 - 2\epsilon_t[g(t) - \mathsf{D}]. \quad (75)$$

Proof: Consider the square of the 2-norm distance to the optimum $||\boldsymbol{\Lambda}(t+1) - \boldsymbol{\Lambda}^*||$ at iteration $t+1$. Using (74), it can be bounded through the 2-norm at iteration t as

$$\begin{aligned}&||\boldsymbol{\Lambda}(t+1) - \boldsymbol{\Lambda}^*||^2 \\ &\quad = ||[\boldsymbol{\Lambda}(t) - \epsilon_t \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]]^+ - \boldsymbol{\Lambda}^*||^2 \quad (76) \\ &\quad \leq ||\boldsymbol{\Lambda}(t) - \epsilon_t \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)] - \boldsymbol{\Lambda}^*||^2 \quad (77) \\ &\quad = ||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||^2 + \epsilon_t^2 ||\mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]||^2 \\ &\qquad - 2\epsilon_t \mathbf{f}^T[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)](\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*) \quad (78) \\ &\quad \leq ||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||^2 + \epsilon_t^2 G^2 \\ &\qquad - 2\epsilon_t \mathbf{f}^T[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)](\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*) \quad (79)\end{aligned}$$

Equation (77) is true because $\boldsymbol{\Lambda}^* \geq \mathbf{0}$; and thus, when some components of $\boldsymbol{\Lambda}(t) - \epsilon_t \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]$ are negative, the distance to $\boldsymbol{\Lambda}^*$ is reduced by setting them to 0. Equivalently, allowing for negative components as in (77) increases the distance to $\boldsymbol{\Lambda}^*$ with respect to when this negative components are set to 0 as in (76). Equation (78) follows after expanding the square in (77), and (79) by using the bound G on the norm of $\mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})]$.

As per (73), $\mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]$ is a subgradient of $g(\boldsymbol{\Lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\Lambda}(t)$ [cf. Theorem 5]. Substituting $\boldsymbol{\Lambda}_0 = \boldsymbol{\Lambda}(t)$ and $\breve{\boldsymbol{g}}(\boldsymbol{\Lambda}_0) = \breve{\boldsymbol{g}}[\boldsymbol{\Lambda}(t)] = \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)]$ into (69) yields

$$\mathbf{f}^T[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}, t)](\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*) \geq g(t) - \mathsf{D} \quad (80)$$

where we used the definition $g(t) := g[\boldsymbol{\Lambda}(t)]$. The last term in (79) can be replaced by the bound in (80) to arrive at (75). ■

Since all primal variables are constrained to the bounded regions $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{P}}(\boldsymbol{h})$, the bound G on the subgradient norm is finite. Given that D denotes the minimum of $g(t)$, it clearly holds that $g(t) - \mathsf{D} \geq 0$. Thus, at each iteration the distance between the current dual iterate $\boldsymbol{\Lambda}(t)$ and the optimal dual variable $\boldsymbol{\Lambda}^*$ is reduced by (at least) $2\epsilon_t[g(t) - \mathsf{D}]$ and increased by (at most) $\epsilon_t^2 G^2$. For ϵ_t sufficiently small, the reduction $2\epsilon_t[g(t) - \mathsf{D}]$ will dominate the increase $\epsilon_t^2 G^2$ and consequently $\boldsymbol{\Lambda}(t)$ will approach $\boldsymbol{\Lambda}^*$.

With a fixed stepsize $\epsilon_t = \epsilon$ however, there is a limit on how close $\boldsymbol{\Lambda}(t)$ can come to $\boldsymbol{\Lambda}^*$. For any given ϵ, $\epsilon^2 G^2$ will eventually become larger than $2\epsilon[g(t) - \mathsf{D}]$ preventing the gap $[g(t) - \mathsf{D}]$ from converging to zero. This is not a limitation of the approach but a consequence of the fact that for nondifferentiable functions the norm of the subgradient $||\breve{\boldsymbol{g}}(\boldsymbol{\Lambda})||$ does not necessarily vanish as $\boldsymbol{\Lambda}$ approaches $\boldsymbol{\Lambda}^*$. Therefore, the iteration in (74) is not convergent. Rather, the iterates $\boldsymbol{\Lambda}(t)$ approach $\boldsymbol{\Lambda}^*$ until $\epsilon^2 G^2$ starts dominating $2\epsilon[g(t) - \mathsf{D}]$.

These considerations motivate the use of vanishing stepsize sequences, i.e., $\lim_{t\to\infty} \epsilon_t = 0$, so that as the duality gap $[g(t) - \mathsf{D}]$ approaches zero, so does ϵ_t. This allows for $2\epsilon_t[g(t) - \mathsf{D}]$ to always dominate $\epsilon_t^2 G^2$ and leads to the following convergence result.

Theorem 7: If stepsizes ϵ_t in the subgradient descent iteration (74) satisfy

$$\sum_{t=1}^{\infty} \epsilon_t = \infty, \quad \lim_{t\to\infty} \epsilon_t = 0 \quad (81)$$

then the limit of the iterates exists and

$$\lim_{t\to\infty} \boldsymbol{\Lambda}(t) = \boldsymbol{\Lambda}^*. \quad (82)$$

Proof: Arguing by contradiction, suppose that $\lim_{t\to\infty} \boldsymbol{\Lambda}(t) \neq \boldsymbol{\Lambda}^*$ which implies $\lim_{t\to\infty} g(t) \neq \mathsf{D}$. Since $g(t) - \mathsf{D}$ cannot be negative, there must exist a time index t_1 and a constant $\delta > 0$ such that $g(t) - \mathsf{D} \geq \delta$ for all $t \geq t_1$. For these time indices Theorem 6 asserts that [cf. (75)]

$$\begin{aligned}||\boldsymbol{\Lambda}(t+1) - \boldsymbol{\Lambda}^*||^2 &\leq ||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||^2 + \epsilon_t^2 G^2 - 2\epsilon_t \delta \\ &= ||\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}^*||^2 + \epsilon_t(\epsilon_t G^2 - 2\delta). \quad (83)\end{aligned}$$

Furthermore, given that $\epsilon_t \to 0$ there exists a time index t_2 such that for all $t \geq t_2$, we have $\epsilon_t < \delta/G^2$. Consequently, for all $t \geq t_1$ and $t \geq t_2$ it holds that

$$||\mathbf{\Lambda}(t+1) - \mathbf{\Lambda}^*||^2 \leq ||\mathbf{\Lambda}(t) - \mathbf{\Lambda}^*||^2 - \epsilon_t \delta. \tag{84}$$

Upon defining time index $t_0 := \max(t_1, t_2)$ application of (84) recursively up until t_0 yields

$$||\mathbf{\Lambda}(t+1) - \mathbf{\Lambda}^*||^2 \leq ||\mathbf{\Lambda}(t_0) - \mathbf{\Lambda}^*||^2 - \delta \sum_{s=t_0}^{t} \epsilon_s. \tag{85}$$

But (81) ensures that the sum of stepsizes is divergent. Thus, $\sum_{s=t_0}^{\infty} \epsilon_s = \infty$ implying that there must exist a time index τ such that $\sum_{s=t_0}^{\tau} \epsilon_s > ||\mathbf{\Lambda}(t_0) - \mathbf{\Lambda}^*||^2/\delta$. Writing (85) for $t = \tau$ yields

$$\begin{aligned} ||\mathbf{\Lambda}(\tau+1) - \mathbf{\Lambda}^*||^2 &< ||\mathbf{\Lambda}(t_0) - \mathbf{\Lambda}^*||^2 - \delta||\mathbf{\Lambda}(t_0) - \mathbf{\Lambda}^*||^2/\delta \\ &= 0. \end{aligned} \tag{86}$$

This is a contradiction because the RHS is clearly nonnegative. Hence, it must hold that $\lim_{t\to\infty} \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*$. ∎

The conditions (81) on the stepsize sequence are certainly minimal. They are satisfied for instance by sequences $\epsilon_t = \epsilon_1/(t+\epsilon_2)^\beta$ with $\beta > 0$ for arbitrary positive constants ϵ_1 and ϵ_2. Nonetheless, constant stepsizes $\epsilon_t = \epsilon$ for all t, are still desirable to enable adaptability. In this case, it can be proved that as $t \to \infty$, $\mathbf{\Lambda}(t)$ "stays close" to $\mathbf{\Lambda}^*$. A formal statement and proof of this claim are provided next.

Theorem 8: Consider the subgradient descent iteration in (74) with constant stepsizes $\epsilon_t = \epsilon$ for all t. With $G := \max_{\boldsymbol{X}\in\mathcal{B},\boldsymbol{p}(\boldsymbol{h})\in\mathcal{P}(\boldsymbol{h})} ||\mathbf{f}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})]||$ denoting the subgradient norm bound in Theorem 6, it holds that:

(i) The best dual value at time t, $g_{\text{best}}(t) := \min_{s\in[0,t]} g(t)$, converges to a value within $\epsilon G^2/2$ of the optimum D, i.e.

$$\lim_{t\to\infty} g_{\text{best}}(t) - \mathsf{D} \leq \epsilon G^2/2. \tag{87}$$

(ii) The average of the dual iterates $\bar{\mathbf{\Lambda}}(t) := (1/t)\sum_{s=1}^{t} \mathbf{\Lambda}(s)$, converges to a point whose optimality gap is less than $\epsilon G^2/2$, i.e.

$$g[\lim_{t\to\infty} \bar{\mathbf{\Lambda}}(t)] - \mathsf{D} \leq \epsilon G^2/2. \tag{88}$$

Proof: Apply (75) recursively up until $t = 1$ to obtain the bound

$$\begin{aligned} &||\mathbf{\Lambda}(t+1) - \mathbf{\Lambda}^*||^2 \\ &\quad \leq ||\mathbf{\Lambda}(1) - \mathbf{\Lambda}^*||^2 + \sum_{s=1}^{t} \epsilon^2 G^2 - \sum_{s=1}^{t} 2\epsilon[g(s) - \mathsf{D}] \end{aligned} \tag{89}$$

where $\epsilon_s = \epsilon$ was used for all s. Note that the LHS of (89) is positive, i.e., $||\mathbf{\Lambda}(t+1) - \mathbf{\Lambda}^*||^2 \geq 0$. Using this, simplifying the sums and rearranging terms in (89) yields

$$2\epsilon \sum_{s=1}^{t} [g(s) - \mathsf{D}] \leq ||\mathbf{\Lambda}(1) - \mathbf{\Lambda}^*||^2 + tG^2\epsilon^2. \tag{90}$$

Results (87) and (88) follow from (90). To obtain (87) note that the definition of the best dual value at time t, namely $g_{\text{best}}(t) := \min_{s\in[0,t]} g(t)$, implies that

$$\sum_{s=1}^{t} [g(s) - \mathsf{D}] \geq \sum_{s=1}^{t} [\min_{s\in[0,t]} g(t) - \mathsf{D}] = t[g_{\text{best}}(t) - \mathsf{D}]. \tag{91}$$

Substituting (91) into (90) and dividing both sides of the inequality by $2\epsilon t$ yields

$$g_{\text{best}}(t) - \mathsf{D} \leq \frac{||\mathbf{\Lambda}(1) - \mathbf{\Lambda}^*||^2}{2\epsilon t} + \frac{\epsilon G^2}{2}. \tag{92}$$

Equation (87) follows after taking the limit of (92) as $t \to \infty$.

To establish (88), use the convexity of $g(\mathbf{\Lambda})$ to write

$$g[\bar{\mathbf{\Lambda}}(t)] := g\left[\frac{1}{t}\sum_{s=1}^{t} \mathbf{\Lambda}(s)\right] \leq \frac{1}{t}\sum_{s=1}^{t} g[\mathbf{\Lambda}(s)] := \frac{1}{t}\sum_{s=1}^{t} g(s) \tag{93}$$

where the first and last equality follow from the definitions of $\bar{\mathbf{\Lambda}}(t)$ and $g(s)$, respectively. Dividing (90) by $2\epsilon t$ and combining the resulting inequality with the one in (93) yields

$$g[\bar{\mathbf{\Lambda}}(t)] - \mathsf{D} \leq \frac{||\mathbf{\Lambda}(1) - \mathbf{\Lambda}^*||^2}{2t\epsilon} + \frac{\epsilon G^2}{2}. \tag{94}$$

Because $g(\mathbf{\Lambda})$ is continuous, we have $g[\lim_{t\to\infty} \bar{\mathbf{\Lambda}}(t)] = \lim_{t\to\infty} g[\bar{\mathbf{\Lambda}}(t)]$. Therefore, the limit of (94) as $t \to \infty$ establishes (88). ∎

As commented after Theorem 6, the subgradient descent algorithm (73)–(74) does not necessarily converge for fixed stepsizes. Nonetheless, a reasonable approximation to $\mathbf{\Lambda}^*$ is achieved by $\mathbf{\Lambda}_{\text{best}}(t)$ defined as the argument for which $g[\mathbf{\Lambda}_{\text{best}}(t)] = g_{\text{best}}(t)$. The quality of this approximation is measured by the optimality gap $g[\mathbf{\Lambda}_{\text{best}}(t)] - \mathsf{D}$ that can be made arbitrarily small with properly selected stepsize ϵ.

By definition $\mathbf{\Lambda}_{\text{best}}(t)$ is the best approximation to $\mathbf{\Lambda}^*$ that can be obtained by (73)–(74) with fixed stepsizes. To find $\mathbf{\Lambda}_{\text{best}}(t)$ though, requires access to the dual values $g(t)$, which may not be available; see Section VI-B. In such circumstances, a reasonably good approximation to $\mathbf{\Lambda}^*$ is the average $\bar{\mathbf{\Lambda}}(t)$ of iterates $\mathbf{\Lambda}(t)$.

B. Layers and Layer Interfaces

The subgradient descent iterations (73) and (74) shed light on the interaction between layers. Returning to the notation in

(53), the Lagrangian $\mathcal{L}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}(t)]$ used in the primal iteration (73) takes the form

$$\begin{aligned}&\mathcal{L}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}(t)] \qquad (95)\\ &= \sum_{i,k}\left(U_i^k\left(a_i^k\right) - \nu_i^k(t)a_j^k\right) + \sum_i\left(\mu_i(t)p_i - V_i(p_i)\right)\\ &+ \sum_{i,j}\left(\xi_{ij}(t) - \lambda_{ij}(t)\right)c_{ij}\sum_{i,j,k}\left(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\right)r_{ij}^k\\ &+ \mathbb{E}_{\mathbf{H}}\left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\left(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})\right) - \mu_i(t)p_{ij}^f(\boldsymbol{h})\right].\end{aligned}$$

Inspection of (95) reveals that the separable structure used to establish Theorems 3 and 4 is not specific to $\boldsymbol{\Lambda}^*$. Therefore, the maximization required for the primal iteration (73) can be likewise decomposed in per-layer and per-fading-state parts. The entries of $\boldsymbol{X}(t)$ and the power function $\boldsymbol{p}(\boldsymbol{h}; t)$ are thus

$$a_i^k(t) = \underset{a_{\min\, i}^k \le a_i^k \le a_{\max\, i}^k}{\arg\max}\ \left[U_i^k\left(a_i^k\right) - \nu_i^k(t)a_i^k\right] \qquad (96)$$

$$r_{ij}^k(t) = \underset{0\le r_{ij}^k \le r_{\max}}{\arg\max}\ \left[\left(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\right)r_{ij}^k\right] \qquad (97)$$

$$c_{ij}(t) = \underset{0\le c_{ij}\le c_{\max}}{\arg\max}\ \left[\left(\xi_{ij}(t) - \lambda_{ij}(t)\right)c_{ij}\right] \qquad (98)$$

$$p_i(t) = \underset{0\le p_i \le p_{\max\, i}}{\arg\max}\ \left[\mu_i(t)p_i - V_i(p_i)\right] \qquad (99)$$

$$\boldsymbol{p}(\boldsymbol{h}; t) = \underset{\boldsymbol{p}(\boldsymbol{h})\in\boldsymbol{\mathcal{P}}(\boldsymbol{h})}{\arg\max}\left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i(t)p_{ij}^f(\boldsymbol{h})\right]. \qquad (100)$$

The subgradient $\breve{\boldsymbol{g}}(t) = \mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}; t)]$ can be likewise separated. Components of the vector function $\mathbf{f}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h}; t)]$ are as specified in (10)–(12). Therefore, the dual iteration (74) can be written explicitly as

$$\nu_i^k(t+1) = \left[\nu_i^k(t) - \epsilon_t\left[\sum_{j\in n(i)}\left(r_{ij}^k(t) - r_{ji}^k(t)\right) - a_i^k(t)\right]\right]^+ \qquad (101)$$

$$\xi_{ij}(t+1) = \left[\xi_{ij}(t) - \epsilon_t\left[c_{ij}(t) - \sum_k r_{ij}^k(t)\right]\right]^+ \qquad (102)$$

$$\lambda_{ij}(t+1) = \left[\lambda_{ij}(t) - \epsilon_t\left[\mathbb{E}_{\mathbf{H}}\left[\sum_{f\in\mathcal{F}}C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}; t))\right] - c_{ij}(t)\right]\right]^+ \qquad (103)$$

$$\mu_i(t+1) = \left[\mu_i(t) - \epsilon_t\left[p_i - \mathbb{E}_{\mathbf{H}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}p_{ij}^f(\boldsymbol{h}; t)\right]\right]\right]^+ \qquad (104)$$

The argument to be optimized in (96) is solely parameterized by $\nu_i^k(t)$. Thus, $a_i^k(t)$ can be determined once the multiplier $\nu_i^k(t)$ associated with the flow conservation constraint is available. Likewise, $r_{ij}^k(t)$ is determined by flow conservation multipliers $\nu_i^k(t)$ and $\nu_j^k(t)$ and link capacity multipliers $\xi_{ij}(t)$. In general, all the primal iterations (96)–(100) depend on multipliers associated with no more than two types of constraints. Similar comments apply to the dual iterations (101)–(104). The update of $\mu_i(t)$ in (104) for instance, depends on the total power $p_i(t)$ and the power function $\boldsymbol{p}(\boldsymbol{h}; t)$. In general, the multiplier updates depend on no more than two different types of primal variables.

The fact that primal and dual variable updates depend on only two types of variables suggests an interpretation of (96)–(104) in terms of layers and layer interfaces. As in Section V, the flow control problem (96) is associated with the transport layer, the link rate problem (97) with the routing layer, link capacity (98) and power control (99) problems are solved at the link layer, while the power assignment (100) pertains to the physical layer. Because the dual variables in (96)–(100) are not optimal, it becomes necessary to communicate variables across layer interfaces. These interfaces are defined by the dual variable updates (101)–(104). Thus, the update of multipliers $\nu_i^k(t)$ in (101) defines the interface between the network and transport layers, while the update (102) characterizes the link to network layer interface. Because there are two problems being solved at the link layer, (103) defines the interface between the physical layer and the link capacity subproblem, and (104) interfaces the physical layer with the power control subproblem.

Fig. 5 shows a schematic representation of the layers and their interfaces. At the bottom of the stack the physical layer solves (100) to find the power distribution $\boldsymbol{p}(\boldsymbol{h}; t)$. Due to coupling that in general is introduced by the function $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$, the physical layer optimization cannot be separated in per-terminal optimization problems and is therefore represented as a common substrate supporting per-terminal stacks. To compute $\boldsymbol{p}(\boldsymbol{h}; t)$, the physical layer receives multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ from the physical-link interface; see also Fig. 6 for a detailed account of variable communications between layers and interfaces.

At the link layer each terminal i maintains variables representing the average link capacities $c_{ij}(t)$ to neighbors T_j, $j \in n(i)$ and the average transmit-power $p_i(t)$. These variables are computed by solving (98) and (99). In turn, this requires dual variables $\lambda_{ij}(t)$ and $\mu_i(t)$ communicated from the physical-link interface and $\xi_{ij}(t)$ communicated from the link-network interface.

As is true for physical and link, all layers compute network variables of interest based on dual variables received from adjacent interfaces. That way, the network layer maintains variables r_{ij}^k for neighbors $j \in n(i)$ and flows k that determine local routing decisions. These are updated as per (97) using multipliers $\xi_{ij}(t)$ received from the link-network interface and ν_i^k and ν_j^k, $j \in n(i)$, from the network-transport interface. The transport layer, finally, keeps variables a_i^k determining the average rate at which packets belonging to the k-th flow are accepted into the network by terminal T_i. These are updated as per (96) using multipliers ν_i^k received from the network-transport interface.

Interfaces in turn, update dual variables using information received from adjacent layers. The physical-link interface computes dual variables $\lambda_{ij}(t)$ for $j \in n(i)$ and $\mu_i(t)$. This is because the multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ are respectively associ-

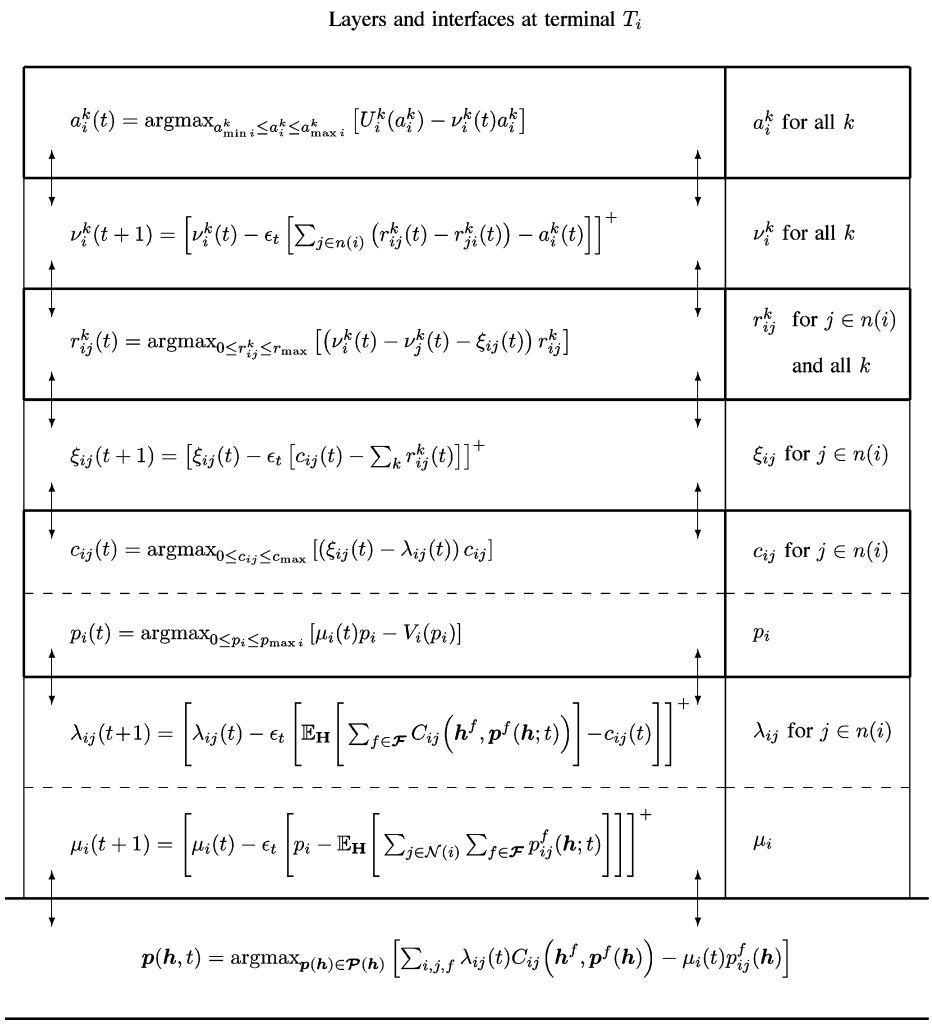

Fig. 5. The subgradient descent iteration (96)–(104) can be interpreted in terms of layers and layer interfaces. Layers keep variables of interest to the network, e.g., link transmission rates r_{ij}^k at the network layer, that they update according to primal iterations (96)–(100). Layer interfaces maintain (auxiliary) dual variables updated as per the dual iterations (101)–(104). Communication of variables across layers and interfaces is restricted to adjacent entities; i.e., layers receive variables from, and transmit to, adjacent interfaces. Interfaces exchange variables with adjacent layers. Note that in general the physical layer optimization problem cannot be separated in per-terminal problems.

ated with the link capacity (10) and power (11) constraints that relate physical-level variables $\boldsymbol{p}(\boldsymbol{h})$ with link-level quantities c_{ij} and p_i. The updates (103) and (104) carried at the physical-link interface require variables $\boldsymbol{p}(\boldsymbol{h};t)$ communicated from the physical layer and variables $c_{ij}(t)$ and $p_i(t)$ from the link layer.

Likewise, the link-network interface keeps track of one multiplier $\xi_{ij}(t)$ per neighbor T_j, $j \in n(i)$. These are associated with the rate constraints in (12) that couple link variables c_{ij} with network variables r_{ij}^k. Updates of $\xi_{ij}(t)$ are specified in (102), being determined by variables $c_{ij}(t)$ and $r_{ij}^k(t)$, respectively communicated from the link and network layers. The network-transport interface, finally, maintains dual variables $\nu_i^k(t)$ associated with the flow conservation constraints in (12) that couple network r_{ij}^k and transport a_i^k variables. These $\nu_i^k(t)$ variables are updated as per (101) using $r_{ij}^k(t)$ and $a_i^k(t)$ received from the network and transport layer, respectively.

As time progresses, the interface variables $\lambda_{ij}(t)$, $\mu_i(t)$, $\xi_{ij}(t)$ and $\nu_i^k(t)$ converge to the optimal multipliers λ_{ij}^*, μ_i^*, ξ_{ij}^* and ν_i^{k*} [cf. Theorem 8]—or a point close to them if the stepsize ϵ_t is fixed [cf. Theorem 8]—enabling computation of optimal network variables $\boldsymbol{p}^*(\boldsymbol{h})$, p_i^*, c_{ij}^*, $r_{ij}^{k\,*}$ and a_i^{k*}.

As depicted in Fig. 6, variable updates at the network layer and the network-transport interface require communication with neighboring nodes. Indeed, note that the update of $r_{ij}^k(t)$ in (97) depends on the local—i.e., kept at T_i—multipliers $\xi_{ij}(t)$ and $\nu_i^k(t)$ as well as multipliers $\nu_j^k(t)$. The latter need to be communicated from the network-transport interface of neighboring terminals T_j, $j \in n(i)$. Analogously, to update $\nu_i^k(t)$ as per (101) requires local variables $a_i^k(t)$ and $r_{ij}^k(t)$ and neighboring—i.e., kept at neighbors T_j with $j \in n(i)$—variables $r_{ji}^k(t)$.

The algorithmic complexity of the subgradient iteration (96)–(104) is determined by the complexity of the optimal power allocation problem (100). Problems (96)–(98) involve maximization of simple single-variable expressions while the dual variable updates (103)–(104) entail a few simple algebraic operations. The power allocation problem (100) at the physical layer, however, requires joint optimization across terminals. An important challenge in wireless networking is cross-terminal optimization at the physical layer. As we will discuss in forth-

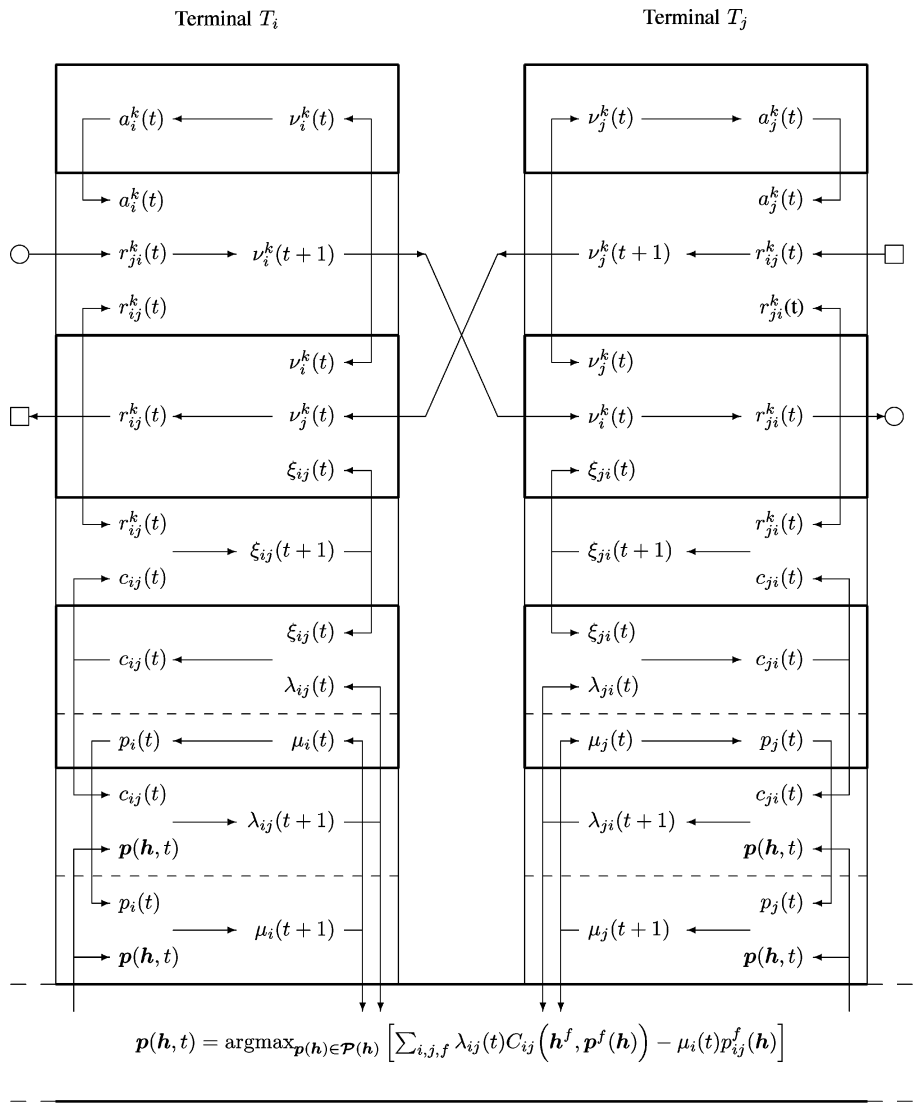

Fig. 6. Communication of variables across layers and interfaces. At the network layer and the network-transport interface exchange of variables between neighboring terminals is necessary. Multipliers $\nu_j^k(t)$ ($\nu_i^k(t)$) are communicated from T_j's (T_i's) network-transport interface to T_i's (T_j's) network layer. Link rate variables $r_{ji}(t)$ ($r_{ij}(t)$) are sent from T_j's (T_i's) network layer to T_i's (T_j's) network-transport interface as represented by the circles (squares).

coming submissions depending on the model determining the function $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$, (100) ranges from simple decomposable problems to provably intractable formulations. A preview of these practical considerations is presented in the next section.

VII. Power Allocation at the Physical Layer

In either the layered architecture of Fig. 4 or the layers and interfaces architecture of Fig. 5 it is necessary to solve power allocation problems respectively given by (62) and (100). Both of these problems comprise a weighted sum rate maximization minus a linear term and are further separable on the per-frequency subproblems

$$\boldsymbol{p}^f(\boldsymbol{h}) = \arg \max_{\boldsymbol{p}(\boldsymbol{h})\in\mathcal{P}(\boldsymbol{h})} \left[\sum_{i,j} \lambda_{ij} C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})) - \mu_i p_{ij}^f(\boldsymbol{h})\right]. \quad (105)$$

To obtain the problem in (62) replace λ_{ij} and μ_i in (105) by λ_{ij}^* and μ_i^*. To obtain (100) substitute by $\lambda_{ij}(t)$ and $\mu_i(t)$.

Because the remaining operations required to find the optimal network are simple [cf. Figs. 4 and 5] it can be said that the practical importance of Theorem 1 is to reduce the difficulty of solving (9)–(12) to that of solving (105). Most functions $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$ used as physical layer models are not concave and for that reason the problem in (105) may be difficult to solve. Still the reduction in computational complexity is significant.

Consider, e.g., the network of Fig. 1 when all nodes are destinations of some flows and with 3 frequencies available. Considering that there are 30 links in the network it is not difficult to see that there are 56 admission control variables a_i^k, 210 routing variables r_{ij}^k, 30 link rates c_{ij}, 8 average powers p_i and *for each* fading state $\boldsymbol{h}$, the total number of powers $\boldsymbol{p}(\boldsymbol{h})$ that needs to be determined is 90. The complexity of a constrained optimization problem in a space with thousands of variables [cf. (9)–(12)] is reduced to a maximization in a space with 30 variables [cf. (105)].

Interestingly, (9)–(12) is also simpler to solve than its deterministic version obtained by removing the expected value operators from (10) and (11). Finding the optimal wireless network

for the deterministic version of (9)–(12) implies solving a non-convex constrained optimization problem with 394 variables. To solve (9)–(12) the significant complexity is reduced to the solution of (105) which involves only 30 variables.

While this reduction in complexity is significant, from a practical perspective it is of interest in medium sized networks only. In general, the problem is intractable in primal form and significantly less but still intractable in dual form. An interesting question is wether there exist problems that are intractable in primal form, or equivalently, with intractable deterministic counterparts, but tractable in dual form.

An example where the deterministic version of (9)–(12) is intractable but (105) is simple to solve is when terminals use orthogonal channels but the function $C(h_{ij}^f p_{ij}^f)$ that maps received power to capacities is not concave. If, e.g., M adaptive modulation and coding (AMC) modes are used $C(h_{ij}^f p_{ij}^f)$ is the staircase function

$$C\left(h_{ij}^f p_{ij}^f\right) = \sum_{m=1}^{M} C_m \mathbb{I}\left(h_{ij}^f p_{ij}^f/\sigma_{ij}^f > P_m\right) \tag{106}$$

where the m-tm mode is used for received powers $h_{ij}^f p_{ij}^f/\sigma_{ij}^f > P_m$ with corresponding rate $\sum_{n=1}^{m} C_n$. Under these assumptions the optimal power allocation problem in (105) reduces to the solution of

$$p_{ij}^f(\boldsymbol{h}) = \arg \max_{0\le p_{ij}^f(\boldsymbol{h})\le p_{\max}} \left[\lambda_{ij} C\left(h_{ij}^f p_{ij}^f\right) - \mu_i p_{ij}^f(\boldsymbol{h})\right]. \tag{107}$$

The solution of (107) is easy to find by evaluating the argument for the AMC transition powers $p_{ij}^f = P_m \sigma_{ij}^f/h_{ij}^f$.

A more practical situation of wireless networks using FDMA with spatial reuse as a physical layer is considered in [34]. In these networks, terminals cannot share frequencies with neighbors but can share them with terminals sufficiently far apart. It is possible to show that in some networks using FDMA with spatial reuse the resource allocation problem (105) is equivalent to a linear program.

In scenarios where (105) cannot be reduced to a tractable optimization problem, approximate solutions are necessary. This is the case, e.g., of interference limited physical layers. It is worth noting that finding an approximate solution to (105) is not equivalent to relaxing the non-convex constraint in (10). Also, the lack of duality gap implies that in a subgradient descent algorithm primal and dual values get close to each other. This property can be used to gauge the performance penalty of using approximate solutions of (105). A comprehensive development of these observations is beyond the scope of this paper but is undertaken in [35], [36].

VIII. Concluding Remarks

This work has shown that optimal wireless network designs with random links adhering to a continuous fading distribution exhibit zero duality gap, even if the underlying optimization problems are generally nonconvex. Pretty much in the spirit of Shannon's separation principle of optimal source and channel coding, this result implies that separating wireless network designs in layers and per-fading state subproblems is optimal. Separability into layers and layer interfaces was further established as a consequence of subgradient descent iterations for the dual function.

It is important to clarify that no argument has been made here on the computational tractability of optimal wireless network designs. It remains open for future research to delineate which sub-classes of problems are intractable in primal form but tractable in their dual formulations. Our preliminary research suggests that certain classes of FDMA networks with spatial reuse fall in this category.

Appendix
On Lyapunov's Theorem and Blackwell's Extension

The proof of Theorem 1 uses Lyapunov's convexity Theorem [29], which has been also invoked in [28] to prove the lack of duality gap in non-convex DSL rate optimization problems. Despite the apparent similarity, the proofs of Theorem 1 and the corresponding [28, Theorem 7] use different methods. Even though Lyapunov's convexity theorem is cited in [28], the proof that the duality gap is null presented by [28, Theorem 7] uses an extension by Blackwell. Blackwell's extension cannot be used to prove Theorem 1 of our manuscript. Rather, our proof is actually based on Lyapunov's theorem.

To elaborate the differences in the method of proof it is worth noting that the problem in [28] is a particular case of (9)–(12) and, as is often the case, tools used in the particular case cannot be extrapolated verbatim to the more general formulation. Using the notation introduced in our manuscript, DSL rate optimization can be written as

$$\begin{aligned} \mathsf{P} = \max & \sum_i U_i(c_i) \\ c_i = & \int_{\mathcal{F}} C_i(\mathbf{f}, \boldsymbol{p}(\mathbf{f}))d\mathbf{f}, \quad \int_{\mathcal{F}} p_i(\mathbf{f})d\mathbf{f} \le p_i \le p_{\max} \end{aligned} \tag{108}$$

where c_i denotes the communication rate of the i-th user, $p_i \le p_{\max}$ the average power budget and $\boldsymbol{p}(\mathbf{f})$ the power allocation across frequencies. In DSL, the number of frequencies available is infinite implying that the sums over frequencies in (10) and (11) are replaced by the integrals in (108). While there is no fading in DSL, the infinite number of frequencies plays an analogous role.

There are no variables a_i^k or r_{ij}^k in (108); the power p_i does not appear in the objective and can be removed from the problem formulation; and there are no spectral mask constraints of the form $\boldsymbol{p}(\mathbf{f}) \in \boldsymbol{\mathcal{P}}(\mathbf{f})$. The first step in the proof of [28, Theorem 7] is to reformulate (108) as

$$\begin{aligned} \mathsf{P} = \max & \sum_i U_i\left(\int_{\mathcal{F}} C_i(\mathbf{f}, \boldsymbol{p}(\mathbf{f}))d\mathbf{f}\right) \\ & \int_{\mathcal{F}} p_i(\mathbf{f})d\mathbf{f} \le p_{\max}. \end{aligned} \tag{109}$$

The proof proceeds to the use of Blackwell's extension to show that the perturbation function, with respect to a perturbation in

the power constraint, is concave. The proof also exploits the linearity of the constraint in (109) and the lack of constraints on individual values of $\boldsymbol{p}(\mathbf{f})$. The substitution in going from (108) to (109) cannot be used for the wireless networking problem in (9)–(12). The linearity of the remaining constraints cannot be exploited and as a consequence Blackwell's extension is not directly applicable here.

REFERENCES

[1] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, 1998.

[2] S. H. Low and D. E. Lapsley, "Optimization flow control, i: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1998.

[3] S. H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Control Syst. Mag.*, vol. 22, no. 1, pp. 28–43, Feb. 2002.

[4] S. H. Low, "A duality model of TCP and queue management algorithms," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 525–536, Aug. 2003.

[5] R. Srikant, *The Mathematics of Internet Congestion Control*, 1st ed. Boston, MA: Birkhauser, 2004.

[6] 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.

[7] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and MAC for stability and fairness in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1514–1524, Aug. 2006.

[8] 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.

[9] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle, "Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2005, pp. 1–13.

[10] Y. Yi and S. Shakkottai, "Hop-by-hop congestion control over a wireless multi-hop network," *IEEE/ACM Trans. Netw.*, vol. 15, no. 133–144, pp. 1548–1559, Feb. 2007.

[11] B. Johansson, P. Soldati, and M. Johansson, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1535–1547, Aug. 2006.

[12] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[13] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[14] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.

[15] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[16] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[17] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[18] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proc. IEEE INFOCOM*, Miami, FL, Mar. 13–17, 2005, vol. 3, pp. 1723–1734.

[19] M. J. Neely, "Energy optimal control for time-varying wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006.

[20] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends in Netw.*, vol. 1, no. 1, pp. 1–144, 2006.

[21] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Opportunistic power scheduling for dynamic multi-server wireless systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1506–1515, Jun. 2006.

[22] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[23] L. Tassiulas, "Adaptive back-pressure congestion control based on local information," *IEEE Trans. Autom. Control*, vol. 40, no. 2, pp. 236–250, Feb. 1995.

[24] L. Tassiulas, "Scheduling and performance limits of networks with constantly changing topology," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 1067–1073, May 1997.

[25] X. Wang and G. B. Giannakis, "An adaptive signal processing approach to scheduling in wireless networks—Part II:: Optimal power allocation with full CSI," *IEEE Trans. Signal Process.*, 2008, to be published.

[26] X. Wang and G. B. Giannakis, "An adaptive signal processing approach to scheduling in wireless networks—Part i: Uniform power allocation with quantized CSI," *IEEE Trans. Signal Process.*, 2008, to be published.

[27] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[28] Z.-Q. Luo and S. Zhang, Dynamic Spectrum Management: Complexity and Duality Dep. ECE, Univ. of Minnesota, Tech. Rep., 2007, submitted for publication.

[29] A. A. Lyapunov, "Sur les fonctions-vecteur complètement additives," *Bull. Acad. Sci. URSS. Sèr. Math.*, vol. 4, pp. 465–478, 1940.

[30] S. Hayashi and Z.-Q. Luo, Spectrum Management for Interference-Limmited Multiuser Communication Systems Dept. of ECE, Univ. of Minnesota, Tech. Rep., 2006, submitted for publication.

[31] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[32] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. New York: Athena Scientific, 1999.

[33] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Berlin, Heilderberg, Germany: Springer-Verlag, 1985.

[34] A. Ribeiro and G. B. Giannakis, "Optimal FDMA over wireless fading mobile ad-hoc networks," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, Mar. 30–Apr. 4 2008, pp. 2765–2768.

[35] N. Gatsis, A. Ribeiro, and G. B. Giannakis, "A class of convergent algorithms for resource allocation in wireless fading networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1808–1823, May 2010.

[36] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking," *IEEE Trans. Signal Process.* Apr. 2010 (to appear) [Online]. Available: http://www.seas.upenn.edu/aribeiro/preprints/eso main.pdf.

Alejandro Ribeiro (M'07) 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 Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

He is currently an Assistant Professor in the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, where he started in 2008. From 2003 to 2008, he was with the Department of Electrical and Computer Engineering, University of Minnesota. From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. His research interests lie in the areas of communication, signal processing, and networking. His current research focuses on wireless communications and networking, distributed signal processing, and wireless sensor networks.

Dr. Ribeiro is a Fulbright scholar and received the best student paper awards at ICASSP 2005 and ICASSP 2006.

Georgios B. Giannakis (F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1981, the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

Since 1999, he has been a professor with the University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the Electrical and Computer Engineering Department, and serves as Director of the Digital Technology Center. His general interests span the areas of communications, networking and statistical signal processing—subjects on which he has published more than 300 journal papers, 500 conference papers, two edited books, and two research monographs. His current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile *ad hoc* networks, wireless sensor, and social networks. He is the (co-)inventor of 16 patents issued.

Dr. Giannakis is the (co-)recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP, has served the IEEE in a number of posts, and is also as a Distinguished Lecturer for the IEEE-SP Society.