Layer Separability of Wireless Networks

Alejandro Ribeiro and Georgios B. Giannakis
Dept. of ECE, University of Minnesota
200 Union Street S.E., Minneapolis, MN 55455

***Abstract*—A general wireless networking problem is formulated to jointly optimize end-to-end user rates, routes, link capacities, transmitted power, frequency and power allocation across subcarriers and fading states. It is proved that wireless networking problems in the presence of fading, while non-convex, have zero Lagrangian duality gap. A consequence of this result is the optimality of separating the problem in (the conventional) layers.**

I. Introduction

The goal of a communication network is to transport information from a generating source to an intended destination. Of the many ways this can be done, there are some that optimize properly chosen criteria. This paper is concerned with structural properties of the associated optimization problem for wireless communication networks in the presence of fading.

Optimization as a mathematical tool to understand networking protocols appears first in the form of the network utility maximization (NUM) framework [1]. The insight provided in this paper is that congestion control algorithms can be understood as distributed implementations of algorithms that solve utility maximization problems. Extending the NUM framework to wireless networking problems is not a simple pursuit. Different from wired networks where links between nodes are determined by physical connections and capacities are fixed, links in wireless networks are not predetermined and their capacities are not fixed. Rather, connections and link capacities are part of the optimization variables. Nevertheless, wireless physical layer models have been incorporated into NUM formulations, and arguably represent one of the most promising research directions on cross-layer design; see e.g., [2].

Most of the NUM literature relies on dual decomposition. This is because the Lagrangian dual problem exhibits a separable structure loosely correlated with the division of networking problems into layers. This has been observed in, e.g., [3] and recently surveyed in [4]. It has to be noted that since wireless networking problems are not convex, the duality gap is not null in general. Thus, the dual optimum is different from the primal optimum and layering comes at the price of performance loss. A related approach to the wireless networking problem is the work on stochastic network optimization; see [5] for a comprehensive treatment. This work builds on the back-pressure algorithm [6], and extends it to deal with random network states as those generated by fading.

Different from NUM, e.g., [2] and stochastic network optimization, e.g., [5] the goal here is not to propose and analyze a specific algorithm but to look at a structural property – the duality gap – of the optimization problem. Different from [4] the aim here is not to argue that layered architectures can be formally understood as decompositions of optimization problems but to prove that layering is optimal in wireless networks.

The paper starts modeling the relation between variables in the wireless networking problem in Section II. The effect of fading in determining link capacities and power consumption is discussed in Section II-A. Using these relations the optimal wireless networking problem is defined in Section III. This problem is non-convex and therefore lies on the unfortunate side of the watershed between easy and difficult optimization problems. The Lagrangian dual of this problem is introduced in Section III-A. The main result of the paper is presented in Section IV. As asserted by Theorem 1, the optimization problem defined in Section III, while non-convex exhibits zero duality gap.

This result has far reaching implications that we have yet not completely explored. These come from the fact that it is easier to work with the dual problem than with the original primal problem. With zero duality gap ensured, working with the dual problem incurs no loss of optimality. In this paper, we devote Section V to show that the separable structure of the Lagrangian implies that it is possible to define separate optimization problems to obtain optimal routes, optimal link capacity allocations, and optimal power/frequency assignments. In this sense Theorem 1 implies the optimality of conventional layering in general wireless networking problems in the presence of fading.

II. Problem formulation

Consider an ad-hoc wireless network composed of J user terminals $\{T_i\}_{i=1}^J$. Terminal T_i wants 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 might be associated with different flows to, e.g., accommodate different types of traffic (video, voice or data). The destination of the flow k is denoted by 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.

We model network connectivity with a graph $\mathcal{G}(v,e)$ with vertices $v := \{1,\dots,J\}$ 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 adjacency of i is denoted $\mathcal{N}(i) := \{j : (i,j) \in \mathcal{E}\}$. Each terminal $\{T_j\}_{j\in\mathcal{N}(i)}$ that can communicate with T_i will be referred to as a neighbor and the set of all neighbors as T_i's neighborhood. Given this model, terminals rely on multi-hop 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 k-th flow packets to T_j. Assuming that packets are not discarded and that queues are stable throughout the network, we can write a flow conservation equation relating rates a_i^k of exogenous packet arrivals (from the application layer) and endogenous (to the network layer) average rates r_{ij}^k of transmission to and from neighboring nodes

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

Consider now the rates r_{ij}^k of all flows traversing the link $T_i \to T_j$. Letting c_{ij} denote the information capacity of this link, queue stability is ensured by requiring

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

The constraints in (1) and (2) are sufficient to describe traffic flow in a wireline network with fixed capacities c_{ij}. In a wireless network however, c_{ij} is not a fixed resource given to terminals. In fact, operating conditions are determined by a set of available frequencies $\mathcal{F}$ and power budgets $p_{\max i}$. Thus, in addition to a_i^k and r_{ij}^k, T_i has to decide how to split a power budget $p_{\max i}$ between tones $f \in \mathcal{F}$ and neighbors T_j, $j \in \mathcal{N}(i)$. Matters are further complicated by fading as described next.

978-1-4244-2247-0/08/$25.00 ©2008 IEEE.

A. Link capacities and transmitted power

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

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

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

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

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

where $\mathrm{E}_{\mathbf{h}}[\cdot]$ denotes expectation over the channel probability distribution.

The rate of information transmission in the $T_i \to T_j$ link is a function of the power distributions $p_{ij}^f(\mathbf{h})$ and the channel realizations $\mathbf{h}$. To keep analysis general define a function $C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)$ to map channels and powers to link capacities so that the capacity $c_{ij}^f(\mathbf{h})$ of the link $T_i \to T_j$ on the tone f is

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

The function $C_{ij}(\cdot)$ is determined by terminal's capabilities and operating conditions. The capacity c_{ij} of the wireless link $T_i \to T_j$ is obtained after integrating over all possible channel realizations:

$$c_{ij} := \mathrm{E}_{\mathbf{h}}\left[\sum_{f \in \mathcal{F}} c_{ij}^f(\mathbf{h})\right] = \mathrm{E}_{\mathbf{h}}\left[\sum_{f \in \mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)\right]. \tag{6}$$

The average link capacity and power expressions in (4) and (6) along with the flow and rate constraints in (1) and (2) describe information flow in a generic wireless network. They are used to define a wireless networking problem in the next section.

III. Optimal wireless networking

Consider concave $U_i^k(a_i^k)$ and convex $V_i(p_i)$ functions, respectively, representing the value of rate a_i^k and the cost of power p_i. We can then define the optimal networking problem as [cf. (1), (2), (4) and (6)]

$$P = \max \sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \tag{7}$$

$$c_{ij} \le \mathrm{E}_{\mathbf{h}}\left[\sum_{f \in \mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)\right] \tag{8}$$

$$p_i \ge \mathrm{E}_{\mathbf{h}}\left[\sum_{j \in \mathcal{N}(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\mathbf{h})\right] \tag{9}$$

$$a_i^k \le \sum_{j \in \mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k\right), \qquad \sum_k r_{ij}^k \le c_{ij} \tag{10}$$

where the constraints (1), (4) and (6) have been relaxed, which we can do without loss of optimality. Note that all problem variables have to be non-negative, but this is left implicit in (7). We have also left implicit power constraints $p_i \le p_{\max i}$ and $p_{ij}^f(\mathbf{h}) \le p_{\max}$, arrival rate requirements $a_{\min i}^k \le a_i^k \le a_{\max i}^k$ and upper bound constraints $c_{ij} \le c_{\max}$ and $r_{ij}^k \le r_{\max}$ on link capacities and link flow rates. For future reference, define the vector valued power distribution $\mathbf{p}(\mathbf{h})$ with components $p_{ij}^f(\mathbf{h})$ and $\mathbf{X}$ the set of primal variables c_{ij}, p_i, r_{ij}^k and a_i^k for all possible subindexes – i.e., all i and all $j \in \mathcal{N}(i)$ for c_{ij}, all i for p_i and so on. The box constraints define a set of feasible variables as follows

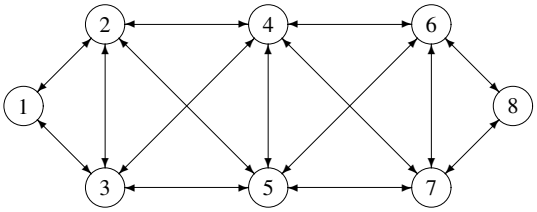

Fig. 1. Connectivity graph of an example wireless network.

$$B := \big\{(\mathbf{X}, \mathbf{p}(\mathbf{h})) :\ 0 \le p_{ij}^f(\mathbf{h}) \le p_{\max},\ 0 \le p_i \le p_{\max i},\ a_{\min i}^k \le a_i^k \le a_{\max i}^k,\ 0 \le c_{ij} \le c_{\max},\ 0 \le r_{ij}^k \le r_{\max}\big\}. \tag{11}$$

The constraints in (11) will be implicit in general and made explicit when demanded by clarity.

The functions $C_{ij}(\cdot)$ are not concave in general, therefore, problem (7) is a difficult optimization problem. This difficulty notwithstanding, there are fundamental properties of wireless networking problems to be derived from properties of (7). This requires the introduction of the Lagrangian dual problem. This is done in the next section.

A. Lagrangian dual problem

To define the dual problem, associate Lagrange multipliers λ_{ij} with the capacity constraints in (8), μ_i with the power in (9), and ν_i^k and ξ_{ij} with the flow and rate constraints in (10). To simplify notation call $\mathbf{\Lambda}$ the set of all dual variables –i.e., λ_{ij}, μ_i, ν_i^k, ξ_{ij} – and write the Lagrangian as

$$\begin{aligned}\mathcal{L}\left[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}\right] = &\sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \\ &+ \sum_{i,j} \lambda_{ij}\left[\mathrm{E}_{\mathbf{h}}\left[\sum_{f \in \mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)\right] - c_{ij}\right] \\ &+ \sum_i \mu_i\left[p_i - \mathrm{E}_{\mathbf{h}}\left[\sum_{j \in \mathcal{N}(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\mathbf{h})\right]\right] \\ &+ \sum_{i,k} \nu_i^k\left[\sum_{j \in \mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k\right) - a_i^k\right] \\ &+ \sum_{ij} \xi_{ij}\left[c_{ij} - \sum_k r_{ij}^k\right].\end{aligned} \tag{12}$$

The dual function is obtained as

$$g[\mathbf{\Lambda}] = \max_{(\mathbf{X}, \mathbf{p}(\mathbf{h})) \in B} \mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}] \tag{13}$$

where the maximization is performed over the set of primal variables satisfying the box constraints in (11). We finally define the dual problem as

$$D = \min_{\mathbf{\Lambda} \ge 0} g[\mathbf{\Lambda}]. \tag{14}$$

Since (7) is non-convex, the duality gap is, in principle, nonzero and we have $D \ge P$. Solving (14) is thus a relaxation in the sense that it yields an upper bound D on the maximum achievable utility. The usefulness of (14) depends therefore in the proximity of D and P. This is explored in the next section.

IV. Optimality of dual relaxation

The challenges in solving (7) are now clear. For deterministic channels, the problem is gnerally NP-hard. The Lagrangian dual problem discussed in Section III-A is certainly useful in establishing upper bounds on the achievable utility, but might or might not be close to the actual utility. One would expect that introducing fading should complicate matters further. It is thus a remarkable fact that in the presence of fading the duality gap vanishes, i.e., $P = D$. We state this result in the following theorem.

Theorem 1 *Let P denote the solution of the primal problem* (7) *and D that of its dual in* (14)*. If the channel cumulative distribution function (cdf) is continuous, then*

$$P = D. \tag{15}$$

The proof relies on Lyapunov's convexity theorem [7] first used in the context of spectrum management for deterministic DSL channels in [8]. Note that the link capacity function $C(\cdot)$ is not necessarily concave in Theorem 1. Therefore the optimization problem might be non-convex. The duality gap, however, is null. Continuity of the channel cdf ensures that no channel realization has positive probability. This is satisfied by practical channel models including Rayleigh, Rice and Nakagami fading.

Before proceeding to the proof of Theorem 1 we state Lyapunov's convexity theorem and introduce pertinent definitions. Start with the concept of non-atomic measure as defined next.

Definition 1 Non-atomic measure. *Let w be a measure defined on the Borel field $\mathcal{B}$ of subsets of a space $\mathcal{X}$. We say w is non-atomic if for any measurable set $E_0 \in \mathcal{B}$ with $w(E_0) > 0$, there exist a subset E of E_0; i.e., $E \subset E_0$, such that $w(E_0) > w(E) > 0$.*

Lyapunov's convexity theorem that we introduce next refers to the range of values taken by (vector) non-atomic measures.

Theorem 2 Lyapunov's convexity theorem. *Let $w_1, \ldots, w_n$ be non-atomic measures on the Borel field $\mathcal{B}$ of subsets of a space X and define the vector measure $\mathbf{w}(E) := [w_1(E), \ldots, w_n(E)]^T$. The range of the vector measure $\mathcal{W} : \{\mathbf{w}(E) : E \in \mathcal{B}\}$ is convex. I.e. if $\mathbf{w}(E_1) = \mathbf{w}_1$ and $\mathbf{w}(E_2) = \mathbf{w}_2$, then for any $t \in [0, 1]$ there exist $E_0 \in \mathcal{B}$ such that $\mathbf{w}(E_0) = t\mathbf{w}_1 + (1-t)\mathbf{w}(E_2) = \mathbf{w}_2$.*

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 (7) obtained by perturbing, the constraints used to define the Lagrangian in (12). The perturbation function $P(\mathbf{\Delta})$ assigns to each (perturbation) parameter set $\mathbf{\Delta} := (\Delta c_{ij}, \Delta p_i, \Delta a_i^k, \Delta r_{ij})$ the solution of the (perturbed) optimization problem

$$P(\mathbf{\Delta}) = \max \sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \tag{16}$$

$$c_{ij} - \mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)\right] \leq \Delta c_{ij} \tag{17}$$

$$\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} p_{ij}^f(\mathbf{h})\right] - p_i \leq \Delta p_i \tag{18}$$

$$a_i^k - \sum_{j\in\mathcal{N}(i)}\Big(r_{ij}^k - r_{ji}^k\Big) \leq \Delta a_i^k$$

$$\sum_k r_{ij}^k - c_{ij} \leq \Delta r_{ij} \tag{19}$$

where the box constraints are implicit as in (7). The perturbed problem (16) can be interpreted as a modified version of (7), where we allow the constraints to be violated by $\mathbf{\Delta}$ amounts. To prove that the duality gap is zero, it suffices to show that $P(\mathbf{\Delta})$ is a concave function of $\mathbf{\Delta}$.

Let $\underline{\mathbf{\Delta}} := (\Delta\underline{c}_{ij}, \Delta\underline{p}_i, \Delta\underline{a}_i^k, \Delta\underline{r}_{ij})$ and $\overline{\mathbf{\Delta}} := (\Delta\overline{c}_{ij}, \Delta\overline{p}_i, \Delta\overline{a}_i^k, \Delta\overline{r}_{ij})$ denote arbitrary sets of perturbations with respective optimal values $\underline{P} := P(\underline{\mathbf{\Delta}})$ and $\overline{P} := P(\overline{\mathbf{\Delta}})$. Further, let $\underline{\mathbf{X}}$ denote (likewise, $\underline{p}_{ij}^f(\mathbf{h})$, $\underline{c}_{ij}$, $\underline{p}_i$, $\underline{r}_{ij}^k$, $\underline{a}_i^k$) the variables achieving the optimum under perturbation $\underline{\mathbf{\Delta}}$; and $\overline{\mathbf{X}}$ (likewise, $\overline{p}_{ij}^f(\mathbf{h})$, $\overline{c}_{ij}$, $\overline{p}_i$, $\overline{r}_{ij}^k$, $\overline{a}_i^k$) the ones under $\overline{\mathbf{\Delta}}$. For arbitrary $\alpha \in [0, 1]$, we are interested in the solution of (16) under perturbation $\overline{\underline{\mathbf{\Delta}}} := \alpha\underline{\mathbf{\Delta}} + (1-\alpha)\overline{\mathbf{\Delta}}$. In particular, we wish to show that $P(\overline{\underline{\mathbf{\Delta}}})$ is concave; i.e.,

$$\begin{aligned} P\left(\overline{\underline{\mathbf{\Delta}}}\right) &= P\left[\alpha\underline{\mathbf{\Delta}} + (1-\alpha)\overline{\mathbf{\Delta}}\right] \\ &\geq \alpha P\left(\underline{\mathbf{\Delta}}\right) + (1-\alpha)P\left(\overline{\mathbf{\Delta}}\right). \end{aligned} \tag{20}$$

Key in establishing (20) is the following lemma.

Lemma 1 *Consider perturbations $\underline{\mathbf{\Delta}}$ and $\overline{\mathbf{\Delta}}$ and let $\underline{\mathbf{X}}$ and $\overline{\mathbf{X}}$ be the corresponding optimal arguments that solve* (16)*. Under perturbation $\overline{\underline{\mathbf{\Delta}}} := \alpha\underline{\mathbf{\Delta}} + (1-\alpha)\overline{\mathbf{\Delta}}$ with $\alpha \in [0, 1]$, the average link capacity and power variables*

$$\begin{aligned} \overline{\underline{c}}_{ij} &:= \alpha\underline{c}_{ij} + (1-\alpha)\overline{c}_{ij} \\ \overline{\underline{p}}_i &:= \alpha\underline{p}_i + (1-\alpha)\overline{p}_i \end{aligned} \tag{21}$$

are feasible for the problem (16)*.*

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 if $\overline{\underline{c}}_{ij}$ is feasible for (16) under $\overline{\underline{\mathbf{\Delta}}}$, then $\overline{\underline{r}}_{ij}^k := \alpha\underline{r}_{ij}^k + (1-\alpha)\overline{r}_{ij}^k$ will also be. Indeed, since $(\underline{r}_{ij}^k, \underline{c}_{ij})$ and $(\overline{r}_{ij}^k, \overline{c}_{ij})$ are feasible for the problem (16) under respective perturbations $\underline{\mathbf{\Delta}}$ and $\overline{\mathbf{\Delta}}$, we have [cf. (19)]

$$\sum_k \underline{r}_{ij}^k - \underline{c}_{ij} \leq \Delta\underline{r}_{ij} \tag{22}$$

$$\sum_k \overline{r}_{ij}^k - \overline{c}_{ij} \leq \Delta\overline{r}_{ij}\ . \tag{23}$$

Multiplying (22) by α, (23) by $(1-\alpha)$, summing up and rearranging terms, yields

$$\begin{aligned} \sum_k \Big[\alpha\underline{r}_{ij}^k + (1-\alpha)\overline{r}_{ij}^k\Big] - \Big[\alpha\underline{c}_{ij} - (1-\alpha)\overline{c}_{ij}\Big] \\ \leq \Big[\Delta\underline{r}_{ij} + (1-\alpha)\Delta\overline{r}_{ij}\Big]. \end{aligned} \tag{24}$$

Using the definitions $\overline{\underline{r}}_{ij}^k := \alpha\underline{r}_{ij}^k + (1-\alpha)\overline{r}_{ij}^k$, $\overline{\underline{c}}_{ij} := \alpha\underline{c}_{ij} + (1-\alpha)\overline{c}_{ij}$ and $\Delta\overline{\underline{r}}_{ij} := \alpha\Delta\underline{r}_{ij} + (1-\alpha)\Delta\overline{r}_{ij}$, the inequality in (24) is equivalent to

$$\sum_k \overline{\underline{r}}_{ij}^k - \overline{\underline{c}}_{ij} \leq \Delta\overline{\underline{r}}_{ij}. \tag{25}$$

Lemma 1 ensures that $\overline{\underline{c}}_{ij}$ is feasible for the problem (16) under perturbation $\overline{\underline{\mathbf{\Delta}}}$. Hence, (25) shows that $\overline{\underline{r}}_{ij}^k$ also is. Likewise, we can prove that since $\overline{\underline{r}}_{ij}^k$ is feasible for (16) under $\overline{\underline{\mathbf{\Delta}}}$, the same should be true for $\overline{\underline{a}}_i^k := \alpha\underline{a}_i^k + (1-\alpha)\overline{a}_i^k$ (simply repeat steps (22)-(25) for the flow conservation constraint in (19)).

In summary, $\overline{\underline{a}}_i^k := \alpha\underline{a}_i^k + (1-\alpha)\overline{a}_i^k$ and $\overline{\underline{p}}_i := \alpha\underline{p}_i + (1-\alpha)\overline{p}_i$ are feasible points of (16) under perturbation $\overline{\underline{\mathbf{\Delta}}}$. Since $P(\overline{\underline{\mathbf{\Delta}}})$ is the maximum of (16) among all feasible (a_i^k, p_i), it must satisfy

$$\begin{aligned} P(\overline{\underline{\mathbf{\Delta}}}) &\geq \sum_{i,k} U_i^k\big(\overline{\underline{a}}_i^k\big) - \sum_i V_i\big(\overline{\underline{p}}_i\big) \\ &= \sum_{i,k} U_i^k\big[\alpha\underline{a}_i^k + (1-\alpha)\overline{a}_i^k\big] - \sum_i V_i\big[\alpha\underline{p}_i + (1-\alpha)\overline{p}_i\big] \end{aligned} \tag{26}$$

where in the second equality we used the definitions of $\underline{a}_i^k$ and $\underline{p}_i$. Finally, since by assumption the $U_i^k(\cdot)$ functions are concave and the $V_i(\cdot)$ functions are convex, it follows that

$$\begin{aligned} P(\overline{\underline{\mathbf{\Delta}}}) &\geq \sum_{i,k}\Big[\alpha U_i^k\big(\underline{a}_i^k\big)+(1-\alpha)U_i^k\big(\overline{a}_i^k\big)\Big] \\ &\qquad -\sum_i\Big[\alpha V_i\big(\underline{p}_i\big)+(1-\alpha)V_i\big(\overline{p}_i\big)\Big] \\ &= \alpha\left[\sum_{i,k}U_i^k\big(\underline{a}_i^k\big)-\sum_i V_i\big(\underline{p}_i\big)\right] \\ &\qquad +(1-\alpha)\left[\sum_{i,k}U_i^k\big(\overline{a}_i^k\big)-\sum_i V_i\big(\overline{p}_i\big)\right] \\ &= \alpha P(\underline{\mathbf{\Delta}})+(1-\alpha)P(\overline{\underline{\mathbf{\Delta}}}). \end{aligned} \tag{27}$$

where the second equality holds true because $(\underline{a}_i^k,\underline{p}_i)$ and $(\overline{a}_i^k,\overline{p}_i)$ are optimal arguments of (16) under perturbations $\underline{\mathbf{\Delta}}$ and $\overline{\underline{\mathbf{\Delta}}}$, respectively. Comparing (20) with (27), we deduce that the perturbation function $P(\mathbf{\Delta})$ is concave. Therefore, the duality gap is null, i.e., $D=P$. ■

We proceed now to the proof of Lemma 1.

Proof: (of Lemma 1). To prove that $\overline{\underline{c}}_{ij}$ and $\overline{\underline{p}}_i$ as defined in (21) are feasible for (16) under perturbation $\overline{\underline{\mathbf{\Delta}}}$, we will show that there exist feasible power distributions $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ for which the following holds true

$$\overline{\underline{c}}_{ij}-\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\overline{\underline{\mathbf{p}}}^f(\mathbf{h})\Big)\right]\leq\Delta\overline{\underline{c}}_{ij} \tag{28}$$

$$\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\overline{\underline{p}}_{ij}^f(\mathbf{h})\right]-p_i\leq\Delta\overline{\underline{p}}_i\ . \tag{29}$$

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

$$\begin{aligned} w_{ij}(E) &:= \left[\mathbb{E}_{\mathbf{h}\in E}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\underline{\mathbf{p}}^f(\mathbf{h})\Big)\right],\right. \\ &\qquad \left.\mathbb{E}_{\mathbf{h}\in E}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\overline{\mathbf{p}}^f(\mathbf{h})\Big)\right]\right]^T \end{aligned} \tag{30}$$

$$\begin{aligned} v_i(E) &:= \left[\mathbb{E}_{\mathbf{h}\in E}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\mathbf{h})\right],\right. \\ &\qquad \left.\mathbb{E}_{\mathbf{h}\in E}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\overline{p}_{ij}^f(\mathbf{h})\right]\right]^T \end{aligned} \tag{31}$$

where the expectation is taken over the set of channel realizations E. A vector of channel realizations $\mathbf{h}\in\mathcal{H}$ is a point in the space $\mathcal{H}$. The set E is a collection of vectors $\mathbf{h}$. Each of these sets is assigned vector measures $w_{ij}(E)$ and $v_i(E)$ defined in terms of the power distributions $\underline{p}_{ij}^f(\mathbf{h})$ and $\overline{\underline{p}}_{ij}^f(\mathbf{h})$. The entries of $w_{ij}(E)$ represent the contribution of realizations $\mathbf{h}\in\ E$ to the capacity of the link $T_i\to T_j$. The first entry measures such a contribution when the distribution is $\underline{p}_{ij}^f(\mathbf{h})$, 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 $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ associated with perturbation $\overline{\underline{\mathbf{\Delta}}}$. Correspondingly, the entries of $v_i(E)$ denote the power consumed by T_i for channel realization $\mathbf{h}\in E$ expressed in terms of $\underline{p}_{ij}^f(\mathbf{h})$ and $\overline{\underline{p}}_{ij}^f(\mathbf{h})$.

The measures $w_{ij}(E)$ and $v_i(E)$ are non-atomic. This follows from the fact that the channel cdf is continuous and the power distributions $p_{ij}^f(\mathbf{h})$ are upper bounded by $p_{\max}$, implying that there are no channel realizations with positive measure, i.e., $w_{ij}(\mathbf{h})=0$ and $v_i(\mathbf{h})=0$ for all $\mathbf{h}\in\mathcal{H}$.

Two particular sets that are important for this proof are the empty set, $E=\emptyset$, and the entire space $E=\mathcal{H}$. For $E=\mathcal{H}$, the expected value operators $\mathbb{E}_{\mathbf{h}\in E}(\cdot)=\mathbb{E}_{\mathbf{h}\in\mathcal{H}}(\cdot)=\mathbb{E}_{\mathbf{h}}(\cdot)$ in (30) and (31) coincide with the expected value operators in (17) and (18). We write this explicitly as

$$\begin{aligned} w_{ij}(\mathcal{H}) &:= \left[\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\underline{\mathbf{p}}^f(\mathbf{h})\Big)\right],\right. \\ &\qquad \left.\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\overline{\mathbf{p}}^f(\mathbf{h})\Big)\right]\right]^T \end{aligned} \tag{32}$$

$$\begin{aligned} v_i(\mathcal{H}) &:= \left[\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\mathbf{h})\right],\right. \\ &\qquad \left.\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\overline{p}_{ij}^f(\mathbf{h})\right]\right]^T. \end{aligned} \tag{33}$$

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 $\mathbf{w}(E):=\big[\{w_{ij}(E)\}_{i\in[1,J],j\in\mathcal{N}(i)},\{v_i(\emptyset)\}_{i\in[1,J]}\big]$. Since the scalar measures are non-atomic, $\mathbf{w}(E)$ is also non-atomic. Theorem 2 then implies that the range of $\mathbf{w}$ is convex; hence,

$$\mathbf{w}_0:=\alpha\mathbf{w}(\mathcal{H})+(1-\alpha)\mathbf{w}(\emptyset)=\alpha\mathbf{w}(\mathcal{H}) \tag{34}$$

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

$$\mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\underline{\mathbf{p}}^f(\mathbf{h})\Big)\right]=\alpha\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\underline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \tag{35}$$

$$\mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\mathbf{h})\right]\ =\alpha\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\underline{p}_{ij}^f(\mathbf{h})\right] \tag{36}$$

and this holds true for all (i,j). The same relation holds for the second entries of w_{ij} and v_i, i.e., (35) and (36) are valid if $\underline{a}_{ij}^f(\mathbf{h})$ and $\underline{p}_{ij}^f(\mathbf{h})$ are replaced by $\overline{a}_{ij}^f(\mathbf{h})$ and $\overline{p}_{ij}^f(\mathbf{h})$, but this is not important for the proof.

Consider now the complement set E_0^c defined as the set for which $E_0\cup E_0^c=\mathcal{H}$ and $E_0\cap E_0^c=\emptyset$. Given this definition and the additivity property of measures, we arrive at $\mathbf{w}(E_0)+\mathbf{w}(E_0^c)=\mathbf{w}(\mathcal{H})$. Combining the latter with (34), yields

$$\mathbf{w}(E_0^c)=\mathbf{w}(\mathcal{H})-\mathbf{w}(E_0)=(1-\alpha)\mathbf{w}(\mathcal{H}). \tag{37}$$

Mimicking the reasoning leading from (34) to (35)-(36), we further deduce that (37) implies that for all (i,j), $w_{ij}(E_0^c)=(1-\alpha)w_{ij}(\mathcal{H})$ and $v_i(E_0^c)=(1-\alpha)w_{ij}(\mathcal{H})$. Restricting focus to the *second* entries of w_{ij} and v_i, it follows that

$$\begin{aligned} &\mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\overline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \\ &\qquad =(1-\alpha)\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f,\overline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \end{aligned} \tag{38}$$

$$\mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\overline{p}_{ij}^f(\mathbf{h})\right]=(1-\alpha)\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}\overline{p}_{ij}^f(\mathbf{h})\right] \tag{39}$$

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

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

$$\overline{\underline{p}}_{ij}^f(\mathbf{h}) = \begin{cases} \underline{p}_{ij}^f(\mathbf{h}) & \mathbf{h} \in E_0 \\ \overline{p}_{ij}^f(\mathbf{h}) & \mathbf{h} \in E_0^c \end{cases} . \tag{40}$$

The distribution $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ satisfies the box constraints $0 \leq \overline{\underline{p}}_{ij}^f(\mathbf{h}) \leq p_{\max}$ because $\underline{p}_{ij}^f(\mathbf{h})$ and $\overline{p}_{ij}^f(\mathbf{h})$ are feasible in their respective problems and thus satisfy $0 \leq \underline{p}_{ij}^f(\mathbf{h}) \leq p_{\max}$ and $0 \leq \overline{p}_{ij}^f(\mathbf{h}) \leq p_{\max}$. Hence, $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ is a feasible power distribution for (16).

Using (35) and (38), the average link capacities can be expressed in terms of $\underline{p}_{ij}^f(\mathbf{h})$, $\overline{p}_{ij}^f(\mathbf{h})$ and $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ as

$$\begin{aligned} &\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\underline{\mathbf{p}}}^f(\mathbf{h})\Big)\right] \\ &= \mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\underline{\mathbf{p}}}^f(\mathbf{h})\Big)\right] \\ &\qquad + \mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\underline{\mathbf{p}}}^f(\mathbf{h})\Big)\right] \\ &= \mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \underline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \\ &\qquad + \mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \\ &= \alpha\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \underline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \\ &\qquad + (1-\alpha)\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\mathbf{p}}^f(\mathbf{h})\Big)\right]. \end{aligned} \tag{41}$$

The first equality in (41) holds since the space $\mathcal{H}$ is divided into E_0 and its complement E_0^c. The second equality is true because when restricted to E_0, $\overline{\underline{p}}_{ij}^f(\mathbf{h}) = \underline{p}_{ij}^f(\mathbf{h})$; and when restricted to E_0^c, $\overline{\underline{p}}_{ij}^f(\mathbf{h}) = \overline{p}_{ij}^f(\mathbf{h})$. The third equality follows from (35) and (38).

An analogous relation also holds for the average power consumptions. Indeed, separating $\mathcal{H}$ into E_0 and E_0^c and using the definition of $\overline{\underline{p}}_{ij}^f(\mathbf{h})$ in (40) along with (36) and (39), yields

$$\begin{aligned} &\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{\underline{p}}_{ij}^f(\mathbf{h})\right] \\ &= \mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{\underline{p}}_{ij}^f(\mathbf{h})\right] + \mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{\underline{p}}_{ij}^f(\mathbf{h})\right] \\ &= \mathbb{E}_{\mathbf{h}\in E_0}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\mathbf{h})\right] + \mathbb{E}_{\mathbf{h}\in E_0^c}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{p}_{ij}^f(\mathbf{h})\right] \\ &= \alpha\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\mathbf{h})\right] + (1-\alpha)\mathbb{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{p}_{ij}^f(\mathbf{h})\right] \end{aligned} \tag{42}$$

Based on (41) and (42) we are ready to complete the proof. Considering the link capacity constraints (17) under perturbations $\underline{\boldsymbol{\Delta}}$ and $\overline{\boldsymbol{\Delta}}$, we have that

$$\underline{c}_{ij} - \mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \underline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \leq \Delta\underline{c}_{ij} \tag{43}$$

$$\overline{c}_{ij} - \mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \leq \Delta\overline{c}_{ij} . \tag{44}$$

As with the link transmission rate constraints in (22) and (23), multiply (43) by α and (44) by $(1-\alpha)$. Summing up, rearranging terms and using the definitions $\overline{\underline{c}}_{ij} := \alpha\underline{c}_{ij} + (1-\alpha)\overline{c}_{ij}$ and $\Delta\overline{\underline{c}}_{ij} := \alpha\Delta\underline{c}_{ij} + (1-\alpha)\Delta\overline{c}_{ij}$, we obtain

$$\begin{aligned} &\overline{\underline{c}}_{ij} - \alpha\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \underline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \\ &\qquad - (1-\alpha)\mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\mathbf{p}}^f(\mathbf{h})\Big)\right] \leq \Delta\overline{\underline{c}}_{ij} . \end{aligned} \tag{45}$$

But the terms in (45) involving expectations are precisely the terms in the right hand side of (41). They can be therefore replaced by the expected link capacity for power distribution $\overline{\underline{p}}_{ij}^f(\mathbf{h})$, to obtain

$$\overline{\underline{c}}_{ij} - \mathbb{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \overline{\underline{\mathbf{p}}}^f(\mathbf{h})\Big)\right] \leq \Delta\overline{\underline{c}}_{ij} \tag{46}$$

which coincides with (28).

Likewise, considering the power constraints in (18) under $\underline{\boldsymbol{\Delta}}$ and $\overline{\boldsymbol{\Delta}}$ and mimicking the steps in (43)-(45) yields

$$\begin{aligned} &\alpha\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \underline{p}_{ij}^f(\mathbf{h})\right] \\ &\qquad + (1-\alpha)\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} \overline{p}_{ij}^f(\mathbf{h})\right] - \overline{\underline{p}}_i \leq \Delta\overline{\underline{p}}_{ij}. \end{aligned} \tag{47}$$

Since the summation of expectations in (47) coincides with the right hand side of (42), it can be replaced by the left hand side of (42) to obtain (29). From (46) – equivalently (28) – and (29) it follows that $\overline{\underline{c}}_{ij} := \alpha\underline{c}_{ij} + (1-\alpha)\overline{c}_{ij}$ and $\overline{\underline{p}}_i := \alpha\underline{p}_i + (1-\alpha)\overline{p}_i$ are feasible for the problem (16) with perturbation $\overline{\underline{\boldsymbol{\Delta}}}$, thus establishing Lemma 1. ■

Given that the duality gap of the generic optimal wireless networking problem (7) is zero, the dual problem (14) can be solved instead. Because the dual function $g(\boldsymbol{\Lambda})$ in (13) 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 (7) is a variational problem requiring determination of the *function* $\mathbf{p}(\mathbf{h})$ that maps channel realizations to transmitted powers. In this 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 mean that it is easy to find the minimum of $g(\boldsymbol{\Lambda})$ in (13). Evaluating $g(\boldsymbol{\Lambda})$ as per (13) requires maximizing the Lagrangian $\mathcal{L}(\mathbf{x}, \mathbf{p}(\mathbf{h}), \boldsymbol{\Lambda})$ in (12). This maximization *might* be difficult to perform depending on the link capacity function $C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)$. Nevertheless, Theorem 1 reveals important properties of wireless networks summarized in the next section.

V. Optimality of layering

An 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 (12) and assuming that the optimal dual argument $\boldsymbol{\Lambda}^*$ is available, we can write

$$\begin{aligned} \mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \boldsymbol{\Lambda}^*] &= \sum_{i,k}\Big(U_i^k(a_i^k) - \nu_j^{k*}a_j^k\Big) + \sum_i\Big(\mu_i^* p_i - V_i(p_i)\Big) \\ &\quad + \sum_{i,j}\Big(\xi_{ij}^* - \lambda_{ij}^*\Big)c_{ij} + \sum_{i,j,k}\Big(\nu_i^{k*} - \nu_j^{k*} - \xi_{ij}^*\Big)r_{ij}^k \\ &\quad + \mathrm{E}_{\mathbf{h}}\left[\sum_{i,j,f}\lambda_{ij}^* C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i^* p_{ij}^f(\mathbf{h})\right]. \end{aligned} \tag{48}$$

The zero duality gap implies that if $\mathbf{\Lambda}^*$ is known we can, instead of solving (7), solve the (separable) problem

$$P = D = d[\mathbf{\Lambda}^*] = \max_{\mathbf{X},\mathbf{p}(\mathbf{h})} \mathcal{L}[\mathbf{X},,\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}^*] \tag{49}$$

where the maximization is constrained to the $\mathbf{X}$ and $\mathbf{p}(\mathbf{h})$ that satisfy the box constraints.

Because primal variables are decoupled in the Lagrangian $\mathcal{L}[\mathbf{X},\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}^*]$ [cf. (48)], the maximization required in (49) can be split into smaller maximization problems involving less variables. This separability can be used to prove the next theorem.

Theorem 3 Layer separability. *Let λ_{ij}^*, μ_i^*, ${\nu_i^k}^*$, and ξ_{ij}^* denote the optimal dual variables that solve* (14)*. Consider the sub-problems*

$$P(a_i^k) = \max_{a^k_{\min\, i}\le a_i^k \le a^k_{\max\, i}} \left[U_i^k(a_i^k) - {\nu_i^k}^* a_i^k\right] \tag{50}$$

$$P(r_{ij}^k) = \max_{0\le r_{ij}^k \le r_{\max}} \left[\left({\nu_i^k}^* - {\nu_j^k}^* - \xi_{ij}^*\right) r_{ij}^k\right] \tag{51}$$

$$P(c_{ij}) = \max_{0\le c_{ij}\le c_{\max}} \left[\left(\xi_{ij}^* - \lambda_{ij}^*\right) c_{ij}\right] \tag{52}$$

$$P(p_i) = \max_{0\le p_i \le p_{\max\, i}} \left[\mu_i^* p_i - V_i(p_i)\right]. \tag{53}$$

Define further the optimal power allocation problem

$$P[\mathbf{p}(\mathbf{h})] = \max_{0\le p_{ij}^f(\mathbf{h})\le p_{\max}} \mathrm{E}_{\mathbf{h}}\left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right) - \mu_i^* p_{ij}^f(\mathbf{h})\right] \tag{54}$$

Then, the optimal utility yield P in (7) *is given by*

$$P = \sum_{i,k} P(a_i^k) + \sum_{i,j,k} P(r_{ij}^k) + \sum_{i,j} P(c_{ij}) + \sum_i P(p_i) + P[\mathbf{p}(\mathbf{h})] \tag{55}$$

i.e., the primal problem (7) *can be separated into the (sub-) problems* (50)-(54) *without loss of optimality.*

Proof: Simply note that the maximization in (49) 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 individual terms' maximums:

$$\begin{aligned}\max\mathcal{L}[\mathbf{X},\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}^*] = \\ \sum_{i,k} \max\left[U_i^k(a_i^k) - {\nu_j^k}^* a_j^k\right] + \sum_i \max\left[\mu_i^* p_i - V_i(p_i)\right] \\ + \sum_{i,j} \max\left[\left(\xi_{ij}^* - \lambda_{ij}^*\right) c_{ij}\right] + \sum_{i,j,k} \max\left[\left({\nu_i^k}^* - {\nu_j^k}^* - \xi_{ij}^*\right) r_{ij}^k\right] \\ + \max \mathrm{E}_{\mathbf{h}}\left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right) - \mu_i^* p_{ij}^f(\mathbf{h})\right].\end{aligned} \tag{56}$$

To complete the proof recall that $P = D = \max \mathcal{L}[\mathbf{X},\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}^*]$ and substitute definitions (50)-(54) into (56). ■

The rate problem in (50) dictates the amount of traffic allowed into the network. It therefore solves the flow control problem at the transport layer; see Fig. 2. Likewise, (51) represents the network layer routing problem, (52) determines link-level capacities at the data link layer and (53) is the (link layer) average power control problem. Eq. (54) represents resource allocation at the physical layer. Therefore, it is a consequence of Theorem 3 that layering, in the sense of problem separability as per (55) is optimal in faded wireless networks.

We remark that Theorem 3 assumes availability of the optimal Lagrange multipliers λ_{ij}^*, μ_i^*, ${\nu_i^k}^*$, and ξ_{ij}^*. Finding them, while possible, is a non-trivial problem that we will address in forthcoming contributions. However, it has to be appreciated that Theorem 3 establishes a fundamental property of wireless networks in the presence of fading: the decomposition of the problem into the traditional networking layers can be optimal.

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 $j \in n(i)$ 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 $j \in n(i)$
$p_i^* \leftarrow \max_{0\le p_i\le p_{\max\, i}} \left[\mu_i^* p_i - V_i(p_i)\right]$	p_i^*
$\mathbf{p}^*(\mathbf{h}) \leftarrow \max_{0\le p_{ij}^f(\mathbf{h})\le p_{\max}} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right) - \mu_i^* p_{ij}^f(\mathbf{h})\right]$	

Fig. 2. 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 ${a_i^k}^*$, link rates ${r_{ij}^k}^*$, link capacities c_{ij}^*, and average transmitted power p_i^*. The physical layer problem can be further separated in per-fading-state subproblems. It cannot, alas, be separated in per-terminal problems for general link capacity functions $C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right)$. Thus, the challenge in wireless networking is not as much in cross-layer optimization as in cross-terminal optimization of the physical layer.

VI. Concluding remark

This paper has shown that optimal wireless networking problems in the presence of fading, while non-convex, have zero Lagrangian duality gap. This result implies that separation in (the conventional) layers is optimal.[1].

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] 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, August 2006.

[3] 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, July 2004.

[4] 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, January 2007.

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

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

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

[8] Z.-Q Luo and S. Zhang, "Dynamic spectrum management: complexity and duality," *Tech. report, Dept. of ECE, Univ. of Minnesota (submitted for publication)*, 2007.

[1] Work in this paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.