Adaptive Cross-Layer Resource Allocation for Wireless Orthogonal-Access Networks

(Invited Paper)

Antonio G. Marques* and Georgios B. Giannakis†
*Dept. of Signal Theory and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28943 Madrid, SPAIN
Email: antonio.garcia.marques@urjc.es
†Dept. of Electrical and Computer Engineering. University of Minnesota, Minneapolis, 55455, MN, USA
Email: georgios@umn.edu

***Abstract*—Adaptive cross-layer designs exploit channel state information (CSI) to optimize wireless networks operating over fading channels. Capitalizing on convex optimization, duality theory and stochastic approximation tools, this paper develops channel-adaptive algorithms to allocate resources at the transport, network, link, and physical layers. Optimality here refers to maximizing a sum-utility of the average end-to-end rates, while at the same time minimizing a sum-cost of the average transmit power. Focus is placed on interference-limited access with nodes transmitting orthogonally over a set of parallel channels. The novel optimal resource allocation schemes depend on two variables: the optimum Lagrange multipliers and the available CSI. Two strategies to find the optimum value of the multipliers are investigated. The first one relies on dual gradient iterations and requires knowledge of the channel distribution. The second one relies on stochastic approximation tools, acquires the channel distribution on-the-fly, and exhibits tracking capabilities. Convergence is asserted for both strategies. Interestingly, it is shown analytically that when layers share the proper information, designs implementing a layered strategy, where each layer uses the available CSI to adapt resources separately, can be rendered optimal.**

***Index Terms*—Resource management, cross-layer designs, stochastic approximation.**

I. INTRODUCTION

Non-linear (convex) optimization tools have been successfully adopted to analyze and design cross-layer algorithms for wireless networks; see e.g., [3] and references therein. Using these tools, optimal network designs are obtained by formulating a constrained optimization problem involving variables from different layers, and by exploiting information about the fading channel. Solving this optimization problem dictates how resources are allocated across different layers, while the structure of the solution typically indicates how signalling protocols have to be designed.

In this context, the present paper aims to optimally design a wireless network under the following operating conditions. At the transport layer, nodes implement a flow control mechanism to transmit at the highest possible rate, while keeping the network stable. At the network layer, nodes receive packets from different applications, which entail variable utility levels and are destined for different sink nodes [9]. At the link layer, nodes access orthogonally a set of parallel flat fading channels. Orthogonal here means that if a terminal is transmitting, no other link interfering with this transmission can be active [5], [9]. At the physical layer, nodes can adapt their instantaneous power and rate loadings per fading channel realization.

The optimization problem is formulated as a sum-utility maximization (sum-cost minimization) with *variables averaged over all possible channel realizations*. The optimal cross-layer resource allocation turns out to be a function of the *instantaneous* channel state information (CSI), and the optimum Lagrange multipliers associated with the optimization problem. Using duality theory and smooth optimization techniques [1], [14], schemes to find the optimal multipliers are developed. Convergence and optimality of such schemes is characterized, and the operating conditions required to implement the developed schemes are discussed. Last but not least, the paper briefly outlines an alternative to estimate the optimal Lagrange multipliers. The basic idea is to use stochastic estimates of the optimal multipliers that adapt with the CSI changes. This is a valuable alternative when the computational burden to find the optimal multipliers cannot be afforded, or, when the channel probability density function (pdf) is unknown. For a problem slightly simpler than the one considered in this paper, such stochastic schemes have been investigated recently in [5].

There is a large body of works dealing with cross-layer network optimization and control. The main differences of the results in this paper from most of the state-of-the-art works are: accounting for wireless fading effects, orthogonal-access constraints per fading realization, and development of low-complexity stochastic schemes capable of tracking the channel.

Section 2 introduces notation, describes the per-layer operation, and formulates the optimization problem giving rise to the optimal resource allocation. The optimum solution is presented in Section 3. Adaptive algorithms together with their convergence analysis are provided in Section 4. Stochastic estimation of the Lagrange multipliers is the subject of Section 5. Numerical tests and conclusions wrap up this paper.[1].

The work in this paper was supported by the Spanish Ministry of Science and Innovation grant No. TEC2009-12098. Part of the results presented here appeared in [6].

[1]*Notation*: $|\mathcal{X}|$ denotes cardinality of the set $\mathcal{X}$; x^* the optimal value of variable x; ${}_{\{\cdot\}}$ the indicator function (${}_{\{x\}} = 1$ if x is true and zero otherwise); and $[x]_a^b$ the projection of x onto the interval $[a, b]$, i.e., $[x]_a^b = \min\{\max\{a, x\}, b\}$. Finally, for a function $f(\cdot)$, $(f)^{-1}(\cdot)$ denotes its inverse and $\dot{f}(\cdot)$ its derivative.

II. Problem Statement

Consider a multi-hop wireless network with I nodes. Two nodes $i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, I\}$ are physically linked if they can communicate with each other. The set of nodes that a node i can communicate with constitutes the neighborhood of i. For simplicity, we will consider that the neighborhood of each node is the entire network; that is, each node i is physically linked with all other nodes $j \neq i$. Although this represents a worst-case scenario from an interference perspective, it simplifies scheduling. More details about this assumption will be given later. Nodes can transmit orthogonally over a set of K flat-fading parallel channels. The kth channel's instantaneous power gain from node i to node j is denoted by h_{ij}^k; and represents the noise-normalized squared magnitude of the fading coefficient. The overall CSI is described by the random vector $\mathbf{h}$, which collects all h_{ij}^k gains. Channels are assumed ergodic, and are allowed to be correlated.

For this network, we wish to develop adaptive algorithms based on the instantaneous CSI $\mathbf{h}$ to allocate resources at the transport, network, link, and physical layers so that pre-specified QoS metrics are optimized. Each layer's operation is described next.

Transport and Network layer operation: Packets generated exogenously at each node correspond to possibly different applications (such as video, voice, or file transfer), and are destined for different sink nodes. Packet streams will be referred to as flows, and will be indexed by f. The total number of flows will be F. Each node serves flows having other nodes as destinations. The destination node associated with a flow f is denoted by $d(f)$, while the average arrival rate of exogenous packets of flow f to node i is denoted by $\bar{a}_i^f$. The instantaneous rate of flow f which during the channel realization $\mathbf{h}$ is routed from node i to node j is denoted by $r_{ij}^f(\mathbf{h})$. As is customary in communication systems, nodes are assumed equipped with queues (buffers) capable of storing the incoming packets. For these queues to be stable, the following necessary average flow conservation condition needs to be satisfied

$$\bar{a}_i^f + \sum\nolimits_{\forall j \neq i} \left[r_{ji}^f(\mathbf{h}) \right] \leq \sum\nolimits_{\forall j \neq i} \left[r_{ij}^f(\mathbf{h}) \right] \qquad (1)$$

for all (i, f) such that $i \neq d(f)$.

With regard to the transport layer operation, nodes are assumed to implement flow control. This means that the values of variables $\bar{a}_i^f$ will not be fixed, but will be optimally found as the solution of the optimization problem.

Link layer operation: As in [12] and [9], links at the outset are allowed to access simultaneously but orthogonally (in time or frequency) to any of the channels. Consideration of orthogonal access is well motivated in practice, since it lowers system complexity. Moreover, from an optimality perspective, orthogonal access is (nearly) optimal when the interference is strong.

Let $w_{ij}^k(\mathbf{h}) \in [0, 1]$ denote the nonnegative fraction of time that link (i, j) is scheduled to transmit over channel k during the channel realization $\mathbf{h}$. Since every node interferes with all other nodes in the network, it must hold that

$$\sum\nolimits_{(i,j)} w_{ij}^k(\mathbf{h}) \leq 1, \quad \forall k. \qquad (2)$$

This way, if $w_{ij}^k(\mathbf{h}) = 0.9$ and $w_{i'j'}^k(\mathbf{h}) = 0.1$, link (i, j) transmits on k during 90% of the duration of realization $\mathbf{h}$, link (i', j') during 10%, and all other links remain silent. The mechanism to implement channel sharing will depend on the specific system. For example, in existing OFDMA systems, typical bounds on the coherence and symbol intervals are 5-100ms and 5-500 μs, respectively. This means that during a coherence interval, several hundreds of symbols are transmitted; hence, those symbols can be assigned to different links. Even so, the ensuing analysis will show that in most situations time sharing is not needed because the optimal scheduling will assign the channel to a single link.

To simplify exposition, each node's neighborhood is assumed here to be the entire network. For the general case where a transmission only interferes with a small set a links, the orthogonal scheduling will proceed in two phases. First, sets of nodes that can transmit simultaneously without interfering with any other node need to be identified. Second, a constraint similar to (2) with the summation performed across the sets identified in the first phase needs to be imposed. The identification of such sets of nodes requires graph-theoretic tools, and complicates the mathematical formulation throughout. For simplicity, we keep the original formulation in (2). However, it is worth stressing that the results of this paper concerning layers different from the link layer carry over to networks comprising small neighborhoods.

Physical layer operation: The resources adapted at the physical layer will be power and rate per link, per channel (tone), and per CSI realization. Specifically, $p_{ij}^k(\mathbf{h})$ will denote the instantaneous power transmitted over channel k from node i to node j during the channel realization $\mathbf{h}$, if $w_{ij}^k(\mathbf{h}) = 1$. For the general case where $w_{ij}^k(\mathbf{h}) < 1$, the power effectively transmitted over channel k from node i to node j during realization $\mathbf{h}$ is $p_{ij}^k(\mathbf{h}) w_{ij}^k(\mathbf{h})$. Furthermore, from an implementation point of view, two power constraints will be considered. On the one hand, the instantaneous power per channel $p_{ij}^k(\mathbf{h})$ will be bounded by a maximum pre-specified level $\check{p}_{ij}^k$ (spectrum mask constraint). On the other hand, the maximum average power node i can transmit is $\check{p}_i$. To impose the latter constraint, we will use the variable $\bar{p}_i$ and require the following:

$$\left[\sum\nolimits_k \sum\nolimits_{j \neq i} w_{ij}^k(\mathbf{h}) p_{ij}^k(\mathbf{h}) \right] \leq \bar{p}_i, \text{ and } \bar{p}_i \leq \check{p}_i \ \forall i. \qquad (3)$$

The reason for introducing the average power $\bar{p}_i$ as an explicit variable is to decouple variables in the optimization algorithms. This will not incur loss of optimality because the optimal solution will always set $\left[\sum\nolimits_k \sum\nolimits_{j \neq i} w_{ij}^k(\mathbf{h}) p_{ij}^k(\mathbf{h}) \right] = \bar{p}_i$.

Under bit error rate or capacity constraints, rate and power variables are coupled. This rate-power coupling will be represented by the function $C_{ij}^k(\mathbf{h}, p_{ij}^k(\mathbf{h}))$. It will be assumed throughout that the rate-power function $C_{ij}^k(\mathbf{h}, p_{ij}^k(\mathbf{h}))$ is increasing and strictly concave. For instance, if sufficiently strong error control coding is employed, $C_{ij}^k(\mathbf{h}, p_{ij}^k(\mathbf{h}))$ will be given by Shannon's capacity formula $\log(1 + h_{ij}^k p_{ij}^k(\mathbf{h}))$, which is certainly increasing and strictly concave.

A. Problem formulation

This section formulates the optimization problem for the optimal resource allocation policies. The resulting algorithms will be designed so that lower average power consumption and higher exogenous average arrival rates are promoted. (Recall that the latter constitutes a flow control mechanism.) To this end, the power cost functions $J_i(\cdot)$ will be chosen to be strictly convex and increasing, and the rate utility functions $U_i^f(\cdot)$ will be selected strictly concave and increasing. Note that different flows f may (and in general will) require different utility functions. For example, utility functions that are almost linear are appropriate for best-effort traffic because user satisfaction always increases as rate increases. Under these considerations, the optimal channel-adaptive cross-layer resource allocation will be obtained as the solution of the following optimization problem

$$\min_{\substack{\bar{a}_i^f, \bar{p}_i, r_{ij}^f(\mathbf{h}), \\ w_{ij}^k(\mathbf{h}), p_{ij}^k(\mathbf{h})}} \quad -\sum\nolimits_{(i,f)} U_i^f\left(\bar{a}_i^f\right) + \sum\nolimits_i J_i\left(\bar{p}_i\right) \tag{4a}$$

$$\text{subj. to}: \ (1),\ (2),\ (3),\ \text{and}$$
$$\sum\nolimits_f r_{ij}^f(\mathbf{h}) \leq \sum\nolimits_k w_{ij}^k(\mathbf{h}) C_{ij}^k(\mathbf{h}, p_{ij}^k(\mathbf{h})). \tag{4b}$$

Although not explicitly written in (4), all optimization variables are constrained to be non-negative, and $p_{ij}^k(\mathbf{h})$ cannot exceed $\check{p}_{ij}^k$.

The cross-layer nature of the resource allocation problem is apparent because variables of different layers are jointly optimized. The channel-adaptive attribute is also apparent because the optimization variables $r_{ij}^f(\mathbf{h})$, $w_{ij}^k(\mathbf{h})$, and $p_{ij}^k(\mathbf{h})$ are all functions of $\mathbf{h}$. The interaction among layers is manifested in (4b), which ensures that the number of packets routed during the channel realization $\mathbf{h}$ never exceeds the instantaneous capacity of the wireless link.

From an optimality perspective, problem (4) is nearly convex. Specifically, the only source of non-convexity are the monomials $w_{ij}^k(\mathbf{h})p_{ij}^k(\mathbf{h})$ and $w_{ij}^k(\mathbf{h})C_{ij}^k(\mathbf{h}, p_{ij}^k(\mathbf{h}))$. This source of non-convexity can be eliminated by introducing the auxiliary variables $u_{ij}^k(\mathbf{h}) := w_{ij}^k(\mathbf{h})p_{ij}^k(\mathbf{h})$. It can be shown that if $p_{ij}^k(\mathbf{h})$ in (4) is replaced with $u_{ij}^k(\mathbf{h})/w_{ij}^k(\mathbf{h})$, the problem becomes convex. Equally important, the reformulated problem yields the same Lagrangian, and the same optimality conditions as those of (4). Details on this convexifying transformation can be found in, e.g., [12] and [7]. Since both problems yield the same optimality conditions, the original formulation in (4) will be retained for brevity, without explicitly introducing the auxiliary variables $u_{ij}^k(\mathbf{h})$.

III. Optimum Resource Allocation

The main objective of this section is to find the optimum solution of (4). This will be accomplished using the Lagrange-dual approach, and specifically the Karush-Kuhn-Tucker (KKT) conditions [1] associated with (4). Due to space limitations, the KKT conditions and the mathematical derivations leading to the optimal solution will be omitted, but they follow along the lines of [7]. It will turn out that the optimal allocation of resources at the physical, link, network, and transport layers is expressible in terms of the optimum multipliers associated with the average constraints in (4), and the instantaneous CSI $\mathbf{h}$. The second part of the section is devoted to a slight modification of the optimum scheduling and routing schemes that is asymptotically optimal, and offers advantages relative to the optimal ones.

A. Characterization of the Optimal Solution

To introduce notation, let ρ_i^f and π_i denote the Lagrange multipliers associated with the average constraints in (1) and (3), respectively. Consider also $\rho_{ij}^* := \max_f[\rho_i^{f*} - \rho_j^{f*}]$, which will play an instrumental role in describing the optimal policies. Moreover, $\dot{J}_i^{-1}(\cdot)$, $\dot{U}_i^{f\ -1}(\cdot)$ and $\dot{C}_{ij}^{k\ -1}(\mathbf{h}, \cdot)$ will denote, respectively, the inverse function of the derivative of $J_i(\cdot)$, $U_i^f(\cdot)$ and $C_{ij}^k(\mathbf{h}, \cdot)$. With this notation, the optimum allocation of average power, average arrival rate, and instantaneous power is described in the following three propositions[2].

Proposition 1: *The optimum average power allocation is*

$$\bar{p}_i^* = \left[\dot{J}_i^{-1}(\pi_i^*)\right]_0^{\check{p}_i}. \tag{5}$$

This result follows from minimizing the Lagrangian of (4) w.r.t. $\bar{p}_i$, and projecting the solution onto the feasible set $[0, \check{p}_i]$.

Proposition 2: *The optimum average arrival rate allocation is*

$$\bar{a}_i^{f*} = \left[\dot{U}_i^{-1}(\rho_i^{f*})\right]_0^{\infty}. \tag{6}$$

Similar to Proposition 1, this result follows from minimizing the Lagrangian of (4) w.r.t. $\bar{a}_i^f$. As expected, the optimal flow control takes into account the utility corresponding to each flow as well as the price paid for injecting more exogenous traffic into the network (represented by ρ_i^{f*}).

Proposition 3: *The optimum instantaneous power allocation is*

$$p_{ij}^{k*}(\mathbf{h}) = \left[\dot{C}_{ij}^{k\ -1}\left(\mathbf{h}, \pi_i^*/\rho_{ij}^*\right)\right]_0^{\check{p}_{ij}^k}. \tag{7}$$

Interestingly, when $C_{ij}^k(\mathbf{h}, p_i^k) = \log(1 + h_{ij}^k p_i^k)$, (7) reduces to the well-known water-filling formula $p_{ij}^{k*}(\mathbf{h}) = [\rho_{ij}^*/\pi_i^{k*} - 1/h_{ij}^k]_0^{\check{p}_{ij}^k}$. For this specific example, higher values of π_i^* entail lower power and rate loadings (average power is a limiting factor), while higher values of ρ_{ij}^* entail higher power and rate loadings (satisfaction of average flow conservation constraint is critical, thus high rates are required). In fact, it is easy to see that the previous observations hold for any $C_{ij}^k(\mathbf{h}, \cdot)$ increasing and concave.

While the optimal values $\bar{p}_i^*$, $\bar{a}_i^{f*}$, and $p_{ij}^{k*}(\mathbf{h})$ can be found in closed form, obtaining the optimal expressions for $r_{ij}^{f*}(\mathbf{h})$ and $w_{ij}^{k*}(\mathbf{h})$ is more intricate. The reason is that the Lagrangian

[2]Proofs of propositions are omitted due to space limitations.

of (4) is linear w.r.t. those variables, and dual Lagrangian methods are known to be challenged by linear constraints.

For this reason, to find the optimum $w_{ij}^{k*}(\mathbf{h})$ per link we need to define first the functional

$$\varphi(\mathbf{h}, k, i, j) := -\rho_{ij}^* C_{ij}^k(\mathbf{h}, p_{ij}^{k*}(\mathbf{h})) + \pi_i^* p_{ij}^{k*}(\mathbf{h}) \tag{8}$$

which represents the instantaneous cost of scheduling channel k to link (i, j); that is, the cost of selecting $w_{ij}^{k*} = 1$. Secondly, with $\wedge$ denoting the "and" operator, we define $\mathcal{S}_W(\mathbf{h}, k) := \{(i, j) : (i, j) = \arg\min_{(i',j')} \varphi(\mathbf{h}, k, i', j') \ \wedge \ \varphi(\mathbf{h}, k, i, j) < 0\}$. With these notational conventions, the following result holds.

Proposition 4: *The optimal instantaneous scheduling* $w_{ij}^{k*}(\mathbf{h})$ *satisfies:*
(i) If $(i, j) \notin \mathcal{S}_w(\mathbf{h}, k)$*, then* $w_{ij}^{k*}(\mathbf{h}) = 0$.
(ii) If $|\mathcal{S}_W(\mathbf{h}, k)| > 0$*, then* $\sum_{(i,j)\in\mathcal{S}_W(\mathbf{h},k)} w_{ij}^{k*}(\mathbf{h}) = 1$.

In words, the optimal solution schedules only links with minimum negative cost, which can be viewed as a greedy policy because for a given channel k most links are not selected to be scheduled. This policy is oftentimes referred to as opportunistic allocation or a "winner-takes-all" solution. To find the optimum scheduling percentages among the links that attain the minimum cost, two different cases must be considered. If the minimum cost is attained by at most one link, then the second part of Proposition allows writing the optimum instantaneous link scheduling in closed form as

$$w_{ij}^{k*}(\mathbf{h}) = \ {}_{\{(i,j)\in\mathcal{S}_W(\mathbf{h},k)\}}. \tag{9}$$

If several links attain the minimum cost, i.e., if $|\mathcal{S}_W(\mathbf{h}, k)| > 1$, then the calculation of the percentage for each of the links is more complicated. Section discusses this issue further.

A similar approach will be followed to find the optimum $r_{ij}^{f*}(\mathbf{h})$. The first step is to define the functional

$$\phi(i, j, f) := \rho_j^{f*} - \rho_i^{f*} \tag{10}$$

which represents the cost of routing flow f into link (i, j). The second step is to define the set of optimum flows $\mathcal{S}_F(i, j) := \{f : f = \arg\min_{f'} \phi(i, j, f') \ \wedge \ \phi(i, j, f) < 0\}$, and the optimum instantaneous rate for link i, j as $C_{ij}^*(\mathbf{h}) := \sum_{\forall k} w_{ij}^{k*}(\mathbf{h}) C_{ij}^{k*}(\mathbf{h}, p_{ij}^{k*}(\mathbf{h}))$. Using these notational conventions, it can be shown that:

Proposition 5: *The optimum instantaneous routing* $r_{ij}^{f*}(\mathbf{h})$ *satisfies:*
(i) If $f \notin \mathcal{S}_F(i, j)$*, then* $r_{ij}^{f*}(\mathbf{h}) = 0$.
(ii) If $|\mathcal{S}_F(i, j)| > 0$*, then* $\sum_{f\in\mathcal{S}_F(i,j)} r_{ij}^{f*}(\mathbf{h}) = C_{ij}^*(\mathbf{h})$.

As before, the optimal solution is greedy, meaning that only flows with minimum negative cost are routed. If the minimum cost is attained by at most one flow, we have

$$r_{ij}^{f*}(\mathbf{h}) = \ {}_{\{f\in\mathcal{S}_F(i,j)\}} C_{ij}^*(\mathbf{h}). \tag{11}$$

If several flows attain the minimum cost, the specific value for each flow can be found using the results of Section . Note that Proposition states the operation of the routing protocol. Only routes (hops) that decrease the value of ρ_i^{f*} are allowed. It is also important to stress that the value of the gain of a given link does not affect how different flows share the link, but only the number of packets actually routed through the link; i.e., only $C_{ij}^*(\mathbf{h})$.

B. Tie Resolution: Winner-Takes-Almost-All

The event of having different flows (or links) attaining the minimum cost will be henceforth referred to as a "tie". The main difficulty in dealing with a tie is that Proposition 4-*(ii)* and Proposition III-A-*(ii)* do not specify how the resources have to be split among winners. The underlying reason is that although from a dual perspective any arbitrary splitting minimizes the Lagrangian of (4), from a primal point of view only a subset of those (in many cases a single one) is the actual solution of the original constrained problem.

One way to find the optimal primal solution when a tie occurs consists of selecting, among all possible tied flows (schedulings), the one satisfying the average constraints with equality. However, this approach does not lead to a closed-form solution. Moreover, it requires knowledge of the exact Lagrange multiplier values, hence it is very sensitive to small inaccuracies.

To bypass such problems, we advocate a smooth suboptimum scheme to resolve ties that: a) can be implemented for any number of elements in $\mathcal{S}_F(i, j)$ and $\mathcal{S}_W(\mathbf{h}, k)$; b) is available in closed form (thus incurs reduced computational burden); and c) is continuous w.r.t. the Lagrange multipliers. Equally important, the next section will establish analytically that the proposed scheme is asymptotically optimal. To distinguish the smooth near-optimum schemes developed next from their optimum counterparts, the notation $\tilde{x}^*$ will be used henceforth to denote the near-optimum version of the optimum x^*.

The first step to derive the smooth schemes is to realize that the number of elements in $\mathcal{S}_F(i, j)$ and $\mathcal{S}_W(\mathbf{h}, k)$ is critical for the behavior of the optimal policies. Since the condition for being a member of those sets is very restrictive (the cost has to be exactly equal to the minimum cost), we will bypass the problem by relaxing the definition of sets $\mathcal{S}_F$ and $\mathcal{S}_W$ so that more elements belong to them. Let ε_W denote a small positive number, and $\boldsymbol{\lambda}$ a vector containing all the Lagrange multipliers (dual variables) associated with average constraints. Then, the functional cost in (8) can be used to define the minimum link cost and the modified set of (sub) optimum links, respectively, as

$$\varphi^{k*}(\mathbf{h}, \boldsymbol{\lambda}) \ := \ \min_{(i,j)} \varphi_{ij}^k(\mathbf{h}, \boldsymbol{\lambda}) \tag{12}$$

$$\mathcal{S}_W(\mathbf{h}, \boldsymbol{\lambda}, k) \ := \ \{(i, j) : \varphi_{ij}^k(\mathbf{h}, \boldsymbol{\lambda}) < \min\{0, \varphi^{k*}(\mathbf{h}, \boldsymbol{\lambda}) + \varepsilon_W\}\}. \tag{13}$$

Based on these, the following modified (smooth) link scheduling is proposed

$$\tilde{w}_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}) = \ {}_{\{(i,j)\in\tilde{\mathcal{S}}_W(\mathbf{h},\boldsymbol{\lambda},k)\}} \times \frac{\left(1 - \frac{\varphi_{ij}^k(\mathbf{h},\boldsymbol{\lambda}) - \varphi^{k*}(\mathbf{h},\boldsymbol{\lambda})}{\varepsilon_W}\right)^2}{\sum_{(i',j')\in\tilde{\mathcal{S}}_W(\mathbf{h},\boldsymbol{\lambda},k)} \left(1 - \frac{\varphi_{i'j'}^k(\mathbf{h},\boldsymbol{\lambda}) - \varphi^{k*}(\mathbf{h},\boldsymbol{\lambda})}{\varepsilon_W}\right)^2}. \tag{14}$$

The smooth scheduling in (14) allows links whose associated cost is not minimum but ε_W-close to the minimum to be scheduled for transmission too, but in a proportional way; that is, links with lower cost will transmit during more time. It is important to remark that for most channel realizations, the set $\tilde{\mathcal{S}}_W(\mathbf{h}, \boldsymbol{\lambda}, k)$ contains a single element, which means that a single link will "win" the channel. Strictly speaking, when $|\tilde{\mathcal{S}}_W(\mathbf{h}, \boldsymbol{\lambda}, k)| > 1$, the scheduling in (14) is not optimum, and may incur a small penalty (always smaller than ε_W). However, (14) offers two advantages: it is available in closed form, and it is continuous w.r.t. $\boldsymbol{\lambda}$. These will be exploited in the subsequent sections; see also [7] for a detailed derivation, and justification of (14) for cellular networks.

We will proceed in a similar manner to find the smooth instantaneous routing $\tilde{r}_{ij}^{f*}(\mathbf{h})$. With ε_F denoting a small positive number, we first define the minimum flow cost, and the (sub) optimum set of flows, respectively, as

$$\phi_{ij}^*(\boldsymbol{\lambda}) := \min_f \phi_{ij}^f(\boldsymbol{\lambda}), \tag{15}$$

$$\tilde{\mathcal{S}}_F(\boldsymbol{\lambda}, i, j) := \{f : \phi_{ij}^f(\boldsymbol{\lambda}) < \min\{0, \phi_{ij}^*(\boldsymbol{\lambda}) + \varepsilon_F\}\}. \tag{16}$$

Based on definitions (15)-(16), the smooth suboptimum routing for a specific flow f on link (i,j) is

$$\tilde{r}_{ij}^{f*}(\mathbf{h}, \boldsymbol{\lambda}) = \tilde{C}_{ij}^*(\mathbf{h}, \boldsymbol{\lambda}) \; \mathbb{1}_{\{f \in \tilde{\mathcal{S}}_F(\boldsymbol{\lambda}, i, j)\}} \times \frac{\left(1 - \frac{\phi_{ij}^f(\boldsymbol{\lambda}) - \phi_{ij}^*(\boldsymbol{\lambda})}{\varepsilon_F}\right)^2}{\sum_{f' \in \tilde{\mathcal{S}}_F(\boldsymbol{\lambda}, i, j)} \left(1 - \frac{\phi_{ij}^{f'}(\boldsymbol{\lambda}) - \phi_{ij}^*(\boldsymbol{\lambda})}{\varepsilon_F}\right)^2}, \tag{17}$$

where $\tilde{C}_{ij}^*(\mathbf{h}, \boldsymbol{\lambda}) := \sum_k \tilde{w}_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}) C_{ij}^{k*}(\mathbf{h}, p_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}))$. As before, (17) allows flows whose cost is not minimum but ε_F-close to the minimum to be routed too, but in a way that flows with lower cost will route more packages.

C. Layered Resource Allocation

Although the formulation in (4) allows for arbitrary dependence among variables of different layers, it turns out that the optimal schemes of this section exhibit a layered structure. Specifically, the power and rate loadings at the physical layer depend only on the channel gain, and the Lagrange multipliers [c.f. (7)]. Once the physical layer is fixed, the links to be activated can be found via (8), (12), (13), and (14). Finally, once the link and the physical layers are solved, $\tilde{C}_{ij}^*(\mathbf{h})$ can be readily obtained, and then the instantaneous routing can be found using (17), which requires only knowledge of the Lagrange multipliers. In fact, it is easy to see that the way flows share a specific link does not depend on the lower layers; only the total (aggregate) number of packets to be routed depends on $\tilde{C}_{ij}^*(\mathbf{h})$. The basic idea is that the Lagrange multipliers encapsulate (most of) the cross-layer information, which is relevant from a resource allocation point of view. These findings are consistent with those in [2] (non-fading case), and in [10] (fading case).

IV. Ergodic Resource Allocation

A. Finding the Optimum Lagrange Multipliers

Let $\mathbf{y}$ denote a vector containing all average primal variables, $\mathbf{x}(\mathbf{h})$ a vector containing all the instantaneous primal variables. Clearly, to implement the optimum resource allocation schemes of the previous section, the optimum multiplier vector $\boldsymbol{\lambda}^*$ must be known, which amounts to finding ρ_i^{f*} and π_i^* $\forall i, f$. However, $\boldsymbol{\lambda}^*$ cannot be obtained analytically from the KKT conditions associated with (4), and numerical search is required. This is possible using dual methods. First, let us write[3] a simplified version of the Lagrangian of (4) where only the contribution of the average constraints is included

$$\begin{aligned}\mathcal{L}(\mathbf{y}, \mathbf{x}(\mathbf{h}), \boldsymbol{\lambda}) &:= -\sum_{\forall i,f} U_i^f\left(\bar{a}_i^f(\boldsymbol{\lambda})\right) + \sum_{\forall i} J_i\left(\bar{p}_i(\boldsymbol{\lambda})\right) \\ &+ \sum_{\forall i,f} \rho_i^f \left(\bar{a}_i^f(\boldsymbol{\lambda}) + \sum_{j \neq i} \mathbb{E}[r_{ji}^f(\mathbf{h}, \boldsymbol{\lambda})] - \sum_{j \neq i} \mathbb{E}[r_{ij}^f(\mathbf{h}, \boldsymbol{\lambda})]\right) \\ &+ \sum_{\forall i} \pi_i \left(\mathbb{E}\left[\sum_{\forall k} \sum_{j \neq i} w_{ij}^k(\mathbf{h}, \boldsymbol{\lambda}) p_{ij}^k(\mathbf{h}, \boldsymbol{\lambda})\right] - \bar{p}_i(\boldsymbol{\lambda})\right).\end{aligned} \tag{18}$$

Recall that all the instantaneous constraints in (2) and (4b) as well as all the non-negativity constraints were already satisfied when the solution of the previous section is implemented. With $\mathcal{F}(\mathbf{h})$ denoting the feasible set for the instantaneous primal variables, the dual function is defined as

$$D(\boldsymbol{\lambda}) := \inf_{\mathbf{y}, \mathbf{x}(\mathbf{h}) \in \mathcal{F}(\mathbf{h})} \mathcal{L}(\mathbf{y}, \mathbf{x}(\mathbf{h}), \boldsymbol{\lambda}) = \mathcal{L}(\mathbf{y}^*(\boldsymbol{\lambda}), \mathbf{x}^*(\mathbf{h}, \boldsymbol{\lambda}), \boldsymbol{\lambda}), \tag{19}$$

which is always concave w.r.t. $\boldsymbol{\lambda}$ [1]. Based on (19), the dual problem of (4) is

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} D(\boldsymbol{\lambda}). \tag{20}$$

Since the problem in (4) is convex, as long as (4) is strictly feasible, the duality gap between the primal and dual problems is zero. As a result, the value of $\boldsymbol{\lambda}$ optimizing (20) can be used to find the optimum primal solution[4]. A standard approach to obtain $\boldsymbol{\lambda}^*$ is through a gradient iteration. However, this is impossible here because the linear constraints in (4) render $D(\boldsymbol{\lambda})$ non-differentiable w.r.t. some of the entries of $\boldsymbol{\lambda}$. In this case, one can resort to subgradient iterations. For the function in (19), it is known that a subgradient at a given point $\boldsymbol{\lambda}$ represents the constraint violation when the primal solution $\mathbf{x}^*(\mathbf{h}, \boldsymbol{\lambda})$ and $\mathbf{y}^*(\boldsymbol{\lambda})$ is implemented [1, Secs. 6.1 and 6.3].

For small decreasing but non-absolutely summable stepsizes, subgradient iterations are known to converge in the dual domain. However, for a finite number of iterations finding a (near-) feasible primal solution is not guaranteed. The problem is that $r_{ij}^{f*}(\mathbf{h}, \boldsymbol{\lambda})$ and $w_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda})$ are typically discontinuous at $\boldsymbol{\lambda}^*$; therefore, even small hovering in the dual domain around $\boldsymbol{\lambda}^*$ can give rise to a major difference in

[3] Throughout this section, dependence on $\boldsymbol{\lambda}$ will be made explicit wherever it contributes to clarity.

[4] Remarkably, for a problem very similar to (4), [10] has recently shown that even if the physical layer allocation is not convex (i.e., if C is not concave), as far as the pdf of the fading channel is continuous, the duality gap still is zero.

the primal domain. Fortunately, this problem does not appear in the suboptimal scheduling and routing policies of Section III-B. The reason is that when viewed as a function of $\boldsymbol{\lambda}$, $\tilde{r}_{ij}^{f*}(\mathbf{h}, \boldsymbol{\lambda})$ and $\tilde{w}_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda})$ are Lipschitz continuous. (Recall that while for the optimum scheduling the transition from a tie to a single-winner is abrupt, for the suboptimum scheduling the transition is smooth, thus avoiding any discontinuity.) Lipschitz continuity guarantees that "proximity" in the dual domain implies "proximity" in the primal domain. In the context of optimization algorithms, smoothing techniques have been successfully used as a means of effecting continuity or differentiability; see e.g., [14] and [8].

Before presenting the results for the smooth case, we need to define the smooth version of the dual function

$$\tilde{D}(\boldsymbol{\lambda}) := \mathcal{L}(\tilde{\mathbf{y}}^*(\boldsymbol{\lambda}), \tilde{\mathbf{x}}^*(\mathbf{h}, \boldsymbol{\lambda}), \boldsymbol{\lambda}); \tag{21}$$

and the smooth version of the subgradient as the vector $\boldsymbol{\partial}\tilde{D}(\boldsymbol{\lambda})$ collecting the following scalar values

$$\begin{aligned} \partial \tilde{D}_{\rho_i^f}(\boldsymbol{\lambda}) \;&:=\; \bar{a}_i^{f*}(\boldsymbol{\lambda}) + \sum_{j \neq i} \; [\tilde{r}_{ji}^{f*}(\mathbf{h}, \boldsymbol{\lambda})] \\ &-\; \sum_{j \neq i} \; [\tilde{r}_{ij}^{f*}(\mathbf{h}, \boldsymbol{\lambda})], \qquad (22) \\ \partial \tilde{D}_{\pi_i}(\boldsymbol{\lambda}) \;&:=\; \sum_{\forall k} \sum_{j \neq i} \; [\tilde{w}_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}) p_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda})] \\ &-\; \bar{p}_i^*(\boldsymbol{\lambda}), \qquad (23) \end{aligned}$$

which correspond to the violation of the average constraints in (1) and (3), respectively. Alternatively, (22) and (23) can also be viewed as epsilon-subgradients of the original dual function of (4); see [1, Ch. 6] and [7] for details.

Based on these definitions, the following convergence and optimality result can be established.

Proposition 6: *If μ denotes a small constant stepsize, then for any $\boldsymbol{\lambda}^{(0)}$ there exists μ so that:*
(i) the iteration

$$\boldsymbol{\lambda}^{(l)} = \left[\boldsymbol{\lambda}^{(l-1)} + \mu \boldsymbol{\partial}\tilde{D}(\boldsymbol{\lambda}^{(l-1)})\right]_0^{\infty} \tag{24}$$

converges, i.e., $\boldsymbol{\lambda}^{(l)} \to \tilde{\boldsymbol{\lambda}}^*$; and
(ii) at the convergence point: $D(\boldsymbol{\lambda}^*) \leq \tilde{D}(\tilde{\boldsymbol{\lambda}}^*) < D(\boldsymbol{\lambda}^*) + f(\varepsilon_W, \varepsilon_F)$, where f is a positive increasing function which satisfies $f(\varepsilon_W, \varepsilon_F) \to 0$ as $(\varepsilon_W, \varepsilon_F) \to (0, 0)$.

Proposition 6 has various implications. As far as convergence is concerned, it provides a systematic algorithm to compute $\tilde{\boldsymbol{\lambda}}^*$. From a feasibility perspective, it guarantees that if $\tilde{\mathbf{x}}^*(\mathbf{h}, \boldsymbol{\lambda})$ is implemented at $\tilde{\boldsymbol{\lambda}}^*$, the average flow conservation and power constraints are satisfied with equality (recall that $\partial\tilde{D}(\boldsymbol{\lambda}) = \mathbf{0}$ only if this holds). Finally, from an optimality perspective, it guarantees that the overall price paid for implementing the smooth instead of the optimum policy is asymptotically small[5]. A rigorous proof of this assertion for a problem with similar structure can be found in [7].

[5]In practice, the gap w.r.t. $D(\boldsymbol{\lambda}^*)$ is almost zero for any small value of ε_W and ε_F. This is true because the smooth schemes are suboptimum only when ties occur, which are rare events.

B. Operation Mode: Offline and Online Phases

The proposed cross-layer channel adaptive schemes operate in two phases: (i) an offline phase which takes place before communication starts during the initialization phase; and (ii) an online phase which is executed during the communication process, every time the instantaneous CSI $\mathbf{h}$ is updated.

The main objective of the offline phase is to find the Lagrange multipliers $\tilde{\boldsymbol{\lambda}}^*$. Note that $\tilde{\boldsymbol{\lambda}}^*$ encapsulates the effect of channel distributions, and the effect of the *average* QoS and structural (system-level) constraints. As explained before, $\tilde{\boldsymbol{\lambda}}^*$ is found by running the iterations in (24). Since (24) are basically dual subgradient iterations, they will exhibit linear convergence. In practice, this implies that speed of converge will not be high, and hundreds of iterations will be needed to find values reasonably close to the optimum ones. Of course, the specific number will depend on factors such as the initialization point, the chosen stepsize value, and the accuracy required. More importantly, the computational burden per iteration can be high, especially for large-scale networks. The reason is that the subgradients in (24) involve expectations over the channel distribution [cf. (22) and (23)]. In practice, these expectations will be computed using Monte Carlo simulations. The higher the size of the network, the higher the number of channel realizations needed to obtain a reliable estimate of the actual subgradients. Although the offline phase may incur high computational burden, it only needs to be re-run every time the channel statistics or the system set-up changes.

In contrast, the online phase needs to be executed every time the CSI changes. Based on the current value of the CSI $\mathbf{h}$, and the value of $\tilde{\boldsymbol{\lambda}}^*$ obtained from the offline phase, the resources at different layers are adapted according to (7)-(17). Although the optimal allocation at the physical and network layers only requires local CSI, obtaining the optimum link allocation requires knowledge of the full $\mathbf{h}$ vector. Specifically, to find the optimum scheduling, the link that gives rise to the lowest cost needs to be found. To this end, nodes need to share either the channel gains of their local links, or, the value of their link cost indicators. Exchange of this information can be implemented using different schemes. Using a decentralized approach, nodes will require signalling channels that allow nodes to exchange their local information. Under a hierarchical approach, the network can contain scheduler-node(s) that gather the information (these can be regular or dedicated nodes), find the optimal link allocation, and broadcast it to the transmitting nodes.

V. Stochastic Lagrange Multipliers

The resource allocation algorithms developed in the previous sections are functions of two variables: the current CSI $\mathbf{h}$, and the optimum (smooth) Lagrange multipliers. As mentioned earlier, finding $\tilde{\boldsymbol{\lambda}}^*$ offline requires knowledge of the channel distribution, and incurs considerable computational burden. To bypass these challenges, resources can be allocated using stochastic approximation algorithms. These algorithms learn the unavailable information on-the-fly, exhibit tracking capabilities, and incur moderate computational complexity.

Roughly speaking, they could be understood as "intelligent" least mean-square (LMS) type schemes. From an operational perspective, they operate as *fully online* solutions because they do not require offline calculations, and consume limited computational resources [13]. Different alternatives can be considered to develop such stochastic schemes. In this section, we will summarize some of the results presented in [6], which deal with a slightly simpler version of (4), and focus exclusively on stochastic schemes.

The basic idea in [6] is to replace the optimal Lagrange multipliers with their stochastic estimates that are adapted with time. Specifically, let n index blocks whose duration is the channel coherence interval. Then it is clear that the CSI varies with n (to emphasize this variation $\mathbf{h}$ can be written as $\mathbf{h}[n]$), and thus the optimal resource allocation varies in the same time scale. The approach in [6] consists of making the Lagrange multipliers also dependent on n; i.e., the optimum $\tilde{\boldsymbol{\lambda}}^*$ are replaced with stochastic estimates denoted by $\hat{\boldsymbol{\lambda}}[n]$. To implement the optimum policies, ρ_i^{f*} and π_i^* are replaced with the stochastic estimates $\hat{\rho}_i^f[n]$ and $\hat{\pi}_i[n]$.

The stochastic estimates $\hat{\rho}_i^f[n]$ and $\hat{\pi}_i[n]$, [6] rely on a Robins-Monroe scheme. The idea is to use the offline (ensemble) iterations in (22) and (23), but replace all ensemble/average terms with unbiased one-shot estimates that will depend on n. Specifically, recall that μ is a *constant* stepsize, and let $a_i^{f*}[n]$ denote the instantaneous arrival of flow f at node i during block n (whose expected value is $\bar{a}_i^{f*}(\hat{\boldsymbol{\lambda}}[n])$). Based on these, the following stochastic rules are proposed to estimate the Lagrange multiplier values

$$\hat{\rho}_i^f[n+1] = \Big[\hat{\rho}_i^f[n] + \mu(a_i^f[n] + \sum\nolimits_{j\neq i} r_{ji}^{f*}(\mathbf{h}[n], \hat{\boldsymbol{\lambda}}[n]) - \sum\nolimits_{j\neq i} r_{ij}^{f*}(\mathbf{h}[n], \hat{\boldsymbol{\lambda}}[n]))\Big]_0^\infty \quad (25)$$

$$\hat{\pi}_i[n+1] = \Big[\hat{\pi}_i[n] - \mu(\bar{p}_i^*(\hat{\boldsymbol{\lambda}}[n]) - \sum\nolimits_k \sum\nolimits_{j\neq i} w_{ij}^{k*}(\mathbf{h}[n], \hat{\boldsymbol{\lambda}}[n]) p_{ij}^{k*}(\mathbf{h}[n], \hat{\boldsymbol{\lambda}}[n]))\Big]_0^\infty . \quad (26)$$

As explained before, (25) and (26) are stochastic versions of the smooth gradients in (22) and (23).

Assuming that the updates in (25) and (26) are bounded, the following result can be established.

Proposition 7: *Given* $A > 0$, *there exists a random variable* $W(\mu)$ *and time instant* t *so that*

$$\max_{n\geq t} \Pr\{\|\tilde{\boldsymbol{\lambda}}^{(n)} - \hat{\boldsymbol{\lambda}}[n]\| > A\} \leq \Pr\{W(\mu) > A\} \quad (27)$$

where $W(\mu) \to 0$ *w.p.1 as* $\mu \to 0$.
Proposition 7 asserts that although the dual iterates do not strictly converge to the optimum value, with arbitrarily high probability, they will hover within a small neighborhood of it. Together with convergence results for the dual variables, convergence of the primal averages is also characterized. The following holds for the primal averages.

Proposition 8: *The sample average of the stochastic primal variables converges to the ensemble average of the primal solution of* (4) *with probability one (w.p.1).*

Proposition 8 guarantees optimality of the stochastic schemes from the point of view of (4). To be more specific, consider the transmit power as an example, and let $\hat{\bar{p}}_i[n] := n^{-1}\sum_{r=1}^n \sum_{k,j\neq i} \tilde{w}_{ij}^{k*}(\mathbf{h}[r], \hat{\boldsymbol{\lambda}}[r]) p_{ij}^{k*}(\mathbf{h}[r], \hat{\boldsymbol{\lambda}}[r])]$ denote the sample average of the stochastic instantaneous power allocation in (3). Recall that if the optimum values of the Lagrange multipliers were known, the ensemble average of the optimum instantaneous power at node i would be given by $\tilde{\bar{p}}_i^* := [\sum_{k,j\neq i} \tilde{w}_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}^*) p_{ij}^{k*}(\mathbf{h}, \boldsymbol{\lambda}^*)]$. Then, Proposition 8 establishes that $\hat{\bar{p}}_i[n] \to \tilde{\bar{p}}_i^*$ as $n \to \infty$ w.p.1.

Following the lines of [5], [6] also establishes a relationship between the stochastic Lagrange multipliers in (25), and the size of the queues where packets wait to be transmitted. Such a relationship is meaningful from different points of view: i) to analyze the stability of our resource allocation algorithms; ii) to estimate the queueing delay of the packets; and iii) to establish connections with other well-known cross-layer resource allocation algorithms (e.g., with the celebrated dynamic back-pressure algorithm [3]). We refer readers to [6] for further elaboration on this issue.

VI. Numerical results

Numerical results are presented for an any-to-any wireless network where all nodes are connected. To gain insight about the behavior of the developed schemes, we consider the very simple case of a network with $I = 3$ nodes, $F = 3$ different flows (one for each possible destination), and $K = 4$ parallel channels. The channel SNRs are exponentially distributed (i.e., the channel amplitudes follow the classical Rayleigh model), and their average power gain is 5dB. The utilities to be maximized are $U_i^f(x) = \log(1 + x)$, if $(i, f) = (1, 3)$ or $(i, f) = (3, 2)$; $U_i^f(x) = \log(1/10 + x)$, if $(i, f) = (2, 1)$; while zero for all other node-flow combinations. Logarithmic utilities are typically employed to guarantee fairness among users. To penalize high power consumption, we adopt quadratic power cost functions $J_i(x) = x^2/10 \ \forall i$. The coefficient $1/10$ has been introduced to balance the effect of utilities and costs in the objective of the optimization problem. Moreover, a maximum average transmit-power of $\check{p}_i = 5$ per node is considered, and the constant stepsize is $\mu = 3 \cdot 10^{-4}$.

We first test the behavior of the offline schemes developed in Section IV. Figure 1 depicts the value of different parameters versus the number of off-line iterations, namely: (a) the value of the average power transmitted by each node (top); (b) the average routing variables, where only non-zero values are plotted (middle); and (c) the Lagrange multipliers associated with the power constraint (low). The latter has been plotted to gain insights about the optimal policies. The main observation is that all variables converge within hundreds of iterations. The convergence point will depend on factors such as the initialization point, and the stepsize. We also observe that convergence occurs both for dual and primal variables. As discussed in Section IV, this was not guaranteed if the smooth allocation was not implemented; see [7] for an example illustrating such convergence problems.

The performance of the network during the online phase when the optimal multipliers are used is the following: $\bar{p}_1 =$

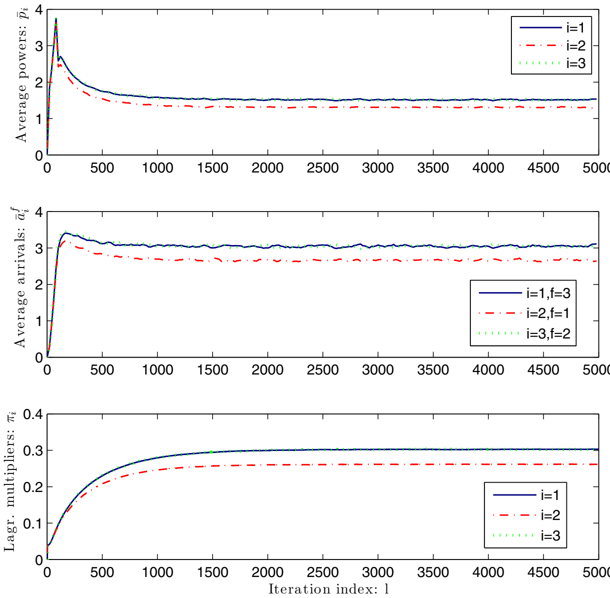

Fig. 1. Trajectories of different primal and dual variables for offline iterations. (Lines for $i = 1$ and $i = 3$ are extremely close.)

1.5, $\bar{p}_2 = 1.3$, $\bar{p}_3 = 1.5$; $\bar{a}_1^3 = 3.0$, $\bar{a}_2^1 = 2.6$, $\bar{a}_3^2 = 3.0$; and $\bar{r}_{1,3}^3 = 3.0$, $\bar{r}_{2,1}^1 = 2.6$, $\bar{r}_{3,2}^2 = 3.0$; while zero for all other variables. Note first that the network conditions are similar for nodes 1 and 3. This is not true for node 2, because it routes packets from flow 1, which according to the utilities considered, yield lower utility. Taking into account these facts, we observe that the numerical results follow the expected behavior since power and rate performance are similar for nodes 1 and 3; and the optimal exogenous rate injected at node 2 is smaller than that at nodes 1 and 3. To further validate the developed schemes, we slightly change the set-up and reduced the average channel gain between nodes 3 and 2 by 10dB. The modified configuration yields the following: $\bar{p}_1 = 1.8$, $\bar{p}_2 = 1.3$, $\bar{p}_3 = 1.2$; $\bar{a}_1^3 = 2.6$, $\bar{a}_2^1 = 2.3$, $\bar{a}_3^2 = 1.5$; and $\bar{r}_{1,3}^3 = 2.6$, $\bar{r}_{2,1}^1 = 2.3$, $\bar{r}_{3,2}^2 = 0.3$, $\bar{r}_{3,1}^2 = 1.2$, $\bar{r}_{1,2}^2 = 1.2$; while zero for all other variables. Since the channel between 3 and 2 is now very poor, most of the packets from 3 destined for 2 are routed through 1. Indeed, this is confirmed by the numerical results. Moreover, we also observe that the power consumed by node 1 increases, the exogenous rate injected at node 3 decreases, and the overall network performance decreases. Clearly, all these changes were caused by the SNR loss between nodes 3 and 2.

VII. Conclusions

This paper dealt with the design of optimal adaptive cross-layer algorithms for multi-hop fading wireless networks that utilize orthogonal access at the link layer. For such networks, channel-adaptive algorithms were developed based on the available CSI to specify power, rate, scheduling, routing and flow control decisions so that a prescribed cost (utility) function is minimized (maximized).

The resultant resource allocation is a function of the current channel realization and the Lagrange multipliers whose values depend on the channel pdf, and the QoS requirements. Different strategies to find the optimum value of such multipliers were proposed, and the operating conditions required to implement the optimal adaptive schemes were analyzed. The analysis reveals that a layered approach, where resources at every layer are optimized separately, is optimal when the proper information is exchanged among layers. This result is in accordance with those previously reported for (fading) wireless networks with different operating conditions than those considered in this paper. The last part of the paper was devoted to briefly analyze stochastic algorithms that acquire the channel statistics on-the-fly, and are provably convergent.

Several extensions of the work presented here are worth investigating. Regarding the link layer, we are currently working on orthogonal access in multi-hop networks (where interference is mostly local and nodes located sufficiently apart can transmit simultaneously) as well as on non-orthogonal access for multi-hop networks (which is challenging since the resultant problem is non-convex). Similarly, consideration of delay as a fundamental measure of QoS and means to incorporate the cost of the signalling protocols into the optimization problem are of high interest.

References

[1] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[2] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures", *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255-312, Jan. 2007.

[3] L. Georgiadis, M. 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, Apr. 2006.

[4] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, 2nd Ed., Springer, 2003.

[5] A. G. Marques, G. B. Giannakis, and F. J. Ramos, "Stochastic resource allocation for orthogonal access based on quantized CSI: Optimality, convergence and delay analysis," in *Proc. of ICASSP*, Taipei, Taiwan, April 9-14, 2009.

[6] A. G. Marques, G. B. Giannakis, and J. Ramos, "Stochastic resource allocation for orthogonal access based on quantized CSI: Optimality, convergence and delay analysis," in *Proc. of ICASSP*, Dallas, TX, March 14-19, 2010.

[7] A. G. Marques, G. B. Giannakis, and J. Ramos, "Optimizing orthogonal multiple access based on Quantized channel state information", submitted to the *IEEE Trans. on Info. Theory*, available at arxiv.org.

[8] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematic Programming, Ser. A*, vol. 103, pp. 127152, 2005.

[9] A. Ribeiro and G. B. Giannakis, "Optimal FDMA over wireless fading mobile ad-hoc networks," *Proc. of ICASSP*, Las Vegas, Nevada, March 30-April 4, 2008.

[10] A. Ribeiro and G. B. Giannakis, "Layer Separability of Wireless Networks," *Proc. of Conf. on Info. Sciences and Systems*, Princeton Univ., NJ, March 19-21, 2008.

[11] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice Hall, 1995.

[12] C.Y. Wong, R.S. Cheng, K.B. Lataief, R.D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[13] X. Wang, G. B. Giannakis, and A. G. Marques, "A Unified Approach to QoS-Guaranteed Scheduling for Channel-Adaptive Wireless Networks," Proceedings of the IEEE, vol. 95, no. 12, pp. 2410-2431, Dec. 2007.

[14] S. A. Zenios, M. C. Pinar, and R. S. Dembo, "A Smooth Penalty Function Algorithm for Network Structured Problems", *European J. of Operational Research*, vol. 83, pp. 220–236, May 1995.