

Minimum-Time Link Scheduling for Emptying Wireless Systems: Solution Characterization and Algorithmic Framework

Vangelis Angelakis, Anthony Ephremides, Qing He and Di Yuan

Linköping University Post Print



N.B.: When citing this work, cite the original article.

©2014 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Vangelis Angelakis, Anthony Ephremides, Qing He and Di Yuan, Minimum-Time Link Scheduling for Emptying Wireless Systems: Solution Characterization and Algorithmic Framework, 2014, IEEE Transactions on Information Theory, (60), 2, 1083-1100.

<http://dx.doi.org/10.1109/TIT.2013.2292065>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-104836>

Minimum-Time Link Scheduling for Emptying Wireless Systems: Solution Characterization and Algorithmic Framework

Vangelis Angelakis, *Member IEEE*, Anthony Ephremides, *Life Fellow IEEE*,
Qing He, *Student Member IEEE*, and Di Yuan, *Member IEEE*

Abstract

We consider a set of transmitter-receiver pairs, or links, that share a wireless medium, and address the problem of emptying backlogged queues with given initial size at the transmitters in minimum time. The problem amounts to determining activation subsets of links, and their time durations, to form a minimum-time schedule. Scheduling in wireless networks has been studied under various formulations before. In this paper, we present fundamental insights and solution characterizations that include: (i) showing that the complexity of the problem remains high for any continuous and increasing rate function, (ii) formulating and proving sufficient and necessary optimality conditions of two baseline scheduling strategies that correspond to emptying the queues using “*one-at-a-time*” or “*all-at-once*” strategies, (iii) presenting and proving the tractability of the special case in which the transmission rates are functions only of the cardinality of the link activation sets. These results are independent of physical-layer system specifications and are valid for any form of rate function. We then develop an algorithmic framework for the solution to this problem. The framework encompasses exact as well as sub-optimal, but fast, scheduling algorithms, all under a unified principle design. Through computational experiments we finally investigate the performance of several specific algorithms from this framework.

Index Terms— algorithm, optimality, scheduling, wireless networks.

I. INTRODUCTION

For multiple communication links with a shared wireless medium, the fundamental aspect of access coordination is called scheduling. It amounts to deciding which links are allowed to transmit simultaneously and for how long they should do so. Usually, the selection of a schedule is driven by the goal of optimizing a cost criterion. Scheduling has a long history of investigation that has ranged from simple transmission models to fully cross-layered ones that combine rate and power control with overall network resource allocation. In this paper, we examine a version of the scheduling problem that arises from the objective of draining in minimum time the bit-contents that reside at the transmitters of a finite number of links. That is, we consider the multiple access or interference channel with finite traffic volume that must be delivered in minimum time, which is also referred to as minimum-time scheduling.

Research of scheduling in wireless networks dates back to the 80s. In [21], a centralized, polynomial-time algorithm was presented for the problem setup in which the network is mapped to an undirected graph and it is assumed that any two links can be successfully activated simultaneously as long as they do not share common vertices of the graph. This type of graph modeling approach was also known as the “protocol” model, with which greedy-type heuristics have been developed for minimum-time scheduling in [29], [31]. Optimal and approximation algorithms for scheduling in networks of trees and planar graphs are provided in [30]. In contrast to the “protocol” model, scheduling with the so called “physical” model (e.g., [6], [20]) accounts for cumulative interference, by using signal-to-interference-and-noise ratio (SINR) constraints for defining feasible activation of multiple links.

Parts of this paper were presented at the 2012 IEEE International Symposium on Information Theory and the 2012 IEICE International Symposium on Information Theory and its Applications.

All authors are with the Department of Science and Technology (ITN), Linköping University, SE-601 74 Norrköping, Sweden (e-mail: vanan@itn.liu.se, antep@itn.liu.se, qing.he@liu.se, diyua@itn.liu.se).

A. Ephremides is also with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA (e-mail: etony@umd.edu).

This work has been supported in part by the Excellence Center at Linköping-Lund in Information Technology (ELLIIT), the Swedish Research Council (Vetenskapsrådet), the EC Marie Curie Actions projects MESH-WISE (FP7-PEOPLE-2012-IAPP: 324515) and Career LTE (FP7-PEOPLE-2013-IOF: 329313), and the NSF grant CCF-0728966.

The authors thank the reviewer and editor for their insightful comments that have greatly helped to improve the manuscript.

Problem complexity has been studied for several specific settings of the scheduling problem. The problem's general hardness under the protocol model is provided in [2]. In [9], [19] \mathcal{NP} -hardness was addressed for the problem of determining a minimum-time schedule under a given traffic demand in a wireless network with SINR constraints. In some special cases the structure of the traffic demand allowed a polynomial algorithm [8]. In [7] it was shown that more fundamental resource allocation problems in wireless networks with SINR constraints, such as node and link assignment, are also \mathcal{NP} -hard.

Scheduling in wireless networks can be represented using a set-covering formulation of integer programming. This enables a column generation method for solving the resulting linear programming relaxation. The notion was introduced in [6], [7]. For the minimum-time scheduling problem, a column-generation-based solution method was also used in [23], which can approach an optimal solution, with the advantage of a potentially reduced complexity. Approximation of minimum-time scheduling with SINR requirements is investigated in [34], [35]. In [26] the minimum-time scheduling problem was formulated as a shortest path problem on directed acyclic graphs, and the authors obtained sub-optimal analytic characterizations. It is also possible to “absorb” the scheduling task in general network resource allocation problems as done in [17]. A recent survey of cross-layer optimization of scheduling, rate adaptation, channel assignment, and routing is provided in [10]. However, basic versions of scheduling remain important, and largely unresolved, both from the theoretical standpoint and from that of specific applications.

It is worth noting the structural differences between scheduling in wireless networks and classical scheduling problems in combinatorial optimization. The latter includes machine and shop scheduling, for which a vast amount of literature is available (e.g., [13], [22]). In machine and shop scheduling, tasks are assigned to machines or processors, possibly with constraints on the processing order, such that the time for completing the tasks is minimized. This type of task assignment is not present in our scheduling problem, where the decision consists of which links should be bundled together for “joint processing” of their queues. Second, and more importantly, for machine and shop scheduling, the processing time of a task on a machine is a constant. In contrast, in our case the scheduling elements interact with each other in processing time. Namely, the rate at which a link's queue is drained is a (continuous or discrete) function of which other links are transmitting simultaneously. Because of this major structural difference, new fundamental understanding of the relation between the schedule-dependent transmission rates and optimality as well as a unified algorithmic approach call for novel research.

In the aforementioned references of scheduling in wireless networks, the problem is studied under specific constraints on feasible grouping of links and the resulting transmission rates. We target new insights into optimal scheduling without restricting to specific system settings. In the following, we outline the main contributions along three lines: complexity, optimality condition of baseline scheduling solutions, and a unified algorithmic framework.

Previous complexity analysis [2], [7], [9], [19] has established the hardness of the scheduling problem with discrete rates, that is, when the rates form a discrete set, each corresponding to an SINR threshold. The result is not a surprise, as determining the optimal rate combinations across links makes the problem combinatorial. A natural follow-up question is whether the complexity changes if the rate is a continuous (and thus much more well behaved) function of SINR. We reveal the fact that the problem's \mathcal{NP} -hardness is preserved for arbitrary continuous rate functions, with a formal proof given in Section V. To see the underlying rationale, note that a schedule is defined by subsets of links with joint transmissions, as well as the transmission duration of each subset. There is a discrete choice in link grouping, i.e., the selection of subsets. In constructing a link group, any scheduling algorithm has to make the binary choice of determining whether or not each link is active in the group. For an active link, the rate depends on interference, which in its turn is determined by group composition. Thus, even when rate is a continuous function of SINR, the problem's combinatorial nature is exhibited in selecting links to form transmission groups, and the rates that can be achieved follow from this discrete choice. This observation is the basis of using fractional graph coloring in the complexity proof in Section V.

Our second line of contributions consists of general optimality conditions for two baseline schedules, namely, obtained by deploying the two straightforward strategies “*one-at-a-time*” and “*all-at-once*”, respectively. The former represents the classic access scheme of time division multiple access (TDMA), that is, one link at a time accesses the channel to empty its backlogged data. The latter strategy organizes transmissions as in the classic interference channel model (i.e., a number of transmitters try to communicate their separate information to their respective receivers via a common channel simultaneously). The resulting scheduling solution amounts to activating the grand set of all links, until (at least) one link has no data left to send, followed by activating the remaining links, and so on. Not only do the two strategies represent classic channel access schemes, but they are also easy to construct

no matter what the rate function is. We are thus interested in understanding under which general conditions these strategies are optimal. Intuitively, “*one-at-a-time*” is optimal if the links pose heavy interference to each other, such that the rate reduction due to interference is significant. This could be the case, for example, in a cellular system where two users are close to each other and on the edges of their respective serving cells. Conversely, “*all-at-once*” is preferable for a low-interference scenario, where simultaneous transmissions have virtually no impact on each others’ rate. This occurs, for example, if the transmitters deploy low power and each receiver is close to its transmitter (e.g., a small-cell environment). Intuitions of optimality, however, give only partial insights. Our contribution lies in providing necessary and/or sufficient conditions that are generally applicable, quantitatively relating the structure of the optimal schedule, in particular “*one-at-a-time*” or “*all-at-once*”, to rates that the links can transmit at, when these rates depend explicitly or implicitly on the set of links that transmit simultaneously. Thus our results strengthen the understanding of how the interaction between link rates affect the performance of scheduling solutions and optimality, and thereby contributing to the tightening of the joint use of physical and MAC layer approaches.

Part of our study deals with the special case of cardinality-based rate. In this case, the transmission rate is determined solely by the cardinality of the group of active links instead of the specific individual elements of the group, and the links in a group share a common rate. An example scenario with such symmetric rates consists of transmitters that are co-located at a central point, resembling a system with a multiple-transmitter base station, and receivers having the same distance (on a circle) from the center with identical geometric channel gains. Consequently the interference is a function of the number of interferers only. The case of cardinality-based rate provides more structure. We investigate this structure and reveal that it has two implications. First, our optimality characterization for “*one-at-a-time*” or “*all-at-once*” become stronger when the rates are symmetric. Second, we prove the implication of cardinality-based rate on tractability, namely, the global optimum can be computed in polynomial time irrespective of the function that relates the SINR to rate. The polynomial-time tractability makes this optimum schedule a reasonable candidate solution for scenarios that approximately exhibit the structure of cardinality-based rate.

On the solution implementation side, our contribution is the development of a unified algorithmic framework with a modular design. The design originates from the natural standpoint that any scheduling algorithm will have to deal with two tasks, namely constructing link activation subsets and determining the transmission duration of each subset. Based on this observation, the framework applies the structural decomposition of deploying two task-specific modules. One module determines the link subsets and the other module the duration of their activation. The modules may apply either exact or various sub-optimal algorithms for accomplishing their respective tasks. Each specific design choice of these modules yields a complete scheduling algorithm. The two modules operate interactively; that is, information of the preference in group construction and the constructed group are exchanged between the two modules. This modular design enables a unified view that includes previous algorithms deploying greedy search (e.g., greedy time-slot assignment in [7]), finding shortest path in a acyclic graph of exponential size [26], and exact algorithm guaranteeing optimality by column generation [23], as well as enabling new algorithms to be constructed. The optimality characterization of the aforementioned specific scheduling solutions has some interesting algorithmic implications. First, when their respective optimality conditions are close to be satisfied, these specific schedules constitute reasonable heuristic solutions. Second and more importantly, the optimality conditions can serve as baseline checkpoints in algorithm construction. Namely, assuming the conditions hold, they can be used to test whether an algorithm is able to deliver the respective optimum schedules, and we will formally provide this validity check of the rationale of a set of specific algorithms in Section X. We evaluate extensively the performance of ten different algorithms chosen from the proposed framework, with the following key findings. First, the overall performance has a high correlation to that of the module for constructing link subsets. Second, the preference in subset selection should account for the remaining queue size if the backlogged queues are gradually drained. Third, if link rate becomes quickly close-to-flat in respect of interference, sub-optimality in either of the modules has noticeable impact.

The paper is structured as follows. In Section II we present the overall system model, formalize the problem, and motivate our investigation on the optimality conditions. We then discuss the models of rate functions we consider in Section III. We provide our first structural results in Section IV, where we also give the Linear Programming (LP) formulations of the problem. In Section V we give the proof of the \mathcal{NP} -hardness of the problem, and in Section VI we give necessary and/or sufficient conditions for the optimality of the two baseline schedules. In Section VII we uncover the tractability of the problem in the case of cardinality-based rates. In Section VIII we introduce the

modular algorithmic framework, and in Section IX we expand it to enable algorithms based on column generation. In Section X we evaluate the complexity and optimality of sample algorithms within the framework, whereas in Section XI we numerically evaluate and compare ten algorithms we developed within our framework. In Section XII we conclude the paper.

II. SYSTEM MODEL

We consider a set $\mathcal{N} = \{1, \dots, N\}$ of links that share a wireless medium. These links are associated with a demand vector $\mathbf{d} = \{d_1, \dots, d_N\}'$ of real and strictly positive numbers, with each d_i representing the amount of bit-traffic stored at the queue of the transmitter of the corresponding link i . Throughout the paper, the transpose of a vector \mathbf{a} is denoted by \mathbf{a}' . Here \mathbf{d} is defined as a column vector for simplifying the mathematical formulations later on. It is assumed that the entries in the demand vector are in descending order. The assumption clearly does not cause any loss of generality, because the link indices can always be chosen with descending demand values. Let \mathcal{H} denote the union of all subsets of \mathcal{N} , excluding the empty set. Thus $|\mathcal{H}| = 2^N - 1$. We use the term *group* to refer to a member $c \in \mathcal{H}$, that is, a subset of the link set. Scheduling a group c means that all elements of c are activated simultaneously for a positive amount of time. For any group, the rate of each of its elements is a function of the group composition. Let F denote the rate function; that is, for $c \in \mathcal{H}$ and link $i \in \mathcal{N}$, $r_{ic} = F(i, c)$ represents the non-negative transmission rate of link i , if c is active. Clearly, the rate values can be positive only for the members of c , i.e., $r_{ic} = 0, i \notin c$. If c is a singleton link i , we use r_{ii} as a more convenient, short-hand notation for the rate instead of $r_{i,\{i\}}$. These rates represent feasible rates that are within the capacity region of the system.

In all applications with meaningful physical interpretations, the rates have the following property: If two elements are served together, the rates of being served cannot be higher than the individual rates, respectively. Thus, throughout the paper, it is assumed that the service rate of any link in a group does not increase if the group is augmented, i.e., for any two groups $c_1 \subset c_2$ and $i \in c_1 \cap c_2$, $F(i, c_1) \geq F(i, c_2)$. We refer to this as the *rate monotonicity property*. No further conditions are imposed on F .

The minimum-time scheduling problem, for given (N, \mathbf{d}, F) , amounts to selecting a set of groups $c_1 \dots c_k$, among the $2^N - 1$ members of \mathcal{H} , along with their respective activation durations $T_j, j = 1 \dots k$, so that $\sum_{j=1}^k T_j$ is minimized, subject to the requirement that all stored traffic is successfully delivered. Thus we do not consider external data arrivals to any transmitter queue. For the latter case, which is in its own right extremely interesting and relevant, a variety of optimization criteria, such as discounted, nondiscounted, and average cost, are appropriate to both finite and infinite-horizon versions of the scheduling problem [15], but that is not investigated here.

It is important to stress that the problem input does not include the explicit knowledge of the $2^N - 1$ rate vectors. If these vectors are all computed a priori, solving the problem reduces to optimizing a linear program (of which the size is exponential in N). What is provided in the problem input is the function F , that can be viewed as a black box, or an ‘‘oracle’’, that returns the rate values for any given $c \in \mathcal{H}$. Thus a scheduling algorithm is regarded of exponential complexity, if, in the algorithm, the number of times that function F is invoked is exponential in N . We assume that the computation of F is practically efficient, that is, one function evaluation $F(i, c)$ of any group $c \in \mathcal{H}$ and $i \in c$ runs in polynomial time in N . Note that from a communication/information-theoretic perspective these rate values represent, as noted earlier, any feasible, or achievable, rates for a given channel with specific coding, modulation and detection structures. Thus the treatment of the problem is decoupled from the physical-layer aspects of it, although it is directly connected to, and dependent on, them. Naturally, solving the scheduling problem and at the same time choosing the ‘‘best’’ rates remains an ultimate cross-layer optimization problem that is not studied here.

One specific special case of interest is the symmetric case in which the rate is determined by the group cardinality. That is, F is a function of $|c|$ but not of the group’s specific composition. This scenario has been considered in [8]. When the rates depend only on the group size, the input can be equivalently defined using an N -dimensional rate vector $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$, each denoting the common rate of every link in a group of size $1 \dots N$ respectively. Rate monotonicity then implies that $\tau_1 \geq \dots \geq \tau_N$. We will subsequently use the input triplet $(N, \mathbf{d}, \boldsymbol{\tau})$ to refer to this problem class.

We end this section by some examples, all of the aforementioned class $(N, \mathbf{d}, \boldsymbol{\tau})$, which are simple and of small size, yet they serve well the purpose of motivating our theoretical investigation of optimality conditions by illustrating that intuition may fail in deriving optimal schedules.

Example 1. Consider a case where $N = 3$, $d_1 = 3d$, $d_2 = 2d$, $d_3 = d$ with any $d > 0$, and $\tau_1 = 6$, $\tau_2 = 4.8$, $\tau_3 = 4$, i.e. $(3, (d, 2d, 3d)', (6, 4.8, 4))$.

For Example 1, relation $3\tau_3 > 2\tau_2 > \tau_1$ holds. That is, the sum-rate (or, equivalently, amount of data drained per time unit) follows the order of group size. In this case, a simple and intuitive schedule is to activate all three links together until link three has no data left, followed by activating the group of links one and two and then single-link activation of link one. One can verify that this solution, derived by intuition, is indeed the optimal one.

Example 2. Consider Example 1, now with the slight change that rate τ_2 is 5 instead of 4.8.

Note that $3\tau_3 > 2\tau_2 > \tau_1$ remains true. Yet, group $\{1, 2, 3\}$ having the highest sum-rate is not part of the optimum. The optimum, which is unique in this case, comprises the two groups $\{1, 2\}$ and $\{1, 3\}$, with time durations $\frac{2d}{5}$ and $\frac{d}{5}$, respectively. Hence preferring the top sum-rate group in designing a schedule fails in this case. In addition, the optimality of $\{1, 2\}$ and $\{1, 3\}$ as well as the uniqueness of this optimum remain as long as rate $\tau_2 > 4.8$. The observation justifies questions of optimality characterizations, e.g., under what conditions the schedule derived in Example 1 remains optimal. We will examine this aspect in Section VI.

In the previous two examples, the time durations are such that at least one link gets its entire demand served in each group. In general, one intuitive algorithmic notion is to iteratively construct N groups, and, for each group, apply a time duration (which is straightforward to compute) such that the remaining queue of (at least) one link in the group is emptied.

Example 3. Consider $(3, (d, d, d), \tau)$, with $2\tau_2 > 3\tau_3$ and $2\tau_2 > \tau_1$.

For Example 3, the unique optimum consists of groups $\{1, 2\}$, $\{2, 3\}$, and $\{1, 3\}$, each with a time duration of $\frac{d}{2\tau_2}$. Observe that *none* of the links has its entire queue emptied in any of the groups that participating in the unique optimum. Thus activating a group “as long as possible” may fail, *even if an exhaustive search* of all group combinations is tried. Note that the order of activation of the groups is unimportant for our purposes, while it might become important if “fairness” considerations were to be introduced.

III. THE RATE FUNCTION F

Thus far, the minimum-time scheduling problem has been presented in a rather generic form. The treatment of the scheduling problem in this paper does not depend on a *specific form* of the rate function F and, hence, it applies to emptying N backlogged queues in minimum time for *any* system for which the rates of draining the queues satisfy the rate monotonicity assumption. For wireless networks, saying that a transmission is successful at some given rate on a link, means that for a fairly broad class of channel models and receiver structures the SINR at the receiver exceeds a certain threshold [18]. Specifically, if a channel matrix \mathbf{G} of dimension $N \times N$ is provided, where its element G_{ij} is the channel gain between the transmitter of link i and the receiver of link j and if P_i denotes the power of link i , and σ^2 the noise variance, then for link i in group c the SINR is given by

$$\gamma_{ic} = \frac{P_i G_{ii}}{\sum_{k \in c, k \neq i} P_k G_{ki} + \sigma^2}. \quad (1)$$

This hypothesis is accurate when interference can be modeled as additive white Gaussian noise; otherwise it is an approximation.

For wireless links with shared medium, two commonly used modeling approaches for defining F are as follows. The first is a one-step function returning either zero (no success) or one (success) as the rate value. Indeed, many of the previous studies of scheduling in wireless networks use implicitly this function (e.g. [6], [7]). In effect, the transmission of a packet is successful if and only if the SINR meets a threshold γ^* . A group c such that all of its links can successfully transmit is sometimes referred to as a *feasible matching*. Clearly, an infeasible matching will not be part of the optimal schedule, since if it were to be used, it would de facto be replaced by the subset of its members that satisfy the SINR condition. An equivalent view is, in the definition of F , to set zero rates for *all* elements of any infeasible matching. Thus the following definition of F provides the SINR-threshold-based model of scheduling. In the sequel, we use $F_{\mathbb{B}}$ to denote this binary function.

$$r_{ic} = F_{\mathbb{B}}(i, \mathbf{c}) = \begin{cases} 1 & \text{if } i \in \mathbf{c} \text{ and } \gamma_{jc} \geq \gamma^*, \forall j \in \mathbf{c}, \\ 0 & \text{otherwise.} \end{cases}$$

The definition can be generalized to account for rate adaptation. In this case, the attainable rate values form a discrete set with cardinality higher than two. Each rate value is associated with a corresponding SINR threshold, often obtained from the available adaptive modulation and coding schemes of a specific wireless channel model [33]. The generalization corresponds to defining F as a step-wise function taking multiple values.

The second commonly used modeling approach is to consider the rate as a continuous function of the SINR [18]. We will use $F_{\mathbb{C}}$ as a general notation of the wide class of continuous functions that are (strictly) monotonically increasing in the SINR. A particular case of interest is the Shannon formula for the additive white Gaussian noise (AWGN) channel. This case will be referred to as $F_{\mathbb{S}}$, and is given by

$$r_{ic} = F_{\mathbb{S}}(i, \mathbf{c}) = \log_2(1 + \gamma_{ic}). \quad (2)$$

The aforementioned property of rate monotonicity clearly holds for both $F_{\mathbb{B}}$ and $F_{\mathbb{C}}$. For $F_{\mathbb{B}}$, we have $r_{ic_1} = r_{ic_2}$, for two groups $\mathbf{c}_1 \subset \mathbf{c}_2$ and $i \in \mathbf{c}_1 \cap \mathbf{c}_2$, if and only if both are feasible matchings or both are infeasible matchings. If \mathbf{c}_1 is feasible but \mathbf{c}_2 is not, $1 = r_{ic_1} > r_{ic_2} = 0, i \in \mathbf{c}_1 \cap \mathbf{c}_2$. For $F_{\mathbb{C}}$, strict inequality $r_{ic_1} > r_{ic_2}$ holds as long as $\mathbf{c}_1 \subset \mathbf{c}_2$.

IV. LINEAR PROGRAMMING FORMULATION

The scheduling problem is easily shown to be equivalent to a linear program (LP). Although formulating the LP does not give a practically feasible solution algorithm, it enables us to gain structural insights. Denote by $\mathbf{T} = T_{\mathbf{c}}, \mathbf{c} \in \mathcal{H}$ the non-negative scheduling decision vector of dimension $2^N - 1$, whose element $T_{\mathbf{c}}$ denotes the time duration of running group $\mathbf{c} \in \mathcal{H}$. We use \mathbf{T}^* to denote an optimal scheduling solution. Notation \mathcal{H}^* is reserved for the set of groups that correspond to an optimum solution, that is, $\mathcal{H}^* = \{\mathbf{c} \in \mathcal{H} : T_{\mathbf{c}}^* > 0\}$.

By the following lemma, all demands will be met exactly at optimum. This is rather intuitive and has been (implicitly) taken for granted before (e.g., [8]). Formalizing this result is useful in our case, as it eliminates any doubt about the validity of the form of LP basic solutions to be discussed later.

Lemma 1. *There exists an optimal schedule such that, before reaching the end of the time duration of a group, none of the link queues in the group is empty.*

Proof: Suppose the opposite is true. Then there exists a group \mathbf{c} run with time duration $T_{\mathbf{c}} > 0$ and link $i \in \mathbf{c}$, such that the demand served on link i in the group, denoted by d_{ic} , satisfies the condition $d_{ic} < r_{ic}T_{\mathbf{c}}$. Let $t = \frac{d_{ic}}{r_{ic}}$. Consider splitting the running time $T_{\mathbf{c}}$ in two segments, with lengths t and $T_{\mathbf{c}} - t$ respectively. In the first segment, group \mathbf{c} is run, and for the second segment, the reduced group $\mathbf{c} \setminus \{i\}$ is run. The lemma follows from two observations. First, the served demand of i in segment one remains d_{ic} . Second, any of the links other than i is served for an overall time of $T_{\mathbf{c}}$, and their rates in $\mathbf{c} \setminus \{i\}$ are not worse, if not better, than those in \mathbf{c} . ■

By Lemma 1, we arrive at the following LP formulation.

$$\min \sum_{\mathbf{c} \in \mathcal{H}} T_{\mathbf{c}}, \quad (3a)$$

$$\text{s. t. } \sum_{\mathbf{c} \in \mathcal{H}} r_{ic}T_{\mathbf{c}} = d_i \quad i = 1, \dots, N, \quad (3b)$$

$$\mathbf{T} \geq 0. \quad (3c)$$

As (3b) are equalities, the formulation is in the so called LP standard form, hence no slack or surplus variables will be involved in constructing matrix bases or the corresponding basic solutions.

Even though there are $2^N - 1$ candidate groups, we can conclude the existence of an optimal scheduling solution using at most N groups. The result follows from the fundamental optimality theory of LP and the structure of (3).

Lemma 2. *There exists an optimal scheduling solution using at most N groups, i.e., $|\mathcal{H}^*| \leq N$.*

Proof: LP (3) has N rows. Thus, setting $2^N - 1 - N$ variables to zeros leads to an equation system with a unique solution, if the remaining N variables are linearly independent, i.e., their corresponding columns in the left-hand side of (3b) have full rank. Clearly, the solution represents a feasible schedule, if the N variables have non-negative values. Such a solution is referred to as a basic feasible solution (BFS) in linear programming. By fundamentals of linear programming (e.g., [25]), there exists an optimal BFS, as long as the solution space is non-empty and the optimum objective value is bounded. Both conditions hold for (3) as the N single-link groups (i.e., TDMA activation of the links one-at-a-time) forms a BFS, and the optimum objective value is bounded from below by zero. For the optimal BFS, the number of activated link groups is at most N , and equals N if the solution is non-degenerate, and the lemma follows. \blacksquare

Instead of applying directly linear programming theory as in the proof of Lemma 2, the optimality characterization can be derived as follows. The derivation sheds light on the rationale of why N groups (among the $2^N - 1$ ones) are sufficient for optimality. Suppose \mathbf{T}^* uses K groups with $K > N$, and, without loss of generality, the K groups have column indices $1, 2, \dots, K$ in the left-hand side of (3b). Denote the K columns by $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K$. Schedule \mathbf{T}^* meets the demand by equation, thus $T_1^* \mathbf{r}_1 + T_2^* \mathbf{r}_2 + \dots + T_K^* \mathbf{r}_K = \mathbf{d}$. Because there are N rows in the equation and $K > N$, the K rate vectors must be linearly dependent, i.e., there exists a not-all-zeros vector $\boldsymbol{\lambda} \in \mathbb{R}_K$, such that $\lambda_1 \mathbf{r}_1 + \lambda_2 \mathbf{r}_2 + \dots + \lambda_K \mathbf{r}_K = \mathbf{0}$. The observations leads to $(T_1^* + \epsilon \lambda_1) \mathbf{r}_1 + (T_2^* + \epsilon \lambda_2) \mathbf{r}_2 + \dots + (T_K^* + \epsilon \lambda_K) \mathbf{r}_K = \mathbf{d}$ for any $\epsilon \in \mathbb{R}$. Thus any choice of ϵ gives a feasible schedule, as long as the activation durations $T_1^* + \epsilon \lambda_1, T_2^* + \epsilon \lambda_2, \dots, T_K^* + \epsilon \lambda_K$ remain non-negative.

We construct two new feasible schedules activating at most $K - 1$ groups as follows. Note that rate vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K$ all contain at least one positive value. Therefore vector $\boldsymbol{\lambda}$ must have both strictly positive and strictly negative elements. Let $\hat{\epsilon} = \min_{i=1, \dots, K: \lambda_i > 0} \frac{T_i^*}{\lambda_i}$ and $\check{\epsilon} = \min_{i=1, \dots, K: \lambda_i < 0} \frac{T_i^*}{|\lambda_i|}$. Then $\hat{\mathbf{T}} = \mathbf{T}^* - \hat{\epsilon} \boldsymbol{\lambda} \geq \mathbf{0}$ and $\check{\mathbf{T}} = \mathbf{T}^* + \check{\epsilon} \boldsymbol{\lambda} \geq \mathbf{0}$, that is, $\hat{\mathbf{T}}$ and $\check{\mathbf{T}}$ are two feasible schedules. In addition, it is clear that $\hat{\mathbf{T}}$ and $\check{\mathbf{T}}$ have no more than $K - 1$ positive activation durations. Next, note that \mathbf{T}^* is in fact a convex combination of $\hat{\mathbf{T}}$ and $\check{\mathbf{T}}$: $\mathbf{T}^* = \frac{\check{\epsilon}}{\hat{\epsilon} + \check{\epsilon}} \hat{\mathbf{T}} + \frac{\hat{\epsilon}}{\hat{\epsilon} + \check{\epsilon}} \check{\mathbf{T}}$. Thus $\sum_{i=1}^K T_i^* = \frac{\check{\epsilon}}{\hat{\epsilon} + \check{\epsilon}} \sum_{i=1}^K \hat{T}_i + \frac{\hat{\epsilon}}{\hat{\epsilon} + \check{\epsilon}} \sum_{i=1}^K \check{T}_i$. The three sums are nothing but the schedule lengths of \mathbf{T} , $\hat{\mathbf{T}}$, and $\check{\mathbf{T}}$, respectively. It follows immediately that at least one of $\hat{\mathbf{T}}$ and $\check{\mathbf{T}}$ must outperform or equal to \mathbf{T}^* in schedule length, as otherwise the equation can not hold. Selecting this schedule and repeating the process, we reach the conclusion of Lemma 2.

The above line of argument reveals the rationale of why optimality does not require more than N groups. Namely, any schedule with more than N activation groups can be reduced to schedules with fewer groups, because the rate vectors in the former are linearly dependent. In addition, because the objective function is linear, there is at least one such reduction leading to better or same schedule length. The reduction is no longer applicable for schedules of N groups, hence one of these schedules is optimal. However, even if there is always a compact representation of N groups at optimality, finding this best combination of N groups, among the $2^N - 1$ candidate ones, remains generally hard (see also Section V). In this regard, optimal scheduling has a combinatorial side, even if formulation (3) is an LP. We finally remark that the result of Lemma 2 can also be obtained by combining the Carathéodory theorem [14] with the fact the the optimum schedule is a point on a face of the convex hull spanned by the $2^N - 1$ rate vectors.

In some of the analysis later on, we utilize the LP dual of (3). Letting π_i denote the dual variable of (3b), the dual formulation is as follows.

$$\max \sum_{i \in \mathcal{N}} d_i \pi_i, \quad (4a)$$

$$\text{s. t. } \sum_{i \in \mathcal{C}} r_{i\mathcal{C}} \pi_i \leq 1 \quad \mathcal{C} \in \mathcal{H}. \quad (4b)$$

The dual variables π are not restricted in sign as (3b) are equalities. We note that the LP dual remains valid with non-negativity requirement on π , because (3b) can be stated as inequalities; however, by Lemma 1 equalities will apply at optimum.

In the remainder of the paper, some of the theoretical analysis use either the Karush-Kuhn-Tucker (KKT) optimality condition or the termination criterion in the LP simplex algorithm. These fundamental results are provided

in Appendix A, and used in the subsequent sections to investigate solution optimality.

V. COMPLEXITY CONSIDERATIONS

Complexity is a fundamental aspect in the treatment of optimization problems. By Lemma 2, obtaining the globally optimal schedule is equivalent to selecting the N “best” groups. The question is how difficult this selection task is. For a discrete rate function $F_{\mathbb{B}}$, the problem is \mathcal{NP} -hard [1], [2], [7], [9], [19]. We consider problem complexity for continuous functions of class $F_{\mathbb{C}}$. For example, is the problem tractable if the rate function is $F_{\mathbb{S}}$, or, even simpler, linear (regardless of the fact that it would not be realistic) in SINR? In the following, we provide a negative answer, stating that the problem in the wireless communications context is in general hard for *all* rate functions that are continuous and strictly increasing in the SINR.

Theorem 3. *Given any function $F_{\mathbb{C}}$ of the SINR, there are NP-hard instances of the minimum-time scheduling problem.*

Proof: Given (N, \mathbf{d}, F) , where F is of type $F_{\mathbb{C}}$, the recognition version of the problem, by Lemma 2, is as follows. Are there N groups, which can be represented using a binary $N \times N$ matrix, such that the total time of satisfying \mathbf{d} using these groups is at most a given positive number? The problem is clearly in class \mathcal{NP} , as checking the validity of a solution (a certificate in form of a square matrix of size N) is straightforward.

We proceed constructing a polynomial reduction of the weighted fractional coloring problem [24] to our problem. Consider a general-topology graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let $N = |\mathcal{V}|$. Thus a link in the scheduling instance corresponds to a vertex in \mathcal{G} . Let $v = F^{-1}(\frac{1}{N})$, and $u = \frac{F^{-1}(1)}{F^{-1}(\frac{1}{N})}$, i.e., $F^{-1}(1) = vu$, with $u > 1$ because F is strictly increasing in SINR. Let $\sigma^2 = \frac{1}{u}$. For each edge (i, j) in the graph, set the coupling element $G_{ij} = G_{ji} = 1$. Moreover, $G_{ii} = G = \min\{v, 1.0\}$, $i \in \mathcal{N}$. All other elements of the channel matrix are zeros. Finally, the transmit power $P_i = \frac{v}{G_{ii}}$, $i \in \mathcal{N}$.

Consider link i and any group that contains i , but not any of the adjacent vertexes in \mathcal{G} . The SINR is $vu = F^{-1}(1)$, thus the rate is 1. If i is put in a group containing at least one adjacent vertex in \mathcal{G} , the SINR is no more than $v/(\frac{v}{G} + \frac{1}{u}) < v = F^{-1}(\frac{1}{N})$, because $\frac{v}{G} \geq 1$ and $u > 0$. Thus the rate of i becomes strictly less than $\frac{1}{N}$. Suppose, at optimum, a group \mathcal{c} containing two links i and j that are adjacent in \mathcal{G} has a positive amount of time duration $T_{\mathcal{c}} > 0$. Note that, in \mathcal{G} , \mathcal{c} corresponds to at least one connected component (because i and j are adjacent). Denote by $\mathcal{m} \subseteq \mathcal{V}$ the component containing i and j , and let $m = |\mathcal{m}|$. Note that $m \geq 2$. By the observation before, for each of the links in \mathcal{m} , including i and j , the demand served in time $T_{\mathcal{c}}$ within group \mathcal{c} is strictly less than $\frac{T_{\mathcal{c}}}{N}$.

Consider splitting group \mathcal{c} into m groups, obtained by combining $\mathcal{c} \setminus \mathcal{m}$ with each of the individual links in \mathcal{m} . Each of the m groups is given time $\frac{T_{\mathcal{c}}}{m}$. For all links in \mathcal{m} , including i and j , the rate grows from less than $\frac{1}{N}$ to 1. Since $m \leq N$, the quantity $\frac{T_{\mathcal{c}}}{m}$ is strictly more than enough to serve demand $\frac{T_{\mathcal{c}}}{N}$ for any link in \mathcal{m} . For the links in $\mathcal{c} \setminus \mathcal{m}$, overall, they are served for the same time duration $T_{\mathcal{c}}$, with rate no less than before. Repeat the argument for the remaining components, if necessary. In conclusion, there is an optimal scheduling solution in which the groups are formed by links corresponding to independent sets of \mathcal{G} . It is therefore apparent that solving the scheduling problem provides the correct answer to the weighted fractional coloring problem, with the demand vector \mathbf{d} being the weights of the vertexes, and the result follows. \blacksquare

Theorem 3 establishes the inherent difficulty of the scheduling problem. The result generalizes the observation made in [9] on the connection between fractional coloring and scheduling under the “protocol” model that uses a conflict graph and disregards the channel matrix. As our result applies to *any* function of $F_{\mathbb{C}}$, one should not expect that the use of smooth rate functions, including linear ones, would reduce complexity.

VI. OPTIMALITY CONDITIONS FOR BASELINE SCHEDULING STRATEGIES

Since the determination of the optimal schedule is, and remains complex, it is of interest to investigate when some basic scheduling strategies are optimal. We consider two basic strategies that we referred to as “one-at-a-time” and “all-at-once”, and denote by \mathcal{H}^1 and \mathcal{H}^N , respectively. In \mathcal{H}^1 , the link queues are emptied completely separately, corresponding to TDMA activation. Thus $\mathcal{H}^1 = \{\{1\}, \{2\}, \dots, \{N\}\}$. Strategy \mathcal{H}^N applies the very opposite philosophy of activating groups of which the sizes are as large as possible. Thus the N -links group is activated until some of the queues becomes empty. The next group consists of the links having positive remaining demand. Continuing in this fashion leads to the complete solution of \mathcal{H}^N .

Note that both strategies are of size N in the number of groups and hence they are BFSs. Given N out of the $2^N - 1$ groups, the computing time of the correct time share of the BFS (or concluding that the N groups do not form a feasible schedule) is normally of complexity $O(N^3)$ due to matrix inversion. Solutions \mathcal{H}^1 and \mathcal{H}^N are simpler to construct – after N calls of function F , computing the \mathcal{H}^1 schedule runs clearly in linear time, whereas for the \mathcal{H}^N schedule the computing time is of $O(N^2)$.

We examine conditions under which it is preferable to activate links separately or jointly with \mathcal{H}^1 and \mathcal{H}^N as representative scheduling strategies in Sections VI-A and VI-B, respectively. Then, in Section VI-C we derive the corresponding conditions for the special case of cardinality-based rates.

A. Optimality Conditions for Separate Link Activation

As the first step of our theoretical treatment of optimality condition, we characterize when separate link activation, in particular schedule \mathcal{H}^1 , is preferred. Intuitively, \mathcal{H}^1 is desirable, if the links, when activated simultaneously with others, experience significant rate reduction. This corresponds to a high-interference environment. The following condition quantifies the notion.

Condition 1. For all $c \in \mathcal{H}$, the sum of the ratios between the elements' rates in c and their respective rates of individual activation, is at most 1.0, that is,

$$\sum_{i \in c} \frac{r_{ic}}{r_{ii}} \leq 1 \quad \forall c \in \mathcal{H}.$$

The above condition is simple in structure. Yet, it is exact in characterizing the optimality of \mathcal{H}^1 .

Theorem 4. \mathcal{H}^1 is optimal if and only if Condition 1 holds.

Proof: We develop the proof by applying the reduced-cost criterion in the simplex algorithm (See Appendix A). For sufficiency, consider LP formulation (3), and the basis matrix \mathbf{B} for BFS \mathcal{H}^1 . The inverse matrix \mathbf{B}^{-1} is diagonal with $\text{diag}(\mathbf{B}^{-1}) = (1/r_{11}, \dots, 1/r_{NN})'$. For any non-basic variable T_c with $|c| \geq 2$, the reduced cost equals $1 - \mathbf{e}'\mathbf{B}^{-1}\mathbf{r}_c$, where \mathbf{e}' is a row vector of N ones and \mathbf{r}_c denotes the column vector corresponding to T_c in (3). The expression leads to reduced cost $1 - \sum_{i \in c} \frac{r_{ic}}{r_{ii}}$ that is non-negative if Condition 1 holds. Since none of the $2^N - N - 1$ non-basic variables has strictly negative reduced cost, \mathcal{H}^1 is optimal in the simplex algorithm.

For necessity, assume that Condition 1 does not hold for some group c . Then the reduced cost of the corresponding non-basic variable is strictly negative. Moreover, for \mathcal{H}^1 , all the basic variables have strictly positive values. Therefore the LP pivot operation of bringing in T_c into the base is not degenerate, meaning that the objective function will strictly improve, and the result follows. ■

Theorem 4 provides a complete answer to the optimality of \mathcal{H}^1 . The condition consists of one inequality per group. From the proof, it is clear that reducing the number of inequalities is not possible. However, if we relax the requirement of necessity, and consider a pair of links, there is a simpler sufficient condition that excludes the activation of both of these links in the same group. This occurs, as formulated below, if the two links generate high interference to each other, but their rates are not much affected by simultaneous transmissions of the other links.

Condition 2. For a pair of links $i, j \in \mathcal{N}$, we define the following inequality.

$$\frac{r_{i,\{i,j\}}}{r_{i,\mathcal{N}\setminus\{j\}}} + \frac{r_{j,\{i,j\}}}{r_{j,\mathcal{N}\setminus\{i\}}} \leq 1.$$

Theorem 5. If Condition 2 is true, then there exists \mathcal{H}^* in which i and j do not appear together in any group, that is, in optimizing the schedule, the condition is sufficient for discarding all groups containing both i and j .

Proof: Suppose an optimal schedule has a group c having both i and j . Without loss of generality, assume the time duration of c is 1. The demands served equal r_{ic} and r_{jc} for the two links, respectively. By the monotonicity property of rate we have $r_{ic} \leq r_{i,\{i,j\}}$ and $r_{jc} \leq r_{j,\{i,j\}}$, which yield the following inequality

$$\frac{r_{ic}}{r_{i,\mathcal{N}\setminus\{j\}}} + \frac{r_{jc}}{r_{j,\mathcal{N}\setminus\{i\}}} \leq 1.$$

Consider, instead of c , two groups $c \setminus \{j\}$ and $c \setminus \{i\}$. The rates of i and j are at least $r_{i, \mathcal{N} \setminus \{j\}}$ and $r_{j, \mathcal{N} \setminus \{i\}}$, respectively. Activating the two groups for time durations $r_{ic}/r_{i, \mathcal{N} \setminus \{j\}}$ and $r_{jc}/r_{j, \mathcal{N} \setminus \{i\}}$ delivers respectively no less than r_{ic} and r_{jc} amounts of demand for i and j , hence the conclusion. ■

Remark 1. Both theorems 4 and 5 significantly extend previous results of the optimality characterization of two links (see [28] and the references therein). In fact, for two links, \mathcal{H}^1 is optimal if Condition 1 holds for $c = \{i, j\}$, otherwise \mathcal{H}^N is optimal.

B. Optimality Conditions for Joint Link Activation

As our next part of investigation, we examine when it is preferable to activate links jointly, in particular scheduling strategy \mathcal{H}^N . Consider augmenting the size of any given group (of any size, except N). Intuitively, one can expect that the group should be augmented with a new link, if the resulting sum-rate is higher than that of any time combination of running the group and the link separately. Conversely, if it is optimal to activate group c , then the rates of c cannot be achieved by any combined use of its $|c|$ subsets of size $|c| - 1$. The insight leads to consider the following condition.

Condition 3. Given group c , let $n = |c|$ and denote them c_1, c_2, \dots, c_n its n subsets of cardinality $n - 1$, obtained by deleting each of the n links of c . Denote by $\mathbf{r}_c \in \mathbb{R}_+^n$ the vector of rates of the links in c , and $\mathbf{r}_i \in \mathbb{R}_+^n$ the corresponding rate vector for c_i (with zero rate for i). We define the following condition: For any $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)' \in \mathbb{R}_+^n$ with $\mathbf{e}'\boldsymbol{\lambda} = 1$, the vector inequality $\sum_{i \in c} \lambda_i \mathbf{r}_i \leq \mathbf{r}_c$ is satisfied for at least one element.

Remark 2. Note that finding whether or not there exists a $\boldsymbol{\lambda}$ vector that violates the condition can be formulated as an LP of size $O(n)$. Thus the condition can be checked efficiently for any given group.

What the above condition states is, in fact, that the rate vector of c cannot be outperformed by the throughput region of the n sub-groups. If group c is active at the optimum, then the condition must be true, as formulated below.

Theorem 6. If $c \in \mathcal{H}^*$, then Condition 3 holds.

Proof: Suppose group c is activated with any positive time T_c . Strict inequality $\sum_{i \in c} \lambda_i \mathbf{r}_i > \mathbf{r}_c$ in all the n elements means that running c_1, c_2, \dots, c_n , with time proportions $\lambda_i, i = 1, \dots, n$, respectively, will serve demand $T_c \mathbf{r}_c$ within less time than T_c , and the result follows. ■

We now turn our attention to the scheduling strategy \mathcal{H}^N . In this solution, the N groups, which are easily identified, are of sizes $N, N - 1, \dots, 1$. To save notation without loss of generality, assume link N has its queue emptied first, followed by link $N - 1$ in the second group, and so on. Applying Theorem 6 yields immediately the following necessary condition for the optimality of \mathcal{H}^N .

Corollary 7. If \mathcal{H}^N is optimal, then Condition 3 must be true for the $N - 1$ groups $\{1, \dots, N\}, \{1, \dots, N - 1\}, \dots$, and $\{1, 2\}$.

To arrive at a sufficient optimality condition for strategy \mathcal{H}^N for the general problem setting (N, \mathbf{d}, F) , we consider maximum and minimum rates of all groups. Denote by r_m^{\max} and r_m^{\min} the maximum and minimum link rates, respectively, of all groups of size m . Their exact values are difficult to calculate in general as the number of all groups is exponential. However, for a given rate function F and channel matrix \mathbf{G} , it is possible to numerically derive an optimistic bound (i.e., upper bound) as well as a pessimistic bound (i.e., lower bound) on the achievable link rate for all groups of size m without enumerating group composition. The bounds can be used as replacements of r_m^{\max} and r_m^{\min} in applying the optimality condition for practical systems.

Condition 4. We define the following $N - 1$ inequalities,

$$\frac{1}{r_m^{\min}} + \frac{1}{r_{m-2}^{\min}} \leq \frac{2}{r_{m-1}^{\max}} \quad m = 2, \dots, N.$$

where, by convention, the term $\frac{1}{r_0^{\min}}$, corresponding to $m = 2$, is taken to be zero.

The inequalities in Condition 4 form a chain for group sizes moving from one to N . By the following theorem, this chain of relations is sufficient for the optimality of schedule \mathcal{H}^N .

Theorem 8. *If Condition 4 holds, then \mathcal{H}^N is optimal.*

Proof: By the assumption on the order in which the queues become emptied, $\mathcal{H}^N = \{\{1, \dots, N\}, \{1, \dots, N-1\}, \dots, \{1\}\}$. In case of any degeneracy, the structure remains, with the only difference that the corresponding time duration of some of these groups is zero. All T -variables other than those for the N groups are zeros.

Our proof utilizes the optimality characterization provided by the KKT condition (See Appendix A). Clearly, LP primal feasibility of (3) is satisfied by \mathcal{H}^N . Consider the LP dual (4), and, for all the groups in \mathcal{H}^N , set the corresponding rows in the dual to equality, that is,

$$\sum_{i=1}^N r_{iN} \pi_i = 1, \quad (5a)$$

$$\sum_{i=1}^{N-1} r_{i, \mathcal{N} \setminus \{N\}} \pi_i = 1, \quad (5b)$$

$$\sum_{i=1}^{N-2} r_{i, \mathcal{N} \setminus \{N, N-1\}} \pi_i = 1, \quad (5c)$$

$$\dots = 1,$$

$$r_{11} \pi_1 = 1. \quad (5d)$$

The above N equalities uniquely determine a solution to the LP dual (4). This, together with \mathcal{H}^N , form a pair of dual and primal solutions. Consider the complementary slackness condition. Recall that the condition for (3b) is always satisfied no matter the values of the dual variables, since constraints (3b) are equalities. For the LP dual, complementary slackness holds for the above N rows. For the rest of rows, the condition is also satisfied because the corresponding T -variables are zeros. In conclusion, the pair of solutions is optimal, if LP dual feasibility holds.

Suppose the derived dual solution is not dual feasible, i.e., at least one of the remaining constraints in (4) is violated. We prove a contradiction, assuming a violated constraint concerning group $\{1, \dots, N-2, N\}$ of size $N-1$. The construction for arriving at a contradiction for other groups of size $N-1$ as well as other group sizes is similar.

The above assumption of constraint violation means that $\sum_{i=1}^{N-2} r_{i, \mathcal{N} \setminus \{N-1\}} \pi_i + r_{N, \mathcal{N} \setminus \{N-1\}} \pi_N > 1$. This implies the following inequality.

$$r_{N-1}^{max} \sum_{i=1}^{N-2} \pi_i + r_{N-1}^{max} \pi_N > 1. \quad (6)$$

Note that (5b) implies the inequality $r_{N-1}^{max} \sum_{i=1}^{N-1} \pi_i \geq 1$. This, together with (6), result in

$$r_{N-1}^{max} \sum_{i=1}^N \pi_i > 2 - r_{N-1}^{max} \sum_{i=1}^{N-2} \pi_i. \quad (7)$$

Next, from (5c), we obtain $\sum_{i=1}^{N-2} \pi_i \leq \frac{1}{r_{N-2}^{min}}$. This observation and (7) lead to the inequality below.

$$\sum_{i=1}^N \pi_i > \frac{2}{r_{N-1}^{max}} - \frac{1}{r_{N-2}^{min}}. \quad (8)$$

Scaling (8) by r_N^{min} along with applying Condition 4 result in the following.

$$r_N^{min} \sum_{i=1}^N \pi_i > r_N^{min} \left(\frac{2}{r_{N-1}^{max}} - \frac{1}{r_{N-2}^{min}} \right) \geq 1. \quad (9)$$

The strict inequality $r_N^{min} \sum_{i=1}^N \pi_i > 1$, however, contradicts (5a), and the theorem follows. \blacksquare

C. Optimality Conditions of Baseline Scheduling Strategies for Cardinality-Based Rates

We now consider the special, symmetric case of cardinality-based rates, referred to as $(N, \mathbf{d}, \boldsymbol{\tau})$, which was introduced in Section II. For $(N, \mathbf{d}, \boldsymbol{\tau})$, the theoretical characterization in Sections VI-A and VI-B become simpler and stronger. Consider first \mathcal{H}^1 . The number of inequalities in Condition 1 becomes N . This, together with the proof of Theorem 4, lead to the following corollary.

Corollary 9. For $(N, \mathbf{d}, \boldsymbol{\tau})$, a sufficient and necessary condition for the optimality of \mathcal{H}^1 is

$$m\tau_m \leq \tau_1 \quad m = 2, \dots, N.$$

The structure of Condition 2 also simplifies for $(N, \mathbf{d}, \boldsymbol{\tau})$. In addition, by augmenting the line of arguments in the proof of Theorem 5, we arrive at a sufficient condition for excluding the use of any group of a specific size m . This fact is summarized in the corollary below.

Corollary 10. For $(N, \mathbf{d}, \boldsymbol{\tau})$ and any given group size $m \in [2, N]$, there is an optimal schedule not using any group of size m if the condition below holds for at least one $m' < m$.

$$m\tau_m \leq m'\tau_{m'}.$$

Noting that sum-rate of a group is the aggregated rate of serving all the queues of the group, Corollaries 9 and 10 have the following interpretations. The former corollary indicates that it is beneficial to use a TDMA-based schedule (i.e., single-link activation) if bundling together any number of links results in a loss in the amount of data being drained from the queues per time unit. The latter corollary states that a group of a certain size should not be used if there is a smaller group size that dominates the original group in aggregated service rate.

Having concluded the optimality condition of \mathcal{H}^1 for cardinality-based rates $(N, \mathbf{d}, \boldsymbol{\tau})$, we treat \mathcal{H}^N for this problem class, and prove that the results in Section VI-B become strengthened. Consider the implication of Condition 3. Because of the rate symmetry, the quantity $\sum_{i \in \mathcal{C}} \lambda_i r_i^\gamma$ can attain maximum simultaneously in all the n elements, only if $\lambda_i = \frac{1}{n}$ for all $i = 1, \dots, n$. For this $\boldsymbol{\lambda}$, all elements of $\sum_{i \in \mathcal{C}} \lambda_i r_i^\gamma$ equal $\frac{n-1}{n} \tau_{n-1}$ for $(N, \mathbf{d}, \boldsymbol{\tau})$, resulting in the observation below.

Corollary 11. For $(N, \mathbf{d}, \boldsymbol{\tau})$, a necessary condition for the optimality of \mathcal{H}^N is

$$(m-1)\tau_{m-1} \leq m\tau_m \quad m = 2, \dots, N.$$

The inequalities in the above corollary form a hierarchy of relations with a clean interpretation. Namely, if \mathcal{H}^N is optimal, then the sum-rate must be monotonically increasing in group size. Conversely, if this monotonicity is violated, we conclude \mathcal{H}^N is not optimal. However, the reverse formulation does not hold, i.e., the hierarchy of relations is not sufficient for ensuring that \mathcal{H}^N is optimal. This is indeed shown by Example 2 in Section II.

Next, we show that Theorem 8 has stronger implications for $(N, \mathbf{d}, \boldsymbol{\tau})$ than the general case. Namely, Condition 4 is not only sufficient, but also necessary for the optimality of \mathcal{H}^N , as long as \mathcal{H}^N is non-degenerate, that is, all the N groups are run with strictly positive time durations. This condition is equivalent to requiring that all links have different demands.

Theorem 12. For $(N, \mathbf{d}, \boldsymbol{\tau})$ and assuming all links have different demands, a necessary condition for the optimality of \mathcal{H}^N is

$$\frac{1}{\tau_m} + \frac{1}{\tau_{m-2}} \leq \frac{2}{\tau_{m-1}} \quad m = 2, \dots, N. \quad (10)$$

Proof: As in the proof of Theorem 8, assume without loss of generality that the sequence in which the queues are emptied in \mathcal{H}^N is $N, N-1, \dots, 1$. Consider the group of size $m-1$ consisting of $\{1, 2, \dots, m-2, m\}$. Note that the group is not part of the \mathcal{H}^N solution. We examine this group using the reduced-cost optimality criterion in Appendix A.

The base matrix \mathbf{B} of solution \mathcal{H}^N is triangular for $(N, \mathbf{d}, \boldsymbol{\tau})$, where column k consists of k consecutive elements of value τ_k , followed by $N-k$ zeros. Hence the basis inverse, \mathbf{B}^{-1} , has the following form.

$$\begin{pmatrix} \frac{1}{\tau_1} & -\frac{1}{\tau_1} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\tau_2} & -\frac{1}{\tau_2} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \frac{1}{\tau_{N-1}} & -\frac{1}{\tau_{N-1}} \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\tau_N} \end{pmatrix}$$

For the aforementioned group, the linear programming reduced cost, for the basis \mathbf{B} , is given by the expression below.

$$1 - \mathbf{B}^{-1}(\tau_{m-1}, \tau_{m-1}, \dots, 0, \tau_{m-1}, 0, \dots, 0)' = 2 - \left(\frac{\tau_{m-1}}{\tau_m} + \frac{\tau_{m-1}}{\tau_{m-2}} \right)$$

If the opposite of (10) holds, then the reduced cost is strictly negative. Thus group $\{1, 2, \dots, m-2, m\}$ is an incoming variable in the simplex algorithm. Because the links all have different demands, \mathcal{H}^N is non-degenerate. Hence the pivot operation that brings the group into the basis will strictly improve the objective function, and the theorem follows. ■

Comparing the condition in Theorem 12 with Condition 4, it is clear that the latter reduces to the former for $(N, \mathbf{d}, \boldsymbol{\tau})$. For this problem class, it has been commented earlier that the condition of strictly improving sum-rate in group size, used in Corollary 11, is not sufficient for the optimality of \mathcal{H}^N , whereas the inequalities given in Theorem 12 are. Thus the former is implied by the latter (in a strict sense, because they are not equivalent). This fact is formally established below.

Corollary 13. For $(N, \mathbf{d}, \boldsymbol{\tau})$, $\frac{1}{\tau_m} + \frac{1}{\tau_{m-2}} \leq \frac{2}{\tau_{m-1}}$, $m = 2, \dots, N$, implies $(m-1)\tau_{m-1} \leq m\tau_m$, $m = 2, \dots, N$.

Proof: For $m = 2$, the inequality gives $\tau_1 \leq 2\tau_2$, as $\frac{1}{\tau_0}$ is effectively zero by the aforementioned convention. For the induction step, assume $(k-1)\tau_{k-1} \leq k\tau_k$ and consider $k+1$. The inequality $\frac{1}{\tau_{k+1}} + \frac{1}{\tau_{k-1}} \leq \frac{2}{\tau_k}$ and the induction hypothesis together yield $\frac{1}{\tau_{k+1}} + \frac{k-1}{k\tau_k} \leq \frac{1}{\tau_{k+1}} + \frac{1}{\tau_{k-1}} \leq \frac{2}{\tau_k}$. Comparing the left and right sides, we obtain $k\tau_k \leq (k+1)\tau_{k+1}$, and the corollary follows. ■

Remark 3. The inequalities in the derived conditions do not involve \mathbf{d} . Except from the assumption in Theorem 12, all optimality conditions are valid completely independent of the demand values. This observation is confirmed by the discussion in Appendix A: Given a feasible schedule in form of a (non-degenerate) BFS, its optimality depends only on the left-hand side of (3b), which does not contain \mathbf{d} .

VII. COMPLEXITY OF SCHEDULING WITH CARDINALITY-BASED RATES

From Section VI-C, one can observe that the optimality conditions for \mathcal{H}^1 and \mathcal{H}^N are more structured and stronger for $(N, \mathbf{d}, \boldsymbol{\tau})$. This raises the question whether or not reaching optimality for $(N, \mathbf{d}, \boldsymbol{\tau})$ is more tractable than in the general case. In this section, we provide a positive answer to the question.

Consider first a more restrictive case, where the demand values are uniform. For this setting, we provide an analytic solution requiring only linear time to compute, and prove it is globally optimal.

Theorem 14. For $(N, \mathbf{d}, \boldsymbol{\tau})$, let $m^* = \operatorname{argmax}_{m=1}^N m\tau_m$. If all demand values are uniform and equal to d , then the N groups, $\{1, 2, \dots, m^*\}$, $\{2, 3, \dots, m^* + 1\}$, \dots , $\{N, 1, \dots, m^* - 1\}$, each scheduled for a time duration of $\frac{d}{m^* \tau_{m^*}}$, is optimal.

Proof: For all the links, the given schedule clearly meets demand d exactly. For any feasible scheduling solution (not restricted to the case in question) of length T , the total demand, $\sum_{i \in \mathcal{N}} d_i$, divided by T , gives the average throughput per time unit. As $\sum_{i \in \mathcal{N}} d_i$ is a constant, a schedule is minimum in time if $(\sum_{i \in \mathcal{N}} d_i)/T$ attains the maximum possible value. By the assumption in the theorem, the instantaneous throughput of any feasible schedule can never exceed $m^* \tau_{m^*}$. This throughput is achieved during the entire duration of the scheme in the theorem, and the result follows. ■

Remark 4. Theorem 14 generalizes a result in [8] that is derived for the much more restrictive case of $F_{\mathbb{B}}$, where m^* corresponds to the size of the largest feasible matching. In [8], however, all matchings of size m^* are used for constructing the optimal schedule. In our analysis, only N groups are needed.

For non-uniform demand, $(N, \mathbf{d}, \boldsymbol{\tau})$ does not admit an optimal schedule in closed form, yet we are able to conclude its polynomial-time tractability. This fundamental insight is established in the following theorem.

Theorem 15. $(N, \mathbf{d}, \boldsymbol{\tau})$ is in class P, that is, the global optimum of any instance can be computed in polynomial time.

Proof: Consider the LP dual given in (4). For problem class $(N, \mathbf{d}, \boldsymbol{\tau})$, the dual has the following form.

$$\max \sum_{i \in \mathcal{N}} d_i \pi_i \quad (11a)$$

$$\text{s. t. } \tau_{|c|} \sum_{i \in c} \pi_i \leq 1 \quad c \in \mathcal{H}. \quad (11b)$$

Observe that there is a symmetry among the occurrences of the dual variables in (11b). As a result, given any feasible solution, swapping the values of any two dual variables will preserve feasibility. Recall that the demand vector \mathbf{d} is arranged in descending order and the validity of this assumption is given in Section II. It follows that there must exist an optimal solution with $\pi_N \leq \pi_{N-1} \leq \dots \leq \pi_1$, because otherwise the objective function value can be improved, or kept the same, by swapping variable values so that the condition holds.

Based on the above observation, one concludes that, among all constraints of (11b) with m variables, the inequality $\tau_m \sum_{i=1}^m \pi_i \leq 1$ is the most stringent one in defining the optimum. Therefore, the number of constraints required to define optimum can be reduced from $2^N - 1$ to N , implying that, at the optimum, the scheduling problem is equivalent to the following LP.

$$\max \sum_{i \in \mathcal{N}} d_i \pi_i \quad (12a)$$

$$\text{s. t. } \sum_{i=1}^m \tau_m \pi_i \leq 1, \quad m = 1, 2, \dots, N, \quad (12b)$$

$$\pi_N \leq \pi_{N-1} \leq \dots \leq \pi_1. \quad (12c)$$

In conclusion, the optimal solution to problem class $(N, \mathbf{d}, \boldsymbol{\tau})$ is found by solving an LP of size $O(N)$, and the theorem follows. \blacksquare

The above theorem is significant not only for the special symmetric case $(N, \mathbf{d}, \boldsymbol{\tau})$, but also for all scenarios where the transmitters have similar distances (and hence close-to-uniform channel gains) to their receivers, and the latter are located close to each other. For these cases, one can expect that solving $(N, \mathbf{d}, \boldsymbol{\tau})$, which can be done fast, will provide a good approximate solution to the exact global optimum.

VIII. A UNIFIED ALGORITHMIC FRAMEWORK

Since optimal scheduling is in general complex, it is important to develop algorithms which trade optimality against reduced complexity and yield decent performance. To this end we propose several algorithm variations that range from sub-optimal ones with low complexity to one that is actually optimal but has high complexity. They are all based on a common framework that uses a natural view of the problem and that is based to some degree on some of the optimality conditions and insights derived in the previous sections. In fact we will demonstrate that the proposed modular structure eventually leads to exploiting tools from optimization theory so that we may come close to, or even achieve, full optimality, with reduced complexity.

As is evident from the LP formulation of the problem, any scheduling algorithm will have to have two basic components:

- (i) a method for generating the N link groups that will be part of the proposed final schedule, and
- (ii) a method for deciding the duration of activation for each of these sets.

Later, we will confirm that this structural decomposition actually leads to a powerful toolset for eventual optimization.

Our proposed algorithms use a variety of criteria for implementing the two aforementioned requirements. Each algorithm uses what we call a *Group Generation Module* to select the activation sets and an *Activation Duration*

Module to decide the length of the activation of each set. The two modules do not operate independently, but are closely coupled. Before proceeding to the description and evaluation of these algorithms, we should emphasize that they are not based on ideas like those that govern the so-called approximation algorithms (e.g. [34]), or algorithms that impose structural restriction or additional assumptions on the problem. Instead, our concepts are completely generally applicable. In fact we will show that some of these algorithms do achieve the optimal solution when some of the conditions that were mentioned earlier hold (i.e. in the cases where either \mathcal{H}^1 , or \mathcal{H}^N is optimal).

- Regarding the activation duration module, we consider two possibilities. Either we activate the chosen group until one of its links empties its queue or we activate it for a fixed amount of time Δ , chosen a priori as a parameter. Clearly, in the latter case, it is possible that the time that a queue empties is less than Δ . In that case the termination of the activation period occurs at that time instant, rather than continuing on until time Δ has passed. Therefore for large values of Δ the two criteria become less and less distinguishable. We refer to the first criterion as *TF* (for “time at which the first queue of the group empties”) and to the second as *T Δ* (for “time Δ , unless a queue empties earlier”).
- Regarding the group generation module, some care needs to be exercised because here is the main source of high complexity. Namely, there are $2^N - 1$ possible groups. So to choose groups we must utilize some heuristic in the selection or, without any knowledge of the rate function F , we must endure the full consideration of all groups. In either case, we want to reduce the complexity by avoiding the solution of the full LP in (3). Therefore we must choose a metric by which we will evaluate the candidate groups. To this effect we either consider the sum-rate metric (*SR*), i.e. the quantity $\sum_{i \in c} r_{ic}$, or the weighted sum-rate (*WSR*), i.e. the quantity $\sum_{i \in c} q_i r_{ic}$, where q_i is the “current” queue size at the transmitter of link i . Clearly, at the start we have $q_i = d_i$; however, as different links get activated at different times, each initial d_i keeps diminishing until it reaches zero. Whether we choose the *SR* or the *WSR* metric, we have two choices for selecting a group. Either we look at all $2^N - 1$ groups (or all the remaining groups, after some links have emptied) or we look at a judiciously chosen group that requires a much reduced search. In the first case, we call the selection method *exact*, while in the second we call it *heuristic*. In fact, we will later see that the “*exact*” choice can be achieved without necessarily looking at all $2^N - 1$ possible groups. Thus we have four possible group generation methods: (i) *SR-exact*, (ii) *SR-heuristic*, (iii) *WSR-exact* and (iv) *WSR-heuristic*. Since we may pair any one of these four group generation methods with either the *TF* criterion or the *T Δ* criterion in the activation module, we obtain a total of eight algorithms.

It remains to describe some method for choosing a group. We propose the following. We rank the N links according to their currently remaining queue size q_i in descending order. To form a group we start with the singleton having the link at the top of the list. We then visit the second link of the ranking and pair it with the first one. Doing so, we obtain the rates of concurrent activation of both links, which are, in general, different than those before the links were paired. If the updated metric (*SR* or *WSR*) increases as a result of pairing the two links into a group, we keep the second link in the group. Otherwise, we skip it. We then visit the third link in the ranking and repeat the same process. We proceed in this fashion until all links are visited. Thus one group will emerge at the end of this process. To hedge against this process being highly sub-optimal, we repeat this entire construction for two additional rank-permutations, where we start with the second and third link in the ranking respectively. Thus, in the end, we have three possible candidate groups from which we select the one with the highest metric.

Therefore we now have the following algorithms: (1) *TF-SR-exact*, (2) *TF-SR-heuristic*, (3) *TF-WSR-exact*, (4) *TF-WSR-heuristic*, (5) *T Δ -SR-exact*, (6) *T Δ -SR-heuristic*, (7) *T Δ -WSR-exact*, (8) *T Δ -WSR-heuristic*.

IX. GENERALIZING THE FRAMEWORK WITH OPTIMIZATION TOOLS

As we hinted earlier, the simple algorithms that were described in the preceding section can be embedded into a considerably more general setting that can exploit a variety of optimization techniques to yield a better combination of performance and complexity. Specifically, we may now consider the activation duration module as a more sophisticated process. Instead of choosing a simplistic criterion for activation (like the *TF* and the *T Δ* strategies), the module can actually obtain a “tentative” set of activation times that are actually optimal for a much reduced set of groups. That is, it can be thought of as solving the LP over a small, limited, and restricted set of link groups. Once it does this, it feeds back to the group generation module a “metric” that is based on the dual variables of the LP. This metric is then used (in lieu of the *SR*, or the *WSR* metric) to select a new group. The new

group is fed to the activation duration module that proceeds to resolve the LP and obtain a new tentative set of activation times. It is possible that this new set improves on the previous one (that is, it does yield a shorter schedule length). If it does, the process is repeated until no more improvement is achieved and the final activation times are then the ones that are obtained by the last iteration of the process. This is the so-called Column Generation (CG) and its first application to the scheduling problem was presented in [6].

The remaining issue is how to select the new group to be added to the partial LP solution at each step of the iteration. The dual variable values scaled by the rates are used as the metric in the group generation module. There are again two possibilities: Either a heuristic can be used (like the one proposed for the four of the eight algorithms of the previous section, with the only difference that now the dual variable π_i is used to sort the links), or an exact determination of the next “optimal” group based on the dual variable metric. If the latter option is chosen, then there are two further possibilities: either an exhaustive search is performed over all remaining groups or a more efficient determination is done. For the use of the second possibility one needs to know the rate function F . Then, a variety of optimal (yet efficient) searches, from techniques of convex optimization, to branch-and-cut or branch and bound and other methods [4], [5], [11], [12], can be performed to determine the best group capitalizing on the knowledge and properties of the function. If on the other hand only the rate values can be retrieved (but not the rate function that produced them) then the only option for exact determination of the best group at each step is the exhaustive search.

Thus, in the arsenal of the eight previously described algorithms we add two more. The first uses the CG method along with the rank-based heuristic for “next group” selection at each step (we call this the *CG–heuristic* algorithm). The second uses the CG method as well, but with an “exact” selection of the next group at each step. We use the exhaustive search method for that, while we note the possibility of dramatic expansion of the problem sizes (in terms of number of links N) that are possible to solve if we utilize the knowledge of the rate function (we call that the *CG–exact* algorithm).

Note also that even in the earlier eight algorithms the “exact” group generation option can be exercised with significantly reduced complexity if the rate function is known. The only difference there is that the metric for the candidate groups is not the one that depends on the dual variables, but rather the SR or WSR , or possibly even a totally different metric. Of course the use of other metrics, either in association with the LP, or the TF or $T\Delta$ methods, does not in general lead to the minimum-time schedule, while in the case of the *CG–exact* algorithm it generally does. The complexity of the *CG–exact* algorithm though is (in the worst case) exponential.

X. OPTIMALITY PROPERTIES OF THE ALGORITHMS

We present now some properties of the proposed algorithms that strengthen their appropriateness for the solution to the scheduling problem. In the following, we denote an algorithm through the descriptive terms used in the preceding sections. For example, $\langle T\Delta, SR \rangle$ denotes the strategy of selecting the group with maximum sum-rate in every iteration, and the time duration of activation is a constant $\Delta > 0$ (or until the first queue in the group empties, whichever occurs first).

A. Complexity in the Number of Iterations

It is clear that each of the five designs of the activation duration module will empty all queues in a finite number of iterations. Here, one iteration refers to the activation of one group under TF or $T\Delta$, or solving an LP under CG . The running time of one iteration is polynomial in N in all cases. As for the number of iterations to empty all queues, the complexity is summarized in the following theorem.

Theorem 16. *1) The number of iterations under CG –exact is polynomial in N , 2) the number of iterations under $\langle TF, SR \rangle$ and $\langle TF, WSR \rangle$ is $O(N)$, and 3) the number of iterations by $\langle T\Delta, SR \rangle$ and $\langle T\Delta, WSR \rangle$ is $O(N \frac{\max_{i \in \mathcal{N}} d_i}{\Delta r^*})$ and hence pseudo-polynomial in N , where r^* is either a) $\min_{i \in \mathcal{N}} r_{i\mathcal{N}}$ for the case of a continuous rate function, or b) the lowest positive rate level of a discrete rate function.*

Proof: An LP can be solved to optimality using a polynomial number of iterations, or equivalently, separation of constraints in the dual LP, therefore the first statement holds. The second statement follows from the fact that at least one link gets its queue emptied in every iteration in $\langle TF, SR \rangle$ and $\langle TF, WSR \rangle$. The last statement follows

from $\sum_{i \in \mathcal{N}} d_i \leq N \max_{i \in \mathcal{N}} d_i$, and that $\langle T\Delta, SR \rangle$ and $\langle T\Delta, WSR \rangle$ drain at least an amount proportional to Δr^* from one or several queues per iteration. ■

Note that the complexity of *CG-exact* by Theorem 16 does not contradict the general \mathcal{NP} -hardness of the scheduling problem. The reason is that exact group selection is a hard problem in general.

B. Optimality Properties

Among the above designs, *CG-exact* is an exact algorithm that guarantees global optimality [7]. The correspondence to the general method of column generation consists of the fact that a column in the LP (3) corresponds a variable associated with a group.

Let us consider the other four options of Section VIII. In general, they are sub-optimal, but significantly simpler than *CG*. In the following, we show that, if the corresponding sufficient optimality conditions discussed in Section VI strongly hold (i.e., the inequalities in the conditions are strict), the use of *SR* gives the optimal schedule, i.e., the two baseline scheduling solutions \mathcal{H}^1 and \mathcal{H}^N , respectively.

Theorem 17. *Consider $\langle TF, SR \rangle$ and $\langle T\Delta, SR \rangle$ with any $\Delta > 0$. With exact group selection, the two algorithms produce schedules \mathcal{H}^1 and \mathcal{H}^N if Conditions 1 and 4 strictly hold, respectively.*

Proof: Assume Condition 1 is strictly satisfied and, without loss of generality, that the rates are in descending order for individual links, that is, $r_{11} \geq r_{22} \geq \dots \geq r_{NN}$. For any group \mathcal{c} with $|\mathcal{c}| \geq 2$, we have

$$\sum_{i \in \mathcal{c}} \frac{r_{ic}}{r_{ii}} < 1 \Leftrightarrow \sum_{i \in \mathcal{c}} r_{ic} \frac{r_{11}}{r_{ii}} < r_{11} \Rightarrow \sum_{i \in \mathcal{c}} r_{ic} < r_{11}.$$

The last inequality above follows from $\frac{r_{11}}{r_{ii}} \geq 1, i \in \mathcal{c}$. As a result, $\{1\}$ is the group to be activated. For *TF*, the demand of link one is served for time duration d_1/r_{11} , and the next group to be activated with the *SR* metric is $\{2\}$, and so on. Applying *TΔ*, the demand of link one is gradually served by repeatedly activating $\{1\}$, after which $\{2\}$ is used. Hence both strategies lead to schedule \mathcal{H}^1 .

If we assume Condition 4 strongly holds, we first prove that $(m-1)r_{m-1}^{max} < mr_m^{min}, m = 2, \dots, N$. For $m = 2$, $r_1^{max} < 2r_2^{min}$ follows immediately from the assumption. Suppose $(k-1)r_{k-1}^{max} < kr_k^{min}$ holds. Then, Condition 4 written for $k+1$ leads to the following

$$\frac{1}{r_{k+1}^{min}} + \frac{k-1}{kr_k^{max}} < \frac{2}{r_k^{max}}$$

which immediately implies $kr_k^{max} < (k+1)r_{k+1}^{min}$. As a result, the total sum-rate increases for any group by adding new links. Therefore, the grand group \mathcal{N} has the highest sum-rate under *SR*. For both *TF* and *TΔ*, the group will be activated until one of the links' queue becomes empty. Repeating the above argument, the next group to be activated is the one consisting of all links with positive remaining demands, and the theorem follows. ■

The construction of the proof leads to another observation regarding $\langle TF, SR \rangle$ and $\langle T\Delta, SR \rangle$. Under the exact, or a deterministic heuristic, for group selection with the *SR* metric *TΔ* will be activating the same group until one of its links' queue empties, becoming equivalent to *TF*.

Consider again the examples in Section II. For Example 1, Condition 4 applies. By Theorem 17, *SR* leads to \mathcal{H}^N , which is optimal in this case. However, *SR* fails for Example 2, as the highest sum-rate group does not appear at all in the unique optimum. For Example 3, $\langle TF, SR \rangle$ clearly fails. Also, $\langle T\Delta, SR \rangle$ has only little chance of success, as it may work only if the group selection algorithm returns the three two-link groups in a round-robin fashion. On the other hand, the *WSR* metric enables the construction of the optimal schedule for Example 2: $\langle TF, WSR \rangle$ or $\langle T\Delta, WSR \rangle$ for any $\Delta > \frac{4}{15}d$ yields the optimum. For Example 3, $\langle T\Delta, SR \rangle$ delivers the optimum for all $\Delta = \text{frac}d2k\tau_2$ with any positive integer k , and $\langle T\Delta, WSR \rangle$ is optimal for $\Delta = \text{frac}d2u\tau_2$, where u is a positive integer satisfying $u > \frac{\tau_2}{2(2\tau_2 - \tau_1)}$. Thus Examples 2 and 3 illustrate the merit of *WSR* and that of *TΔ*.

XI. SIMULATION SETUP AND RESULTS

In this section we provide simulation results to illustrate the performance of the algorithms developed within the proposed framework. We consider a set of $N = 15$ links randomly placed in an area of 1000×1000 meters.

The signal propagation follows a distance-based law. The distance between the transmitter and the receiver of a link was restricted to be between 3 and 250 meters to obtain links of practically meaningful SNR values. For the queue sizes, we defined two different sets: (i) uniform demand of 1000 bits, (ii) non-uniform demand, uniformly distributed in [100, 1500] bits. In each setup, 100 link location instances were ran, unless otherwise indicated.

For the rate values, we consider two cases, namely (i) rates given by the Shannon formula as in Eq. (2), (ii) rates given by a combination of uncoded BPSK with symbol rate control, at a fixed error rate, as in [27]. In Appendix B we provide a detailed derivation of the BPSK rate formula used, with the standard assumption that the energy of interference from concurrent transmissions is equivalent to Average White Gaussian Noise (AWGN). For illustration purposes we provide both functions in Figure 4 of Appendix B. Note that although both are approximate, with Shannon’s formula giving an upper bound on the achievable rates, and BPSK giving a more practical flavor in our investigation, they jointly provide a useful insight on how the physical layer affects the algorithms’ performance.

For each of the setups described above, we solved the full LP of (3), using AMPL [16], to establish the optimal schedule length, assuming a unit of bandwidth in Hz and an error rate of 10^{-6} for the BPSK rate calculations. Then, all ten algorithms of the previous sections were ran for all instances. In what follows all results are normalized with respect to the global optimal value.

A. The Effect of Activation Duration

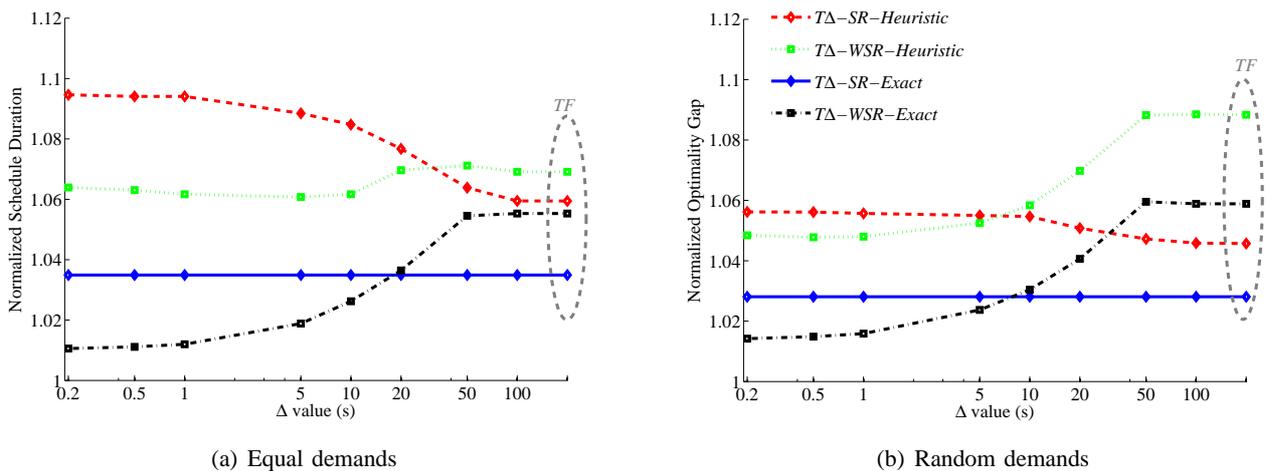


Fig. 1. Effect of activation time Δ on the schedule duration for the four $T\Delta$ -based algorithms, under the Shannon rate function (x-axis in logarithmic scale; each point is an average value over 50 simulation instances).

In Figure 1 we present the effect of the Δ parameter on the performance of the $T\Delta$ -based algorithms. Intuitively, a small Δ enables a more “cautious” design since the algorithms will iterate more times between the two modules, thus having more opportunities to select groups closer to the optimal with more refinement. This is directly confirmed by the green ($T\Delta$ -WSR-Heuristic) and black ($T\Delta$ -WSR-Exact) algorithms, for which the gap increases when Δ grows. Note that for a very large Δ value, the $T\Delta$ strategy would coincide with the TF one, as can be seen in the rightmost of Figure 1. In contrast to $T\Delta$ -WSR-Heuristic and $T\Delta$ -WSR-Exact, the red line for algorithm $T\Delta$ -SR-Heuristic has the opposite trend, as small Δ gives larger optimality gap. This is due to the behavior of the heuristic, namely that it sorts the links by their remaining demand in group generation, even though the sorting may not respond well to the SR metric. The mismatch is however rectified for larger Δ values since the impact of the heuristic will be encountered fewer times. The blue line for algorithm $T\Delta$ -SR-Exact is horizontal, because, with exact group selection and the SR metric, the maximum sum-rate group will remain selected, regardless of the size of Δ , until one of the link queues empties. Thus $T\Delta$ is equivalent to TF, as commented in the discussion after Theorem 17 in Section X.

For $T\Delta$ -WSR-Heuristic and $T\Delta$ -WSR-Exact, the gain of reducing Δ diminishes below some point (~ 0.5 s in Figure 1), since, without significant queue draining, the same group is simply selected over and over again. For the minimum Δ used (i.e., leftmost point in the figure), the best performance is given by $T\Delta$ -WSR-Exact, as also

suggested by Example 2 in Section II. On the rightmost (i.e., TF activation), the WSR metric is inferior to the SR metric, because the former may result in a group with low sum-rate, in the case of very long queues, and the impact cannot be mitigated by group activation which runs the group until one link empties the entire remaining queue. As a result, for large Δ or equivalently TF activation, the best performance is achieved by $T\Delta-SR-Exact$. Finally, as expected, the performance of heuristic selection is consistently inferior to exact selection in Figure 1.

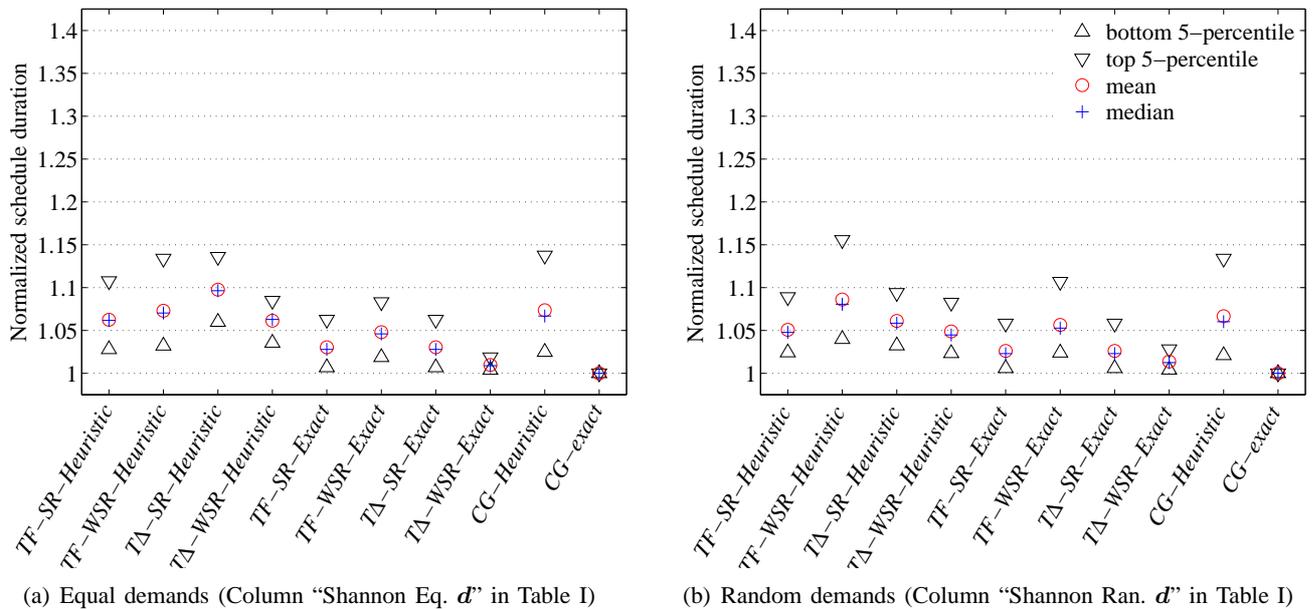


Fig. 2. Normalized schedule durations for all the algorithms under the Shannon-based rate function of Figure 4.

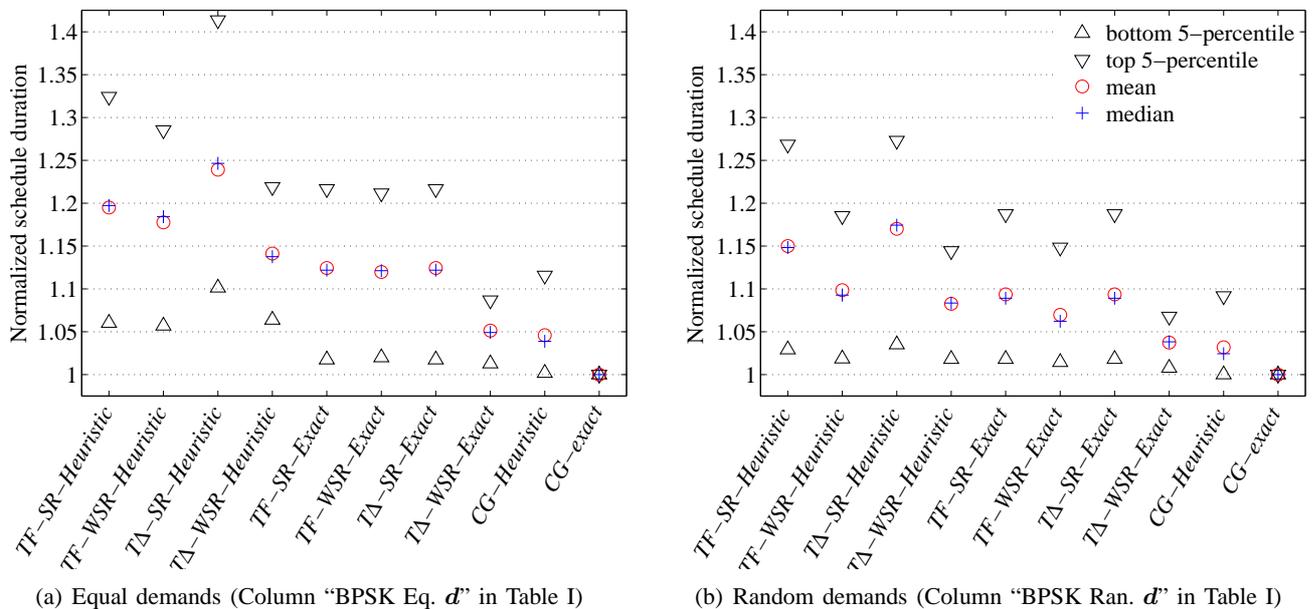


Fig. 3. Normalized schedule durations for all the algorithms under the BPSK-based rate function of Figure 4.

B. Performance Comparison

Figures 2 and 3 provide performance comparison for all the algorithms. For $T\Delta$ -based algorithms, based on our preceding discussion, a small value (0.5s) is chosen for the activation duration parameter Δ . The last algorithm, $CG-exact$, always gives the global optimum, normalized to 1.0 in the figures. Overall, the other nine algorithms

perform reasonably well, with only one giving an average optimality gap larger than 20%. A summary of the performance results, by means of ranking the algorithms by ascending order of average sub-optimality gap, is given in Table I. In addition, the table provides key observations as well as cross-references to this section’s text for more detailed explanation and discussion.

TABLE I
ALGORITHM COMPARISONS.

Alg.	Group Activation	Metric	Group Generation	Shannon Eq. d	Shannon Ran. d	BPSK Eq. d	BPSK Ran. d	Comments
i	TF	SR	heuristic	7	6	9	9	Outperforms ii with Shannon rate function, see 3)
ii	TF	WSR	heuristic	9	10	8	8	–
iii	$T\Delta$	SR	heuristic	10	8	10	10	–
iv	$T\Delta$	WSR	heuristic	6	5	7	5	Best performance among i–iv, see 1)
v	TF	SR	exact	3	3	5	6	Identical to vii, see 4) and Section X
vi	TF	WSR	exact	5	7	4	4	Better performance than v under BPSK rates, see 3)
vii	$T\Delta$	SR	exact	3	3	5	6	–
viii	$T\Delta$	WSR	exact	2	2	3	3	Outperforms vii with small Δ , see 1) and 2)
ix	LP	$dual$	heuristic	8	9	2	2	Good performance under BPSK rates, see 1)
x	LP	$dual$	exact	1	1	1	1	Optimal, high complexity, see Section X.

Notes:

- Exact group selection consistently provides better results than heuristic selection, considering algorithm pairs viii and iv, vii and iii, vi and ii, and v and i. See further discussion in 2).
- $T\Delta$ - WSR -exact (viii) has close-to-optimality performance regardless of the rate function due to Max-Weight Scheduling throughput optimality [32].
- Under Shannon rate function, algorithms i–viii yield better results in comparison to BPSK-based rates. See further discussion in 1).
- Results are inconclusive for the activation duration method, except under the WSR metric. See further discussion in 4).

- 1) *The effect of the rate function:* The first eight algorithms all perform better when the rates are derived from the Shannon formula, in both demand cases. The explanation lies in the shapes of the two rate functions. The BPSK rate is much more robust to interference from concurrent transmissions in comparison to the Shannon function. Hence, at the optimum for the BPSK rate function, large-cardinality groups having similar rates on the links are very likely to be used. With the Shannon formula, smaller groups of higher link rates are more favored in the optimal schedule. Consequently, scheduling with BPSK rate is much more prone to sub-optimality in group selection (as many groups perform similarly for both SR and WSR), and, more importantly, to sub-optimality in group activation (cf. Example 3 in Section II). Thus TF and $T\Delta$ become more sub-optimal for BPSK-derived rates than those derived by the Shannon formula. This conclusion is further supported by the performance of CG -heuristic. In this case, group activation is carried out with the LP , which is the best possible solution of determining the time share among groups, justifying the better performance of this algorithm for the BPSK case, as opposite to the other two heuristic activation strategies.
- 2) *Exact versus heuristic group generation:* For both TF and $T\Delta$, heuristic group generation is always outperformed by the exact one. Hence enhancing the group generation module alone gives noticeable contribution to the overall performance, regardless of the activation strategy. As expected, the impact of sub-optimality in group selection is more striking in BPSK. As was mentioned above, larger groups are expected at optimum for BPSK-derived rates. When the exact solution of group generation contains many elements, it is less probable that greedy selection, as used in our heuristic, is able to approach optimality.
- 3) *The effect of metric for group selection:* Comparing the two metrics SR and WSR , the latter always yields better results for the $T\Delta$ activation strategy, that is, the notion of remaining demand interacts better with emptying queues progressively. This was also observed in the discussion of Figure 1. For BPSK, WSR also outperforms

SR for *TF* activation. The reason is that the optimal schedule with BPSK tends to use groups of similar sizes. With *SR*, there is a higher risk that a small-cardinality group with low sum-rate and high remaining demand will have to be deployed by the end, making the overall schedule inferior in comparison to balancing the remaining queues in group selection. For Shannon-based rates, the structure at optimum is quite the opposite, and *SR* behaves better than *WSR* in *TF*.

- 4) *Activation duration*: From the above two figures, the examples in Section II, and the results in Figure 1, it is inconclusive whether the *TF* or $T\Delta$ activation should be in general preferred. However, if the *WSR* metric is employed in group generation, $T\Delta$ is clearly superior. This is because the group selection metric *WSR* can be seen as a component in the maximum weight scheduling algorithm. Coupling it with the $T\Delta$ activation and exact group generation, we indeed obtain a max-weight scheduling algorithm for a fully deterministic setting. In addition, the two group activation strategies coincide in *TF-SR-Exact* and $T\Delta$ -*SR-Exact*, as justified by our discussion in Section X and the results in Figure 1.
- 5) *Demand distribution*: Finally, the demand structure (uniform versus non-uniform) has a noticeable impact on performance, when heuristic group generation is used. In general, the results show improved performance when the initial demand is non-uniform. This can be attributed to the demand ordering in group construction of the heuristic that we used. Non-uniform demands aid the heuristic to better differentiate among the links, especially in the BPSK case where the effect of non-uniform demand is indeed more prominent, as the links in a group tend to have more similar rates. For the Shannon-based rate function, with exact group generation coupled with *TF* and $T\Delta$, the demand structure has virtually no effect on performance. For BPSK-derived rates, however, non-uniform demand leads to smaller sub-optimality gap. This is because the sub-optimality of *TF* and $T\Delta$ in group activation is more crucial for uniform demand that resembles the structure of Example 3.

XII. CONCLUSION

We have considered the minimum-time scheduling problem for the case of emptying N queues over a shared channel. The generic consideration of rates, which may be produced by some underlying function, unifies the previously considered formulations. Several fundamental results on the solution characterization have been obtained. First, we have confirmed the hardness of the problem for all continuous and monotonically increasing functions in SINR. Second, optimality conditions of two baseline scheduling strategies are developed and formalized. Third, we have demonstrated how the special case of cardinality-based rates can be solved efficiently. On the algorithmic side, we have presented a framework that accommodates both exact and sub-optimal scheduling solutions. Extensive simulation results have been provided and assessed to quantify the performance of some specific algorithm designs.

The research line of the current paper suggests several extensions. For example, we may consider cooperative methods among the links, including relaying each other's messages. In that case we could have multiple transmitters that transmit to the same receiver draining the same queue simultaneously. Another extension is the fundamental solution characterization under a multi-objective setting that incorporates both efficiency (i.e., time) and energy expenditure, or that includes the aspect of fairness among the links. Multi-hop or multicasting applications are also of interest. Last but not least, it is important to use the insights from this work in the problem of scheduling with continuous arrivals (rather than the queue draining problem). Of course the ultimate cross-layer optimization version that chooses both schedules and rates remains looming large as a major challenge.

REFERENCES

- [1] M. Andrews and M. Dinitz. Maximizing capacity in arbitrary wireless networks in the SINR model: complexity and game theory. In *Proc. of IEEE INFOCOM '09*, 2009.
- [2] E. Arikan. Some complexity results about packet radio networks. *IEEE Transactions on Information Theory*, 30:910-918, 1984.
- [3] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali. *Linear Programming and Network Flows*, 2nd edition. John Wiley & Sons, 1990.
- [4] D. Bertsimas, J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.
- [5] D. P. Bertsekas. *Nonlinear Programming: 2nd Edition*. Athena Scientific, 1999.
- [6] P. Björklund, P. Värbrand, and D. Yuan. Resource optimization of spatial TDMA in ad hoc radio networks: a column generation approach. In *Proc. of IEEE INFOCOM '03*, 2003.
- [7] P. Björklund, P. Värbrand, and D. Yuan. A column generation method for spatial TDMA scheduling in ad hoc networks. *Ad Hoc Networks*, 2:405-418, 2004.
- [8] S. A. Borbash and A. Ephremides. Wireless link scheduling with power control and SINR constraints. *IEEE Transactions on Information Theory*, 52:5106-5111, 2006.

- [9] C. Boyacı, B. Li, and Y. Xia. An investigation on the nature of wireless scheduling. In *Proc. of IEEE INFOCOM '10*, 2010.
- [10] A. Capone et al. Resource optimization in multi-radio multi-channel wireless mesh networks. In: S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic (editors), *Mobile Ad Hoc Networking: The Cutting Directions*, 2nd edition, Wiley and IEEE Press, 2013.
- [11] A. Capone et al. A new computational approach for maximum link activation in wireless networks under the SINR model. *IEEE Transactions on Wireless Communications*, 10:1368–1372, 2011.
- [12] D.-S. Chen, R. G. Batson, and Y. Dang. *Applied Integer Programming: Modeling and Solution*. Wiley, 2010.
- [13] P. Chrétienne et al. *Scheduling Theory and Its Applications*, Wiley, 1995.
- [14] J. Eckhoff. Helly, Radon, and Carathéodory type theorems. In P. M. Gruber and J. M. Wills (editors), *Handbook of Convex Geometry*. North Holland, 1993.
- [15] A. Ephremides and S. Verdu. Control and optimization methods in communication network problems. *IEEE Transactions on Automatic Control*, 34:930–942, 1989.
- [16] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Press, 2002.
- [17] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 1(1), 2006.
- [18] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.
- [19] O. Goussevskaia, Y. A. Oswald, and R. Wattenhofer. Complexity in geometric SINR. In *Proc. of ACM MobiHoc '07*, 2007.
- [20] J. Grönkvist, A. Hansson, and J. Nilsson. A comparison of access methods for multihop ad hoc radio networks. *Proc. of IEEE VTC '00*, 2000.
- [21] B. Hajek and G. Sasaki, Link scheduling in polynomial time. *IEEE Transaction on Information Theory*, 34:910–917, 1988.
- [22] H. Hoogeveen, P. Schuurman, and G. J. Woeginger. Non-approximability results for scheduling problems with minsum criteria. *INFORMS Journal on Computing*, 13:157–168, 2001.
- [23] S. Kompella et al. On optimal SINR-based scheduling in multihop wireless networks. *IEEE/ACM Transactions on Networking*, 18:1713–1724, 2010.
- [24] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41:960–981, 1994.
- [25] K. G. Murty. *Linear Programming*. Wiley, 1983.
- [26] A. Pantelidou and A. Ephremides. Minimum schedule lengths with rate control in wireless networks. In *Proc. of IEEE MILCOM '08*, 2008.
- [27] A. Pantelidou and A. Ephremides. Optimal rate control policies for proportional fairness in wireless networks. In *Proc. of Conference on Information Sciences and Systems (CISS)*, 2008.
- [28] A. Pantelidou and A. Ephremides. Scheduling in wireless networks. *Foundations and Trends in Networking*, 4(4), 2009.
- [29] C. G. Prohazka. Decoupling link scheduling constraints in multihop packet radio networks. *IEEE Transactions on Computers*, 38:455–458, 1989.
- [30] S. Ramanathan and E. L. Lloyd. Scheduling algorithms for multihop radio networks. *IEEE Transactions on Networking*, 1:166–177, 1993.
- [31] D. S. Stevens and M. H. Ammar. Evaluation of slot allocation strategies for TDMA protocols in packet radio networks. *Proc. of IEEE MILCOM '90*, 1990.
- [32] L. Tassiulas and A. Ephremides. Stability properties of constrained queuing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37:1936–1948, 1992.
- [33] J. Proakis. *Digital Communications*. McGraw-Hill Science/Engineering/Math, 2000.
- [34] P. J. Wan, et al. Wireless link scheduling under physical interference model. In *Proc. of IEEE INFOCOM '11*, 2011.
- [35] X. Xu and S. Tang. A constant approximation algorithm for link scheduling in arbitrary networks under physical interference model. In *Proc. of ACM FOWANC '09*, 2009.

APPENDIX A

LINEAR PROGRAMMING OPTIMALITY CONDITIONS FOR THE SCHEDULING PROBLEM

For an LP, it is well-known [3], [25] that the Karush-Kuhn-Tucker (KKT) optimality condition is both sufficient and necessary. For the scheduling problem, this optimality characterization boils down to the existence of a pair of solutions \mathbf{T} and $\boldsymbol{\pi}$ for LP (3) and its dual (4), respectively, such that the following three conditions hold:

- 1) *primal feasibility*, that is, \mathbf{T} is a feasible solution of (3);
- 2) *dual feasibility*, that is, $\boldsymbol{\pi}$ is a feasible solution of (4);
- 3) *complementary slackness*, that is, the product of each primal variable and the corresponding dual constraint is zero and vice versa:
 - $T_c(\sum_{i \in \mathcal{C}} r_{ic}\pi_i - 1) = 0, \forall c \in \mathcal{H}$, and
 - $\pi_i(\sum_{c \in \mathcal{H}} r_{ic}T_c - d_i) = 0, \forall i \in \mathcal{N}$.

By the KKT optimality characterization, a solution pair satisfying the above conditions are optimal for (3) and (4), respectively. Note that, because (3b) are equalities, primal feasibility implies the second part of complementary slackness.

The simplex algorithm is a standard method for solving LPs. The algorithm goes through a sequence of solutions, each being a BFS. As was discussed earlier, for (3), a BFS is characterized by N out of the $2^N - 1$ variables. These

N variables are known as basic variables, of which the corresponding columns form a square matrix \mathbf{B} . For the N basic variables, their values are determined by $\mathbf{B}^{-1}\mathbf{d}$. A BFS means also that $\mathbf{B}^{-1}\mathbf{d} \geq 0$, i.e., primal feasibility and hence the second part of complementary slackness. Define dual solution $\boldsymbol{\pi} = \mathbf{e}'\mathbf{B}^{-1}$, where \mathbf{e}' is the vector of ones. For the basic variables, it is easily verified that the first part of complementary slackness is satisfied by the definition of $\boldsymbol{\pi}$. For the non-basic variables, this part of complementary slackness also holds simply because these variables have value zero. What remains is dual feasibility, that is, $1 - \mathbf{e}'\mathbf{B}^{-1}\mathbf{r}_c \geq 0, \forall i \in \mathcal{N}$, where \mathbf{r}_c is the column vector of variable T_c in (3). The left-hand side is known as reduced cost in the simplex algorithm. For the N basic variables, the reduced cost is clearly zero and thus the corresponding dual feasibility condition is fulfilled. This is not the case for all the non-basic variables, unless the solution is optimal. The termination criterion of the simplex algorithm is non-negative reduced cost (for minimization), and this is in fact dual feasibility in the KKT condition. If the criterion is not satisfied, the simplex algorithm will choose a non-basic variable with negative reduced cost, and perform a pivot operation; namely, to bring this variable into the basis and exclude a variable from the basis. Moreover, if the latter is not degenerate, i.e., it has a positive value before pivoting, the pivot operation is guaranteed to strictly improve the objective function value. Applying a sequence of pivoting operations, the simplex algorithm will eventually arrive at a solution satisfying dual feasibility and conclude optimality.

APPENDIX B THE BPSK RATE FUNCTION FOR A MULTI-USER ENVIRONMENT

This is standard material from basic digital communication theory, summarized here for completeness.

Assuming an uncoded BPSK modulation scheme, in an interference-free environment the bit error probability z is given by

$$z = Q\left(\sqrt{\frac{2E_b}{N_0}}\right),$$

where the function $Q(x)$ is the probability that a Gaussian random variable with zero mean and unit variance exceeds x . The $\frac{E_b}{N_0}$ fraction is the system SNR, where the numerator is the bit energy and the denominator is the power spectral density (psd) of the noise, both in Joules [18].

In the presence of interference by concurrent transmissions, denoting by T_0 the duration of one BPSK symbol the bit rate will be $r_b = \frac{1}{T_0}$. The received bit power is also $P_b = \frac{E_b}{T_0}$ in Watts. Then the error rate can be calculated as

$$z = Q\left(\sqrt{\frac{2E_b}{I_N}}\right) = Q\left(\sqrt{\frac{2P_b T_0}{I_N}}\right) = Q\left(\sqrt{\frac{2P_b}{I_N r_b}}\right),$$

where I_N denotes the sum of all interference energy in Joule (which we consider it to be an AWGN signal) plus the noise psd. Notice that I_N can be approximated by scaling the sum of the interference powers received by a time factor, in order to obtain an appropriate quantity in Joules.

Observe now that any change in the I_N value, under a fixed z and fixed P_b , leaves us with the symbol duration T_0 , i.e. the bit rate, as the only control we have to keep the equation above satisfied. Hence, solving the error rate equation above yields our approximation to the BPSK bit rate:

$$r_b = \frac{2}{Q^{-1}(z)^2} \frac{P_b}{I_N},$$

where Q^{-1} is the inverse Q -function.

Note that for general pulses the BPSK symbol rate must satisfy $r_s = B/k$, where k is the spectral efficiency (which for ideal pulses we can assume to be equal to one) and B is the channel bandwidth [18]. Hence, we limit our BPSK rate function results to a maximum rate value $r_b^{max} = r_s = B$, i.e.

$$r_b = \min\left\{\frac{2}{Q^{-1}(z)^2} \frac{P_b}{I_N}, B\right\}.$$

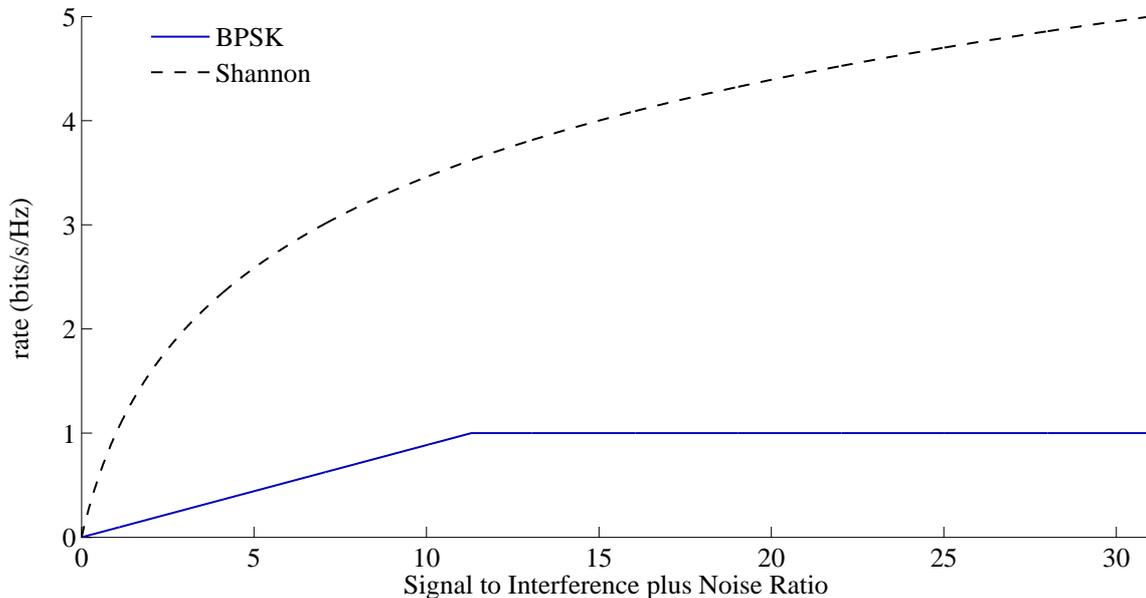


Fig. 4. The Shannon-based and BPSK-based rates, per bandwidth unit, versus the SINR.

Vangelis Angelakis (S05, M09) holds an MSc('04) and a PhD('08) in Computer Science from the University of Crete, Greece. Presently, he is an Assistant Professor in Mobile Telecommunications at the Department of Science and Technology, Linköping University. His current research is centered on Wireless Networks' resource allocation.

Dr Angelakis has been a research assistant at the Institute of Computer Science (ICS) of the Foundation for Research And Technology – Hellas (FORTH), Greece since 2001. In 2005 he was a visiting research associate at the Institute for Systems Research of the University of Maryland in College Park, USA. From 2008 he was a postdoctoral researcher with the Telecommunications and Networks laboratory of FORTH-ICS, until 2009 when he joined Linköping University. Since then he has been a visiting researcher in the UK (RANPLAN Ltd), and China (Huazhong University of Science and Technology). He is an associate editor of the Journal of Communications and has received support from national funding bodies in Greece, Sweden, and the EU within FP7.

Anthony Ephremides (M71, SM77, F84, LF09) holds the Cynthia Kim Professorship of Information Technology at the Electrical and Computer Engineering Department of the University of Maryland in College Park where he is a Distinguished University Professor and has a joint appointment at the Institute for Systems Research, of which he was among the founding members in 1986. He obtained his PhD in Electrical Engineering from Princeton University in 1971 and has been with the University of Maryland ever since.

He is the author of several hundred papers, conference presentations, and patents, and his research interests lie in the areas of Communication Systems and Networks and all related disciplines, such as Information Theory, Control and Optimization, Satellite Systems, Queueing Models, Signal Processing, etc. He is especially interested in Wireless Networks and Energy Efficient Systems.

Qing He (S'12) received her BSc degree in Electronics and Information Systems, and MSc degree in Circuits and Systems from the Department of Electronic Science and Engineering of Nanjing University, China, in 2001 and 2004, respectively. She has a second degree in Finance and had been working as software designer and system engineer in Lucent Technologies Bell Labs, China, until 2010. She is currently working toward the Ph.D degree at the Department of Science and Technology, Linköping University, Sweden. Her current research interests include wireless network optimization and information theory.

Di Yuan received his MSc degree in computer science and engineering, and PhD degree in operations research at Linköping Institute of Technology in 1996 and 2001, respectively. At present he is a full professor in telecommunications at the Department of Science and Technology, Linköping University, and head of a research group in mobile telecommunications. His current research mainly addresses network optimization of 4G systems.

Dr Yuan (M'03) has been guest professor at Technical University of Milan (Politecnico di Milano), Italy, in 2008, and senior visiting scientist at Ranplan Wireless Network Design Ltd, United Kingdom, in 2009 and 2012. In 2011 and 2013 he has been part time with Ericsson Research, Sweden. He is an area editor of the Computer Networks journal. He has been in the management committee of four European Cooperation in field of Scientific and Technical Research (COST) actions, invited lecturer of European Network of Excellence EuroNF, the recipient of a European FP7 Marie Curie IOF career grant award, and Principal Investigator of five European FP7 projects. He is co-recipient of a Best Paper Award in IEEE ICC'12.