Signal Quality Pricing: Decomposition for Spectrum Scheduling and System Configuration

Eric Anderson
Electrical and Computer Engineering
Carnegie Mellon University

Caleb Phillips
Computer Science
University of Colorado

Douglas Sicker
Computer Science
University of Colorado

Dirk Grunwald
Computer Science
University of Colorado

***Abstract*—*Who gets to use radio spectrum, and when, where, and how*? Many problems in traditional radio communication, wireless networking, and cognitive radio are variants of this question. Optimization decomposition based on Lagrangian relaxation of signal quality requirements provides a mathematical framework for solving this type of combined problem. This paper demonstrates the technique as a solution to optimal spatial reuse time-division multiple access (STDMA) scheduling with reconfigurable antennas. The joint beam steering and scheduling (JBSS) problem offers both a challenging mathematical structure and significant practical value.**

We present algorithms for JBSS and describe an implemented system based on these algorithms. We achieve up to 600% of the throughput of TDMA with a mean of 234% in our experiments. The decomposition approach leads to a working distributed protocol which is provably equivalent to our original problem statement while also producing optimal solutions in an amount of time that is *at worst* linear in the size of the input. This is, to the best of our knowledge, the first actually implemented wireless scheduling system based on dual decomposition. We identify and briefly address some of the challenges that arise in taking such a system from theory to reality.

I. INTRODUCTION

Interference is the primary manifestation of spectrum scarcity, and its management is the main technical challenge in realizing efficient dynamic spectrum access (DSA). Most work on spectrum allocation, whether in the context of DSA or scheduling, takes transmitters' and receivers' properties as given and seeks to work around those constraints. Our view is that *how* users access the spectrum is as important as *who* accesses it and *when*: The who and when questions are about dividing up a fixed capacity efficiently and fairly, but the "how" can actively increase that capacity. When users can intelligently minimize the interference they impose on each other – *including on primary users* – more value can be extracted from a fixed amount of resources. This paper presents a framework for jointly optimizing the how with the who and when. We will focus on antenna configuration as a physical configuration parameter because it can add tens of dB isolation between links, and because it presents and especially challenging mathematical structure.

A. Combined Spectrum Scheduling and System Configuration

We define *spectrum scheduling* as assigning users (who can be understood as either transmitters or links) to discrete slots of time in which they may generate radio signals. In general, this is a many-to-many mapping. We define *system configuration* as stipulating the way in which users access the spectrum in each time slot. Each user's transmit power, channel, modulation scheme, and antenna configuration are examples of system configuration variables. For the purposes of this paper, we assume that users are cooperative and follow their assigned schedules and configurations. These assignments can be thought of as upper bounds on how users may affect each other.

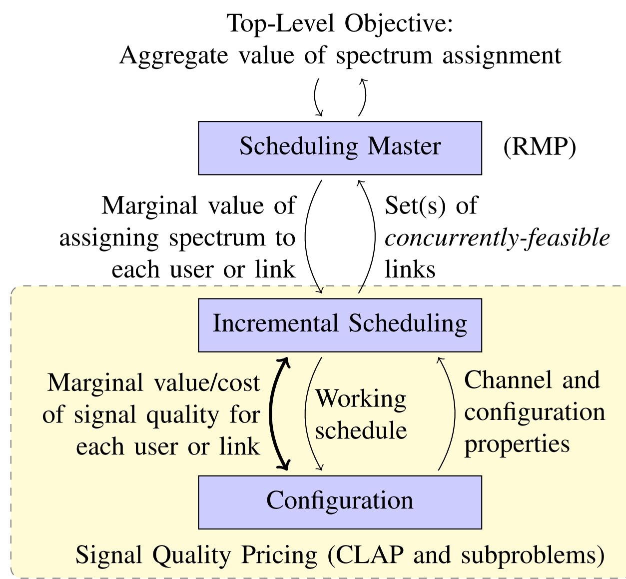


Fig. 1: Problem decomposition model: The lower block shows the decomposition by signal quality pricing.

The combined problem is interesting when the optimal (or feasible) configuration depends on the schedule *and vice-versa*, so that neither problem can be solved independently. In this case, we can define a joint problem and then decompose it into subproblems which are coupled by a Lagrange multiplier which functions as a marginal *value* or *price of signal quality* for each user. This is the model shown in Figure 1.

The intuition behind this approach is simple: A high signal quality price for a given link is an indication to the scheduling process that (given the schedule under consideration) it is difficult to satisfy that link's signal to interference and noise ratio (SINR) requirements, and it might be better to not schedule that link in this time slot. The same high price indicates to the configuration process that the link's signal quality is limiting the overall utility, and it would be good to improve it. By iteratively solving the scheduling and configuration problems, updating the price each time, the system converges to a joint solution.

B. Spatial-Reuse TDMA With Configurable Antennas

This paper demonstrates a novel approach to minimizing interference and maximizing spatial reuse for competing spectrum users. These concerns are significant anytime interference is a limiting factor, including packet radio networking, mobile telephony, and radio repeaters. Here, we specifically consider explicitly scheduled Medium Access Control (MAC) protocols such as Time Division Multiple Access (TDMA). These MACs enable optimizations for spatial reuse and avoid problems that random-access carrier-sense protocols (*e.g.*, CSMA/CA) incur when deployed in large networks where hidden terminal effects limit performance.

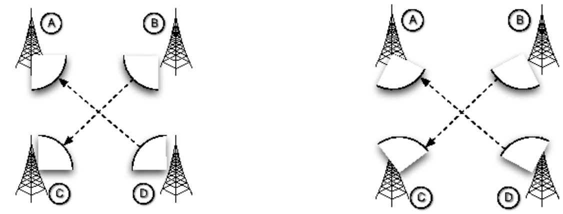
We present a joint optimization process for integrating scheduling and beam steering to achieve greater spatial reuse than is given by solving the two problems separately. Without such coupling between the MAC scheduling and physical antenna configuration processes, a "chicken-and-egg" problem exists: If antenna decisions are made before scheduling, they cannot be optimized for the communication that will actually occur. If the scheduling decisions are made first, the scheduler cannot know what the actual interference and communications properties of the network will be.

Our results show significant gains by integrating scheduling with antenna reconfiguration. An analysis of the performance of our algorithm in simulation shows a mean speedup of 234% with as much as 600% improvement in some scenarios. We also show that simple techniques such as greedy approaches to antenna steering and scheduling result in substantial interference between neighboring links.

Figure 2 illustrates the potential pitfalls of treating scheduling and antenna configuration separately. The best performance case (b) can occur only if the antenna patterns and schedule are chosen jointly – the schedule is impossible with a naïve antenna choice (a), and there's no reason for the better antenna patterns to be chosen unless that schedule is being considered.

To understand the effects of this phenomenon on a real network, we conducted an empirical study using a wide-area phased array testbed of seven nodes [1]. Considering all feasible two-link transmission sets (*e.g.* $\{A \rightarrow B, C \rightarrow D\}$ with each link using its independent best (greedy) antenna patterns, we find significant inter-link interference. The distribution of observed signal to interference ratios (SIRs) is shown in fig. 3. The reference lines mark 10.5 and 26.5 dB, which are theoretical signal to noise (SNR) thresholds[1] to achieve a bit error rate (BER) of 10^{-6} using two common modulation schemes, BPSK and 64 QAM [2]. Pairwise interference is sufficient to preclude BPSK and 64 QAM at this BER in 28% and 74% of cases, respectively.

[1]These SNR thresholds are roughly comparable with SIR numbers, if the interfering signal is close to Gaussian noise and other sources of noise and interference are negligible.



(a) Nodes have their beam pattern main lobes pointed directly at their communicating partner.

(b) Scheduling-aware antenna configuration: Beam patterns chosen *to enable* a denser schedule.

Fig. 2: Example: Links B to C and D to A can be scheduled concurrently, but not with greedy antenna configurations.

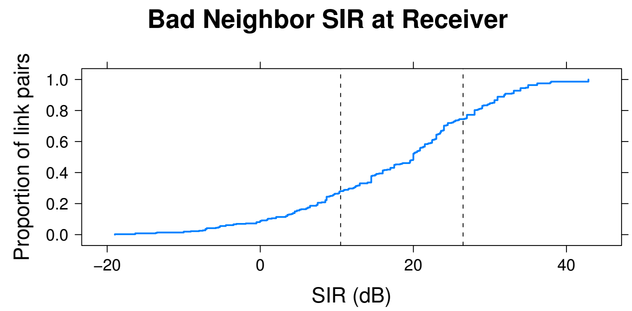


Fig. 3: Interference between neighboring links when greedy antenna patterns are used. Reference lines show theoretical SNR values for 10^{-6} BER with BPSK (10.5 dB) and 64-QAM (26.5 dB) modulation schemes.

Based on the difficulties described, we observe that naïve approaches have deficiencies which can be addressed with a more complete, integrated solution. This motivates our proposal here, which is capable of finding system-wide optimal solutions in a reasonable amount of time.

In the following section, we discuss the background and related work. In §III, we present our formulation along with a series of decompositions which transform this problem into a tractable form. Section IV evaluates this algorithm via numerical experiments, showing that optimal solutions are both achieved quickly and offer substantial speedup over (non-spatial-reuse) TDMA schedules. In §V we discuss a testbed proof-of-concept implementation of our approach, and finally in §VI we summarize our contributions and conclude.

II. Related Work

There are two main areas of closely-related work which bear discussing: Optimization-based wireless scheduling, and wireless networking with directional antennas. At the intersection of the two, there are several proposals which consider antennas in the context of scheduling, but none which do so with significant integration or optimality results.

A. Optimization and Wireless Scheduling

There is a significant body of theoretical work in the area of optimization and wireless scheduling, although ours is one of the first to produce an implementation. We identify a few salient examples here. The principle optimization foundations were laid by Arikan's formulation of $\vec{f}$-feasibility and Toumpis and Goldsmith's analysis of capacity regions [3], [4]. The first explicit treatment of utility maximization and its dual problems in networking is Kelly's work on rate control [5].

Björklund *et al.* introduced the first optimization formulation of wireless scheduling of which we are aware [6]. The authors present a linear column generation formulation. The paper compares the complexity and efficacy of scheduling by link and by node (transmitter), re-establishes NP-completeness results for both problems, and compares an integer formulation with its continuous relaxation. This paper's formulation is the starting point for the present paper.

Xiao *et al.* present the Simultaneous Routing and Resource Allocation (SRRA) problem, which is a joint optimization approach to routing and something similar to scheduling [7]. The authors make, and acknowledge, the assumption that link capacities can be determined completely by sender-local decisions. While this abstracted view does not correspond with any real system, it enables a very clean and logical development of techniques central to multi-layer optimization in networks. This paper presents hierarchical dual decomposition using subgradient solution methods, and the coupling of routing and scheduling by per-node capacity prices.

The general principles of the preceding are further explored in a series of papers by Chiang *et al.* under the moniker of "layering as optimization decomposition" [8]–[10]. These address wireless scheduling specifically, and develop the broader notion of network layers as *computational elements* coupled together to solve some global objective, whether by design or by accident. Of particular import for scheduling is work by Tan *et al.* which shows that many non-convex functions of interest, such as interference-limited Shannon capacity, are log-convex when transfered to the logarithmic domain, and therefore admit equivalent convex formulations [11].

There has also been work on *unscheduled* spectrum allocation, in which the goal is to identify a single set of concurrently-operating users, rather than a time-dependent sequence of such sets [12], [13]. This problem is closely related to finding concurrent link sets in Danzig-Wolfe decomposed scheduling (§ III-D), and we expect that ideas developed in either context will be useful in the other.

B. Scheduling with Antenna Considerations

The remaining related work can be divided into two groups: those that do not consider antenna configuration directly and those that do consider antenna configuration, but separately from scheduling.

The first group of papers assume idealized high-level effects of using directional antennas, rather than deal with the actual RF gains of specific antenna configurations. Such approaches are computationally much easier, but the assumptions are often incorrect. Cain *et al.* assume that an arbitrarily narrow beam width allows interference to be disregarded entirely, and propose scheduling then based on the (only) constraint that each node may have at most one link active at a time [14]. Several other papers replace the geometric circular interference region model with a "pie wedge" version [15], [16]. Sundaresan *et al.* consider real signal strength and interference, but assume that a smart antenna can completely eliminate interference from a given number of stations. Scheduling is then as with fixed antennas, with the addition of choosing a set of interferers to disregard [17]. This is conceptually the closest work to the problem we are addressing, but its assumptions are typically false: A K-element phased array antenna has $K-1$ "degrees of freedom," but they are not arbitrary. The signal strength can only be varied independently in $K-1$ directions if they correspond to mutually-orthogonal antenna vectors, which is in general not the case [18, §10.1]. All of the preceding papers are based on simplifying assumptions which do not hold well in practice.

The second group of papers considers actual antenna and radio effects, but with the configuration determined separately from scheduling, as in the example in Fig. 2a on the preceding page. A series of papers by Sánchez-Garache and Dyberg investigates scheduling with the assumption that the stations in every link beam-form toward each other [19], [20].

Recent work by Liu *et al.* considers partial integration of antenna selection and scheduling, using a conflict graph model based on pairwise interference [21], [22]. This simplification leads to a significant loss of optimality, but enables simple and efficiently implementable protocols. Jorswieck *et al.* present an analytical characterization of the potential benefit of beamforming in a given group of concurrent users based on their channel correlation properties, but do not propose any scheduling process to take advantage of this [23].

III. Model and Algorithms

In this section, we describe a distributed and decomposed mathematical solution to the integrated beam steering and scheduling problem. We begin by presenting a formalization of the objective and constraints. Following from this, we present a series of decompositions to make the problem more computationally tractable. Figure 4 on the next page outlines these decompositions. The basic formulation is the Joint Beam Steering and Scheduling Master Problem (JBSS-MP). This is decomposed into a Restricted Master Problem (RMP) and the Configuration and Link Activation Problem (CLAP). After working through two intermediate forms, CLAP is decomposed into the Fixed-link Antenna Reconfiguration Program (FARP), and the Relaxed Primal Fixed-antenna Link Activation Problem (RP-FLAP). FARP and RP-FLAP are separable, and are split into distributed, per-node versions, the Single Node Antenna Reconfiguration Problem (SNARP) and Single Node Relaxed Primal FLAP (SNRP-FLAP). Lastly, we transform SNRP-FLAP to reduce oscillations, producing what we call the Single-node Dual Quadratic FLAP (SDQ-FLAP).

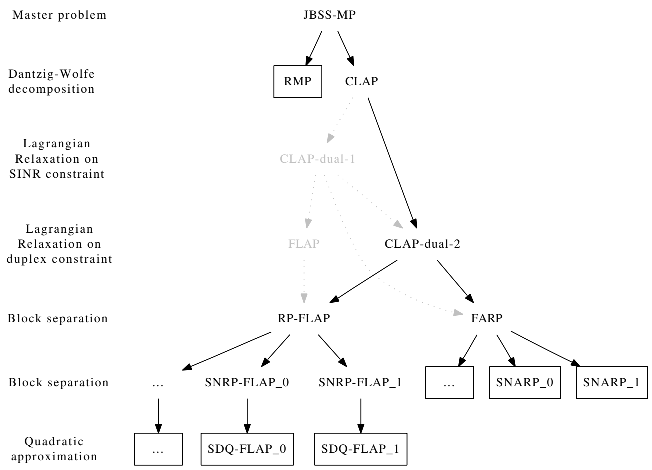


Fig. 4: Problem Decompositions

First, some remarks on notation. Variables and constants are vectors or matrices except where otherwise noted; vectors are regarded as column vectors. Indexing is indicated with subscripts, more than 2 subscripts indicating a multidimensional array. An undecorated variable is a decision variable in the problem at hand, while a bar (*e.g.* $\bar{S}$) indicates an estimate, especially one which is a constant in any given context. The hat (*e.g.* $\hat{S}$) indicates the estimates used in primal solution extraction. The meanings of repeatedly-used symbols are given below. Many of these symbols are logically indexed by the link set l, but outside of the master problems JBSS-MP and RMP, only one link set – the one currently being computed – is considered at any given time, so the subscript l is omitted for simplicity.

symbol	interpretation
A	The set of all links
N	The set of all nodes
L_A	The set of all concurrently-feasible link sets
L_A^t	The generated subset of L_A at time t
x_l	Number of slots assigned to link set $l \in L_A$
q_{ij}	Demand (in slots) for link ij
S_{ij}	Activation of link ij (in current link set)
M_{ij}	Constant s.t. ineq. (16) holds when $S_{ij} = 0$
V_i	Node i is active (in current link set)
P_i	Transmit power of node i
γ_1	Desired SINR threshold
N_r	Receiver noise level
D_{ij}	Directivity of node i in the direction of node j
$Lb(i,j)$	Path loss from node i to node j
G_{ikp}	Gain for node i using pattern p, toward node k
B_{jp}	Beam (antenna, pattern, ...) p used at node j

TABLE I: Notation

A. Formulation

A direct statement of the integrated scheduling and antenna configuration process is given in (JBSS-MP). The objective, (1), is to minimize the time allocated across all link sets. For

[JSBS-MP]

$$\min_{x_l} \quad \sum_{l \in L_A} x_l \tag{1}$$

$$\text{s.t.} \quad \sum_{l \in L_A} S_{ijl} x_l \geq q_{ij} \quad \forall_{i,j} \tag{2}$$

$$\sum_{j:(i,j)\in A} S_{ijl} + \sum_{j:(j,i)\in A} S_{jil} \leq 1 \quad \forall_{i,l} \tag{3}$$

$$\left.\begin{array}{r} \dfrac{P_{il} D_{ijl} D_{jil}}{Lb(i,j) N_r} S_{ijl} \;+ \\ \gamma_1 (1 + M_{ijl})(1 - S_{ijl}) \;\geq \\ \gamma_1 \left(1 + \displaystyle\sum_{k \in N \setminus \{i,j\}} \dfrac{P_{kl} D_{kjl} D_{jkl}}{Lb(k,j) N_r} V_{kl}\right) \end{array}\right\} \quad \forall_{i,j,l} \tag{4}$$

$$S_{ijl} \leq V_{il} \quad \forall_{i,j,l} \tag{5}$$

$$\sum_{p \in P} B_{jpl} = 1 \quad \forall_{j,l} \tag{6}$$

$$D_{ik} = \sum_{p \in P} G_{ikp} B_{ipl} \quad \forall_{i,k,l} \tag{7}$$

$$x_l \geq 0 \quad \forall_{l \in L_A} \tag{8}$$

$$S_{ijl},\ B_{jpl} \in \{0,1\} \tag{9}$$

each link set l in the universe of possible concurrent link sets L_A, x_l is a variable indicating the amount of time for which l is active. Note that at times we overload l to denote instead an index referring to the link set. Constraint (2) specifies that the schedule must "cover" the demand. S_{ijl} is a boolean variable indicating whether link ij is active in link set l, and q_{ij} is the demand for link ij, measured in time. The constraint therefore requires that the total time for which link sets containing ij are activated is sufficient. Constraint (3) specifies that in any given link set l, every node j may be active in *at most* one link. This precludes duplex operation, as well as transmitting to or receiving from multiple partners.

Constraint (4) specifies that minimum SINR requirements are met, taking antenna configuration into account. The formulation of this constraint is patterned after Björklund, and can be somewhat unintuitive. See [24, Chapter 3, eq. (3.12), and Appendix B].

Ignoring the second term, the constraint specifies that if the link ij is used, the received signal strength must exceed the combined interference and noise level at j by factor γ_1. The first term is a product of 0-1 variable S_{ijl}, and the second term is a product of $(1 - S_{ijl})$; the second term ensures that the constraint is satisfied when $S_{ijl} = 0$: The constraint is effectively a no-op when the link ij is not selected. For any given ij, when $S_{ijl} = 1$, the constraint reduces to inequality (10) below. Considering a given link set l, the subscripts can be removed for clarity.

$$\frac{P_i D_{ij} D_{ji}}{Lb(i,j) N_r} S_{ij} \;\geq\; \gamma_1 \left(1 + \sum_{k \in N \setminus \{i,j\}} \frac{P_k D_{kj} D_{jk}}{Lb(k,j) N_r} V_k \right) \quad (10)$$

The left-hand side gives the received SNR in linear units: P_i is node i's transmit power. D_{ij} and D_{ji} are the directional gains of node i and node j toward each other. $Lb(i,j)$ is the path loss between nodes i and j, and N_r is the receiver noise figure. While we use a single N_r for all nodes, having different per-node noise figures does not change the complexity of the solution. The right-hand side is the sum of the contribution above the noise floor of received interfering signals plus 1. The 0-1 variable V_k specifies that node k is (or may be) transmitting in the given time slot.

Constraint (5) couples the decision variables S_{ijl} and V_{il} so that if any link ij is selected, the variable V_{il} reflects that i is transmitting. The V variable is used in (4) to identify sources of interference. The 0-1 variable B_{jpl} indicates whether node j uses beam pattern p in link set l. Constraint (6) specifies that each node must select gain which correspond to a convex combination of its beam patterns. When $B_{jpl} \in \{0,1\}$, the gain must correspond to exactly one pattern. Constraint (7) couples the otherwise free directional gain variables D_{ikl} to the choice of antenna beam B_{ipl}. Constraints (8) and (9) specify positivity and 0-1 requirements for variables.

B. Extensions

The joint beam steering and scheduling problem generalizes several other joint scheduling problems. In particular, transmit power and receiver sensitivity control are achieved by relaxing constraint (6) to allow fractional antenna gain as below, which is mathematically equivalent:

$$\sum_{p \in P} B_{jpl} \leq 1 \quad \forall_{j,l}$$

Additionally, selection from a finite set of modulation schemes is achieved by considering multiple "logical" links for each physical link, with a different $\gamma_{1,r}$ for each rate r and adding a rate constant R_r to constraint (2) as follows:

$$\sum_{l \in L_A} S_{ijlr} R_r x_l \geq q_{ij} \quad \forall_{i,j}$$

Neither extension is considered further in this paper, but both are consistent with the decompositions presented. Adding rate selection is computationally similar to increasing the number of links, while adding power control does not increase the complexity at all.

C. Computational Complexity

The master problem (JBSS-MP) is complete, but a direct solution is computationally intractable. First, the program is mixed-integer cubic, meaning that the objective or constraints involve polynomials of degree 3 and a mixture of continuous and integer variables. There are a number of efficient algorithms for solving linear and quadratic programs, but cubic programs are as difficult as arbitrary non-linear programs. There is no obvious way to reformulate the cubic terms ($D_{ijl} D_{jil} S_{ijl}$ and $D_{kjl} D_{jkl} V_{kl}$) away, as they are the fundamental determinants of SINR and are all real decision variables. The size of the problem is also vast. The subscript l indexes the set of all possible sets of links L_A, having dimension 2^m for m links. Several of the variables are indexed over $L_A \times N \times N$, meaning there are $\Theta(n^2 2^m)$ variables and similarly many constraints.

D. Decompositions

The first transformation we apply is *Danzig-Wolfe decomposition*. This allows the solution technique of *delayed column generation*, or *implicit enumeration* (a term we prefer). This decomposition is used in many previous scheduling works, including [6], [7], [25], [26]. The objective function of JBSS-MP (equation (1) on the preceding page) is quite simple; the complexity lies in defining the region of feasible values. Any set of feasible points defines a convex hull which is a subset of the feasible region. Therefore, given any set of feasible points, the original problem can be replaced with a *restricted master problem* (RMP), in which the only constraint is that the solution must lie within the polytope defined by those points. For a simple objective function and any modest number of such points, the RMP is a computationally simple conservative approximation of the full master problem. The quality of the approximation depends on the how closely this polytope approximates the true feasible region in the area of the master problem's optimal solution. The scheme of implicit enumeration proceeds by iteratively solving the RMP and a sub-problem which searches for additional *objective-improving* feasible points to extend the polytope. If no such points exist, then the approximating polytope matches the true constraint region *at the optimal point* and therefore the solution to the RMP is the optimal solution to the master problem.

Applying this decomposition to JBSS-MP produces the restricted master problem (RMP) and a subproblem which we designate the Configuration and Link Activation Problem (CLAP), shown on the next page. The time allocated to each feasible point l is denoted x_l, and the activation level of each link i,j in each l – an output of the subproblem, not a decision variable here – is denoted $\bar{S}_{ijl}$. The set of feasible points defining the RMP's approximation polytope in iteration t is L_A^t.

[RMP]

$$\min_{x_l} \quad \sum_{l \in L_A^t} x_l \quad (11)$$

$$\text{s.t.} \quad \sum_{l \in L_A^t} \bar{S}_{ijl} x_l \geq q_{ij} \quad \forall_{ij} \quad (12)$$

$$x_l \geq 0 \quad \forall_{l \in L_A^t} \quad (13)$$

The process of solving RMP produces not only a primal solution $X^*_{RMP_t}$ but also dual costs for the constraints (12), $\bar{\beta}_{ij}$. These are inputs to the subproblem CLAP, which produces a set of concurrently feasible links indicated by S and associated antenna gains and configurations D and B. In the context of Figure 1, CLAP is *incremental scheduling* and *configuration* and RMP is the *scheduling master*.

[CLAP]

$$\max_{S} \quad \bar{\beta}^T S \tag{14}$$

$$\text{s.t.} \quad \sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} S_{ji} \le 1 \qquad \forall_i \tag{15}$$

$$\left.\begin{array}{r} \frac{P_i D_{ij} D_{ji}}{Lb(i,j)N_r} S_{ij} \;\; + \\ \gamma_1(1+M_{ij})(1-S_{ij}) \;\; \ge \\ \gamma_1\left(1 + \sum_{k\in N\setminus\{i,j\}} \frac{P_k D_{kj} D_{jk}}{Lb(k,j)N_r} V_k\right) \end{array}\right\} \quad \forall_{i,j} \tag{16}$$

$$S_{ij} \le V_i \qquad \forall_i \tag{17}$$

$$D_{ik} = \sum_{p\in P} G_{ikp} B_{ip} \quad \forall_{i,k} \tag{18}$$

$$\sum_{p\in P} B_{jp} = 1 \qquad \forall_j \tag{19}$$

$$S_{ij},\ B_{jp} \in \{0,1\} \tag{20}$$

The problem RMP is trivial, but CLAP retains most of the original complexity of JBSS-MP. Crucially, however, it is no longer dimensioned over the set of all possible sets of links: For n nodes, the number of variables and the number of constraints are both $\Theta(n^2)$. The primary computational difficulty in CLAP comes from constraint (16) which is still order 3 and mixed-integer. Let us define vector-valued convenience function $d^s(\cdot)$, entry ij of which is given by:

$$\begin{aligned} d^s(S,D,V)_{ij} = -\Big(&\frac{P_i D_{ij} D_{ji}}{Lb(i,j)N_r} S_{ij} + \\ &\gamma_1(1+M_{ij})(1-S_{ij}) - \\ &\gamma_1\big(1 + \sum_{k\in N\setminus\{i,j\}} \frac{P_k D_{kj} D_{kj}}{Lb(k,j)N_r} V_k\big)\Big) \end{aligned} \tag{21}$$

Note that constraint (16) is then $d^s(S,D,V) \le 0$. Let the Lagrangian function with regard to (14) and (16) be:

$$\mathcal{L}(S,\lambda) = \bar{\beta}^T S - \lambda^T d^s(S,D,V) \tag{22}$$

This gives a dual function

$$\phi(\lambda) = \max_{S,D,V} \ \mathcal{L}(S,D,V,\lambda) \tag{23}$$

The corresponding Lagrangian dual problem is CLAP-dual-1. The resulting Lagrangian relaxed primal problem (RPP) of CLAP is given below, where $\bar{\lambda}$ denotes an *estimate* of the optimal multipliers λ^*.

$$\begin{aligned} \max_{S,D,V} \quad & \bar{\beta}^T S + \bar{\lambda}^T d^s(S,D,V) \\ \text{s.t.} \quad & \text{constraints } (15) - (19) \\ & \textbf{except } (16). \end{aligned} \tag{24}$$

[CLAP-dual-11]

$$\begin{aligned} \min_{\lambda} \quad & \phi(\lambda) \\ \text{s.t.} \quad & \sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} S_{ji} \le 1 && \forall_i \\ & S_{ij} \le V_i && \forall_i \\ & D_{ik} = \sum_{p\in P} G_{ikp} B_{ip} && \forall_{i,k} \\ & \sum_{p\in P} B_{jp} = 1 && \forall_j \\ & S_{ij},\ B_{jp} \in \{0,1\} && \forall i,j,p \end{aligned}$$

This RPP is block-structured and separable into two subproblems coupled by the Lagrange multipliers λ. These problems correspond to the *Incremental Scheduling* and *Configuration* tasks in Figure1. We label these the Fixed-antenna Link Activation Problem (FLAP) and the Fixed-link Antenna Reconfiguration Problem (FARP) respectively. FLAP takes estimated Lagrange multipliers and antenna gains $\bar{\lambda}$, $\bar{D}$ as parameters and computes link activations S. Conversely, FARP takes $\bar{\lambda}$ and estimated link activations $\bar{S}$ as parameters and computes antenna gains D:

[FLAP]

$$\begin{aligned} \max_{S,V} \quad & \left\{ \begin{array}{l} \bar{\beta}^T S - \sum_{ij} \bar{\lambda}_{ij} \Big(\frac{P_i \bar{D}_{ij} \bar{D}_{ji}}{Lb(i,j)N_r} S_{ij} + \\ \quad \gamma_1(1+M_{ij})(1-S_{ij}) \ - \\ \gamma_1\Big(1 + \sum_{k\in N\setminus\{i,j\}} \frac{P_k \bar{D}_{kj} \bar{D}_{kj}}{Lb(k,j)N_r} V_k\Big)\Big) \end{array} \right. \\ \text{s.t.} \quad & \sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} S_{ij} \le 1 && \forall_i \\ & S_{ij} \le V_i && \forall_{i,j} \\ & S_{ij} \in 0,1 && \forall i,j \end{aligned}$$

This problem has the integrality property, and so the constraint $S_{ij} \in 0,1 \ \forall i,j$ can be dropped.

Proposition 3.1: The continuous relaxation of FLAP is equivalent to FLAP with integer S.

Proof: The constraint matrix of continuous FLAP is totally unimodular by Ghouila-Houri's Theorem. Therefore every extreme point of the feasible polytope is in $\mathbb{Z}^n$. The function to be maximized is concave, implying that no maximum occurs within the feasible polytope, and therefore that the constrained optimum occurs at an extreme point. Therefore the integer optimum and continuous optimum occur at the same point. The inverse of the constraint matrix is also totally unimodular by Cramer's Rule, and so the same argument holds for the dual problem. ■

[FARP]

$$\max_{D,B} \begin{cases} \bar{\beta}^T\bar{S} - \sum_{ij}\bar{\lambda}_{ij}\Big(\frac{P_i D_{ij} D_{ji}}{Lb(i,j)N_r}\bar{S}_{ij} + \\ \gamma_1(1+M_{ij})(1-\bar{S}_{ij}) \; - \\ \gamma_1\Big(1+\sum_{k\in N\setminus\{i,j\}}\frac{P_k D_{kj} D_{kj}}{Lb(k,j)N_r}\bar{V}_k\Big)\Big) \end{cases}$$

$$\text{s.t.} \quad D_{ik} - \sum_{p\in P} G_{ikp}B_{ip} = 0 \quad \forall_{i,k}$$

$$\sum_{p\in P} B_{ip} = 1 \quad \forall_i$$

Note that $\bar{\beta}^T\bar{S}$ is a constant and is dropped for simplicity in subsequent formulations. The constraints are easily separable by index i. The objective function is also separable, but it is slightly less obvious. To identify the separability, we will introduce the following notation. Let x denote the vector of all antenna gains D. Now let i partition x as: $x_i = \cup_{k\neq i} D_{ik}$.

$$g_i(x) = \begin{cases} \sum_j \left(\frac{1}{2}\bar{\lambda}_{ij}\bar{S}_{ij}\frac{P_i}{Lb(i,j)N_r}D_{ij}\bar{D}_{ji}\right) + \frac{k}{|N|} \\ \quad \text{if } i \text{ is a transmitter} \\ \sum_j \left(\frac{1}{2}\bar{\lambda}_{ji}\bar{S}_{ji}\frac{P_j}{Lb(j,i)N_r}\bar{D}_{ji}D_{ij}\right) + \frac{k}{|N|} \\ \quad \text{if } i \text{ is a receiver} \end{cases}$$

$$h_i(x) = \begin{cases} \sum_j \Big(\sum_{k,l\in N\setminus\{i,j\}}\Big(\frac{1}{2}\gamma_1\bar{S}_{ij}\bar{\lambda}_{kl}\frac{P_i}{Lb(i,l)N_r}D_{il}\bar{D}_{li}\Big)\Big) \\ \quad \text{if } i \text{ is a transmitter} \\ \sum_j \Big(\sum_{k,l\in N\setminus\{i,j\}}\Big(\frac{1}{2}\gamma_1\bar{S}_{ji}\bar{\lambda}_{ji}\frac{P_k}{Lb(k,i)N_r}\bar{D}_{ki}D_{ik}\Big)\Big) \\ \quad \text{if } i \text{ is a receiver} \end{cases}$$

$$f_i(x) = g_i(x_i) - h_i(x)$$

$$f(x) = \sum_i f_i(x) \text{ given } \sum_j \bar{S}_{ij} \leq V_i \quad \forall_i$$

Now, $f(x)$ is a re-arrangement of the inverse of the FARP objective, and is clearly separable by index i. Using this separation, we define an instance of the Single Node Antenna Reconfiguration Problem (SNARP) for every node in the network.

[SNARP$_i$]

$$\max_{D,B} \quad 1 - f_i(D) \tag{25a}$$

$$\text{s.t.} \quad D_{ik} - \sum_{p\in P} G_{ikp}B_{ip} = 0 \quad \forall_k \tag{25b}$$

$$\sum_{p\in P} B_{ip} = 1 \tag{25c}$$

$$B_{ip} \leq 1 \quad \forall_{p\in P} \tag{25d}$$

$$B_{ip} \geq 0 \quad \forall_{p\in P} \tag{25e}$$

Proposition 3.2: SNARP$_i$ with continuous variables has an optimal solution equal to that with boolean B_{ip}.

Proof: SNARP$_i$ is a linear program in D, B, but can be re-written purely in B by substituting $\sum_{p\in P} G_{ikp}B_{ip}$ for D_{ik} in the objective function. So written, it is a linear program with $|P|$ variables and $2|P|+1$ constraints. By the fundamental theorem of linear programming [27, Theorem 3.4] $\exists$ a basic solution in which $|P|$ constraints are satisfied with equality. Constraint (25c) must be one of them. This forces $|P|-1$ out of (25d), (25e) to be satisfied with equality, which means that $|P|-1$ of the variables must be either 0 or 1. Those variables must then sum to either 0 or 1, based on (25c). Those options force the remaining variable to be 1 or 0, respectively, in order to satisfy (25c). ■

In the interest of scalability, it would be desirable to similarly separate FLAP. Unfortunately the duplex constraint prevents this, and is not easily massaged away algebraically. To address this, we extend the Lagrangian relaxation of CLAP to the constraint $\sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} S_{ji} \leq 1 \; \forall_i$. Paralleling equation (21) on the preceding page, let $d^d(S)$ be the function having the i-th element given be equation (26):

$$d^d(S)_i = \sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} S_{ji} - 1 \tag{26}$$

Let $d'^s(S,V)$ be $d^s(S,D,V)$ where the antenna gain variables D are replaced with fixed estimates $\bar{D}$, where element ij is given by:

$$d'^s(S,V)_{ij} = -\Big(\frac{P_i\bar{D}_{ij}\bar{D}_{ji}}{Lb(i,j)N_r}S_{ij} + \gamma_1(1+M_{ij})(1-S_{ij}) - \gamma_1\big(1+\sum_{k\in N\setminus\{i,j\}}\frac{P_k\bar{D}_{kj}\bar{D}_{jk}}{Lb(k,j)N_r}V_k\big)\Big) \tag{27}$$

Let us define a new Lagrangian function $\mathcal{L}'(\cdot)$ as follows:

$$\mathcal{L}'(S,D,V,\lambda,\mu) = \bar{\beta}^T S - \lambda^T d^s(S,D,V) - \mu^T d^d(S) \tag{28}$$

This gives a new dual function $\phi'(\lambda,\mu)$ below and corresponding problem dual problem CLAP-dual-2:

$$\phi'(\lambda,\mu) = \max_{S,D,V} \mathcal{L}'(S,D,V,\lambda,\mu)$$

This produces a new relaxed primal version of FLAP, RP-FLAP. FARP remains unchanged.

[RP-FLAP]

$$\max_{S,V} \quad \bar{\beta}^T S + \bar{\lambda}^T d'_s(S,V) - \bar{\mu}^T d^d(S)$$
$$\text{s.t.} \quad S_{ij} \leq V_i \quad \forall_{ij} \tag{30}$$

[CLAP-dual-2]

$$\begin{aligned} \min_{\lambda,\mu} \quad & \phi'(\lambda,\mu) \\ \text{s.t.} \quad & S_{ij} \le V_i \quad \forall_i \\ & D_{ik} = \sum_{p\in P} G_{ikp} B_{ip} \quad \forall_{i,k} \\ & \sum_{p\in P} B_{jp} = 1 \quad \forall_j \end{aligned} \tag{29}$$

RP-FLAP is separable along the index i. We group the link ij with node i, defining d'^d in equation (31):

$$d'^d(S)_i = \sum_{j:(i,j)\in A} S_{ij} + \sum_{j:(j,i)\in A} \bar{S}_{ji} - 1 \tag{31}$$

Using the preceding definition, we define the following:

$$\begin{aligned} \bar{\beta}_w &= \{\bar{\beta}_{ij} | i = w\} \\ \bar{\lambda}_w &= \{\bar{\lambda}_{ij} | i = w\} \\ \bar{\mu}_w &= \{\bar{\mu}_i | i = w\} \\ d'^d_w(S) &= \{d'^d(S)_i | i = w\} \\ S_w &= \{S_{ij} | i = w\} \\ d'^s_w(S_w, V) &= \{d'^s_w(S,V)_{ij} | i = w\} \end{aligned}$$

The partitioned form of RP-FLAP is the Single Node Relaxed Primal FLAP (SNRP-FLAP$_w$) for each index w:

[SNRP-FLAP$_w$]

$$\max_S \quad \bar{\beta}_w^T S_w + \bar{\lambda}_w^T d'^s_w(S_w, V_w) - \bar{\mu}_w^T d^d_w(S_w) \tag{32a}$$

$$\text{s.t.} \quad S_{wj} \le V_w \quad \forall_j \tag{32b}$$

The preceding series of decompositions replace the relaxed primal problem (RPP) with $2N$ easy subproblems which can be solved in parallel. Each instance of SNARP$_i$ is a linear program with $|P|$ variables and 1 general constraint. By Proposition (3.2), it can be solved by simply enumerating the objective value for each $p \in P$, of which there are a small constant number, and choosing the pattern with the highest value. Therefore the overhead of a general-purpose solver can be avoided. Each instance of SNRP-FLAP is a linear problem with $O(N)$ variables and constraints, although it will be further re-formulated.

E. Economic Interpretation

This formulation can be interpreted in the following way: In the coupling between the restricted master problem (RMP) and CLAP, the dual values $\bar{\beta}_{ij}$ represent the estimated value in terms of improvement to the overall schedule of accommodating more traffic on link ij. In the coupling between Lagrangian subproblems, $\bar{\lambda}_{ij}$ is the *signal quality price*: It represents the value of improving the SINR on link ij, and *duplex price* $\bar{\mu}_i$ represents the value of decreasing the usage of node i.

In SNRP-FLAP, each node activates links to maximize its utility, where $\bar{\beta} \ge 0$ is the reward for activating each link, $\bar{\lambda} \ge 0$ is the penalty for any SINR reduction on each link, and $\bar{\mu} \ge 0$ is the penalty for using each node. In SNARP, each node chooses antenna gains to maximize a different utility, defined solely in terms of $\bar{\lambda}$. When all the constants have their values substituted in, the objective function of SNARP$_i$ is of the form in equation (33), where the actual value of constant k_{ij} is determined by $\bar{\lambda}$, node j's antenna configuration, and RF parameters Lb, P, and Nr.

$$\max_{D_{ij}} \quad \sum_{j\neq i} D_{ij} k_{ij} \tag{33}$$

$$k_{ij} \quad \begin{cases} \ge 0 & \text{if } ij \text{ or } ji \text{ is an active link} \\ \le 0 & \text{if } ij \text{ or } ji \text{ is an ``interference link''} \\ = 0 & \text{otherwise} \end{cases}$$

F. Lagrange Multiplier Updates

The combined problems SNRP-FLAP$_i$ and SNARP$_i$ for all nodes i implement the relaxed primal problem. Solving proceeds by iteratively solving the RPP and updating the Lagrange multipliers λ and μ so that they converge to an optimal solution of the dual problem. We use a *subgradient* method because it lends itself to distributed implementation and because it scales well with the problem size. At time t, let s^t denote the degree of constraint violation, α^t the step size, and let $[\cdot]_+$ denote projection onto the nonnegative orthant. The subscripts λ and μ are used to distinguish the values pertaining to each set of Lagrange multipliers.

$$\begin{aligned} s^t_\lambda &= d^s(S^t, D^t, V^t) \\ s^t_\mu &= d^d(S^t) \\ \bar{\lambda}^{t+1} &\leftarrow \left[\bar{\lambda}^t + \alpha^t_\lambda s^t_\lambda\right]_+ \\ \bar{\mu}^{t+1} &\leftarrow \left[\bar{\mu}^t + \alpha^t_\mu s^t_\mu\right]_+ \end{aligned}$$

We define step size rule $\alpha^t = \dfrac{a}{(t+b)^2}$, $a > 0,\ b \ge 0$. The a and b are tunable parameters, and are not related to Guan's a and b in SDQ-FLAP and [28].

1) Convergence Properties: The subgradient method described above will produce optimal values of the Lagrange multipliers for CLAP-dual-2.

Proposition 3.3: The sequences $\{\bar{\lambda}^t\}$ and $\{\bar{\mu}^t\}$ converge to $\lambda^* \in \boldsymbol{\lambda^*}$ and $\mu^* \in \boldsymbol{\mu^*}$, where $(\boldsymbol{\lambda^*}, \boldsymbol{\mu^*})$ are the optimal sets of CLAP-dual-2.

Proof: Let X refer to the set of all decision variables, x refer to a vector value in X, and x_0 refer to some specific value of x, *not* a scalar component of x. Let G_s be any subgradient of d'^s and G_d be any subgradient of d'^d. Then $\lambda G_s(x_0) + \mu G_d(x_0)$ is a subgradient of $-\bar{\beta}^T S + \lambda^T d^s(S, D, V) + \mu^T d^d(S)$, by Shor's Thm 15 of [29]. This equals equation (28) on the preceding page. $\therefore\ S^t_\lambda + S^t_\mu$ is a subgradient of (28). The sum over all i of the objectives and constraints of SNRP-FLAP$_i$ and SNARP$_i$ equal the objectives and constraints of CLAP-dual-2. Assume that the Slater condition holds, otherwise the problem and JBSS-MP are infeasible.

$\int \frac{a}{(t+b)^n} d(t) = \frac{a(t+b)^{1-n}}{1-n}$, $n > 1$, which diverges as $t \rightarrow \infty$. $\therefore \sum_{t=0}^{\infty} \alpha^t = +\infty$. $\lim_{t\to\infty} \alpha^t = 0$ for $n, a > 0$. Therefore, $\{x\}$ converges to optimal x^* by [29, Thm 31]. ■

It does not follow that the sequence of *primal* values produced will converge to optimal S^*, D^* even though the problems exhibit strong duality. To address this, we define the following sequence:

$$\hat{S}^t = (1 - \alpha^t)\hat{S}^{t-1} + \alpha^t S^t \tag{34}$$

Proposition 3.4: $\{\hat{S}^t\}$ converges to S^*, and the analogous $\{\hat{D}^t\}$ converges to D^*

Proof: We appeal to a result by Larsson *et al.* [30]. $\{S^t\}$ is generated by a dual subgradient process satisfying his criteria (9)-(11). $\{\hat{S}^t\}$ is an ergodic sequence satisfying (7), (13). It follows from [30, Theorem 1] that $\{S^t\}$ converges to the solution set. The same applies to $\{\hat{D}^t\}$. ■

Our formulation exhibits a well-known issue with subgradient methods: small changes in the Lagrange multipliers produce large changes in primal solutions, causing oscillation around the ideal search trajectory. This can slow the solution process. The linear objective function and previously-mentioned integrality property contribute to this behavior in FLAP and its derived problems. Additionally, when the relevant SINR prices λ are 0 and the $\bar{\beta}$ values are the same, links which share a node exhibit the *homogeneous subproblem* property where any given dual price μ will result in the same primal outcome for all links. This issue arises in the context of the hydrothermal *unit commitment* problem, and is commonly addressed using surrogate subgradient methods. To simplify decentralized implementation, however, we instead use a nonlinear approximation method of the form presented in [28]. This is conceptually very similar to an augmented Lagrangian, but the additional quadratic parameter is computed locally for each subproblem, maintaining the separable structure of the original program. Based on this transformation, we introduce the Single-node Dual Quadratic FLAP, or SDQ-FLAP, where a_{ij} and b_{ij} are defined as in [28].

[SDQ-FLAP$_i$]

$$\max_{S} \quad \sum_{j:ij\in A} -a_{ij}{S_{ij}}^2 + (b_{ij} - \bar{\lambda}_{ij}S_{ij}) - \frac{\bar{\mu}_i + \bar{\mu}_j}{2} S_{ij} \tag{35a}$$

$$\text{s.t.} \qquad S_{ij} \le V_i \quad \forall_j \tag{35b}$$

Proposition 3.5: Any stable solution to SDQ-FLAP$_i$ is also a solution to SNRP-FLAP$_i$.

Proof: The constraints of SDQ-FLAP$_i$ are identical to those of FLAP$_i$. At any point x_0, the nonlinear approximation f' generated at x_0 is parallel to f [28]. For both SDQ-FLAP and SNRP-FLAP, the constraints are all differentiable and convex, and the objective is convex (when stated as minimization). Therefore the Karush-Kuhn-Tucker conditions are sufficient for global optimality.

Let x^* be an optimal solution of SDQ-FLAP$_i$ as constructed at x_0. The KKT conditions therefore hold. Assume x^* is stable, therefore $x^* = x_0$. Suppose that x_0 is not an optimal solution of SNRP-FLAP$_i$. The KKT conditions other than stationarity are the same in both cases, so they must hold for SNRP-FLAP$_i$. Therefore the stationarity condition must hold for SDQ-FLAP$_i$ but not for SNRP-FLAP$_i$. That requires, for the same constraints, that $\nabla f'(x^*) \neq \nabla f(x^*)$, $\perp$. ■

The preceding formulation significantly reduces oscillation relative to FLAP or its decomposed analogue, SNRP-FLAP. The constraint (35b) can be ignored: The variable V_i does not appear in the objective function and is otherwise free, so the problem can be solved for S and V_i chosen to be $\max(S_{ij})$. SDQ-FLAP$_i$ is therefore an *unconstrained* (or bound-constrained) quadratic program with at most $2N$ variables.

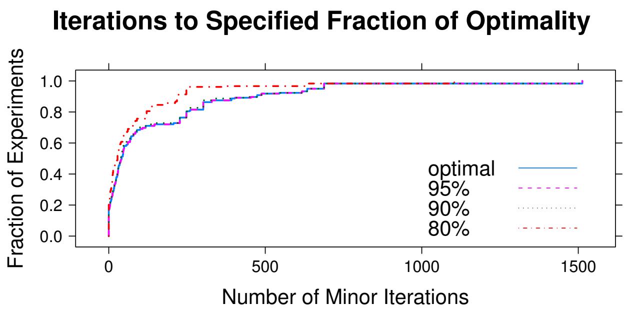
G. Partial Pricing

Recall that the objective of the column generation subproblem is to find improving feasible points for inclusion in the restricted master problem. The optimality of the overall result does not require that the subproblem finds the *most* improving point, only that it finds *an* improving point if one exists. We exploit this by using "partial pricing" and returning the first improving primal feasible result $(\hat{S}^t, \hat{D}^t)$ – which may or may not be the best possible – without waiting for the subgradient process to converge [31]. It is only necessary to allow the subproblem to fully converge to prove that there is no as-yet-undiscovered feasible improving point.

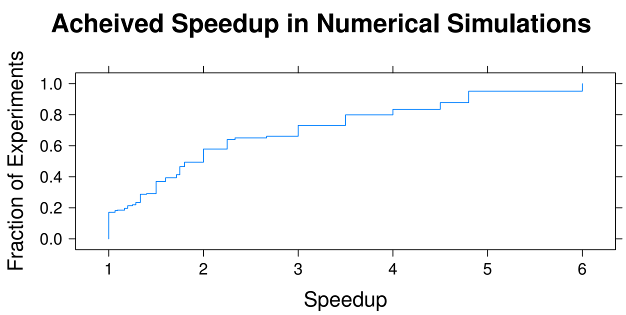
Every solution to an iteration of the restricted master problem is a valid schedule. Each such schedule can be put into place in the network immediately if it is superior to the current schedule, regardless of whether or not it is the final, best schedule. Consequently, terminating the subproblem early and re-solving the RMP yields a useful result sooner than solving the subproblem to optimality, even though it may or may not improve the overall running time.

H. Distributed Consensus

The preceding sections decompose the original problem into a form where $2N$ small problems are solved in parallel for each subgradient update iteration. Going from a parallel algorithm to a distributed one requires some consideration of the communication processes. We make use of a very simple and robust model due to [32]: Every node maintains its own version of every variable, and nodes announce their variable values to other nodes occasionally. A node may actively compute locally-generated values for some, all, or no variables. Upon computing a new value or receiving other nodes' variable values, a node updates its own values according to a weighted averaging scheme. Under surprisingly light requirements on the weights and communication frequencies, it is shown that this scheme has the same convergence properties as its centralized counterpart. The results in [32] are only shown for objective functions which are continuously differentiable and Lipschitz, while the Lagrangian dual function is generally



(a) Distribution of number of (minor) iterations necessary in simulations to first reach and optimal solution.



(b) Empirical cumulative distribution of achieved speedup (ratio of optimal to TDMA performance) across all simulations.

Fig. 5: Performance in Simulations

non-differentiable. Similar results are proven for the non-differentiable case in [33].

IV. Numerical Experiments

This section considers the performance of the optimization process taken in isolation. These experiments emulate a distributed algorithm in that each node's computations are performed separately. Experiments were conducted by running the algorithm over a large number of scenarios constructed with varying initial values. In total, 1396 experiments were run. The following major parameters were varied: Number of nodes (between 0 and 48), the number of links (between $\frac{1}{2}$ and 3 links per node), and the size of the simulated region (between 1 and 16 square km). For each set of parameters, nodes were randomly placed within the simulated area with uniform probabilities, and pairwise path losses were estimated using the Green-Obaidat model [34]. All possible links were identified based on a hypothetical transmission power of 14.7 dBm, a required signal strength of -80 dBm, and the best-case antenna gains given a measured phased array antenna beam pattern. The requested number of links were chosen randomly from the pool of possible links; if enough possible links did not exist then a new layout was generated. The results presented here are aggregates across all of these scenarios—a full factorial analysis is planned for future empirical studies of these algorithms and associated STDMA MAC.

A. Running time

A well-known limitation of subgradient methods for updating Lagrange multipliers is that they are very slow to reach a provably converged state. This means in practice that such algorithms may find optimal values relatively quickly, but then require a longer period to verify that no better values exist. As alluded to in section III-G on the previous page, *termination* may not be the best criterion for an on-line system. It is expected that schedule optimization will be a continuous process, converging and diverging as system parameters change. Consequently, we find it useful to examine the time required to find optimal and near-optimal solutions as well as the time to termination. In our experiments, we find that by either measure, execution time is *at worst* linear in the size of the input.

To quantify the behavior, see Figure 5a which plots the distribution of the number of iterations required to first reach the optimal solution across all of our simulation runs. We can see that in more than 90% of the cases, the optimal solution is found within 500 iterations (the mean is 150 iterations and 91.83% are solved to optimality within 500), yet some scenarios may require as many as 1500 iterations to settle on the optimal solution. On average, we are able to get within 10% of optimal within 146 iterations and within 20% of optimal within only 85 iterations.

B. Schedule Efficiency

In addition to convergence properties, our numerical experiments provide a window into the ability of the algorithm to produce efficient (high-reuse) schedules across a large number of randomly generated scenarios. Figure 5b plots a speedup metric which is the ratio of the time required by a TDMA MAC to service a given demand relative to the time required by our optimized system. This is either an increase in throughput, given a fixed amount of time, or an increase in free spectrum time, given a fixed workload. In our experiments, we see speedup values ranging from 1 (no speedup) to 6 with an average speedup of 2.34 across all scenarios ($\sigma = 1.31$).

V. Deployed System

Thus far, this paper has described and evaluated our *mathematical* design; this subsection addresses its concrete implementation. The system we present here operates in a *fully distributed*, *asynchronous* manner. Nodes maintain and exchange variables as described in § III-H.

In addition to the subproblem solver processes, there is a separate process for the early termination check and restricted master problem (RMP). When this process detects that its current estimates ($\bar{S}^t$, $\bar{D}^t$) constitute a primal feasible solution with negative reduced cost, a corresponding new column is added to the RMP, which is re-solved. The resulting new schedule, updated dual prices $\bar{\beta}$, and step size reference time are sent to all nodes by a flooding protocol.

Fig. 6 on the following page shows the execution of the algorithm on a wide-area phased array antenna testbed, as observed by a single node (C). In this scenario, node B is

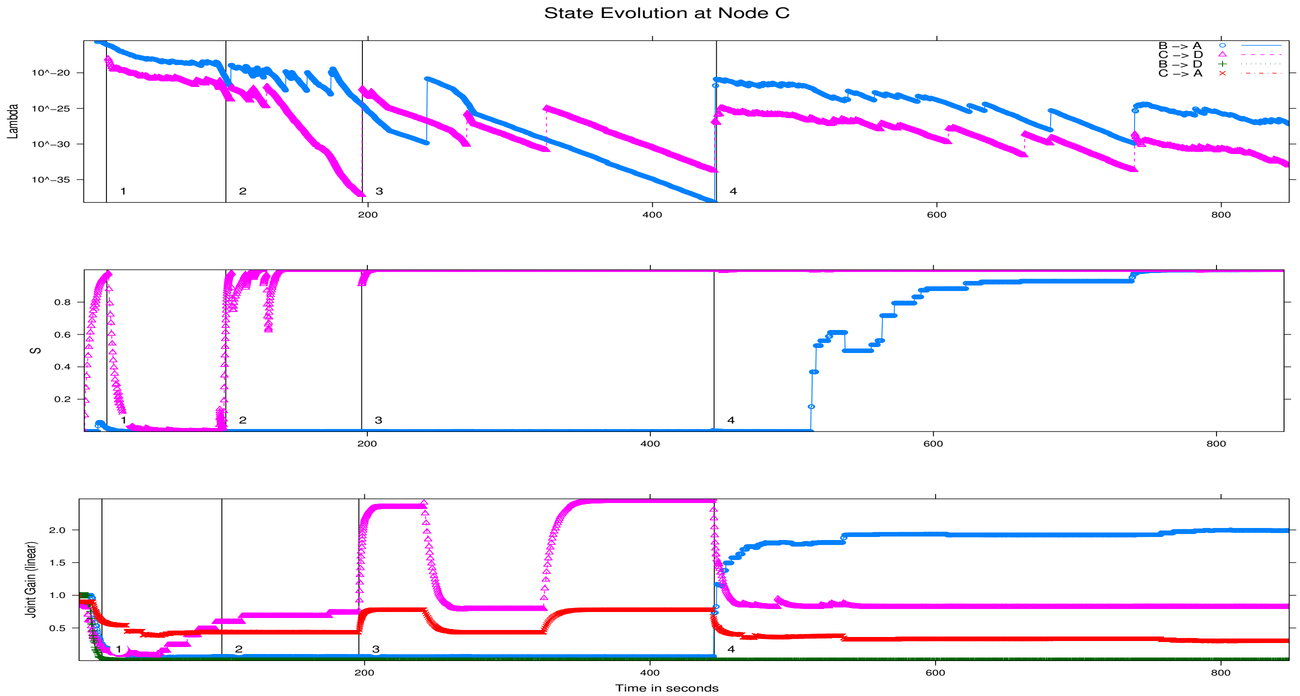


Fig. 6: Trace of algorithm scheduling links B $\to$ A and C $\to$ D concurrently, as seen locally at node C. The top strip shows $\bar{\lambda}$, middle strip shows $\hat{S}$, and the bottom strip shows the *combined* gain $\hat{D}_{ij}\hat{D}_{ji}$. Note that B $\to$ D and C $\to$ A are interference if both data links are active. The aligned x axis is time in seconds.

sending traffic (a stream of UDP packets with 1024 byte payloads) to node A while node C sends a similar volley to node D. With naïve beam-steering, these links are bad neighbors—node B will cause substantial interference at node C (62.5 dBm on average). Hence, these links can be activated simultaneously, but only carefully.

The top strip of this figure shows the evolution of the SINR Lagrange multiplier estimates $\bar{\lambda}$ – the *signal quality prices*, the second strip shows the consensus estimated link activations $\hat{S}$, and the third strip shows the combined antenna gains for the signals and interference. Times of interest are marked with a vertical bar and labelled (1, 2, ...) on all strips. No change to actual system state occurs until a new execution of the RMP; the link activation and antenna configurations referred to are variable values. Qualitatively, the execution of the algorithm can be understood in the following stages:

Prior to time 1, node C's estimate $\bar{\lambda}_{\text{CD}}$ is 0 and does not appear on the log-scale. This drives the link activation $\hat{S}_{\text{CD}}$ toward 1, while the SNARP objective is undefined and the resulting gains are low.

At time 1, high activation and low gain causes the SINR constraint for link C $\to$ D to be violated. The price $\bar{\lambda}_{\text{CD}}$ takes a large step to $> 10^{-20}$. This increase drives $\hat{S}_{\text{CD}}$ back toward 0, and causes $\hat{D}_{\text{CD}}\hat{D}_{\text{DC}}$ to start increasing. The price $\bar{\lambda}_{\text{CD}}$ decreases as the low activation and higher gain stay within the constraints.

At time 2, the combination of low λ_{CD} and higher gain allows $\hat{S}_{\text{CD}}$ to increase to near 1. At time 3, $\hat{S}_{\text{CD}}$ gets close enough to 1 to violate the SINR constraint again and drive up $\bar{\lambda}_{\text{CD}}$ The increase in $\bar{\lambda}_{\text{CD}}$ relative to $\bar{\lambda}_{\text{BA}}$ drives the antennas to favor $\hat{D}_{\text{CD}}\hat{D}_{\text{DC}}$. Note that this antenna configuration at node C has high gain toward A, raising the unwanted gain $\hat{D}_{\text{CA}}\hat{D}_{\text{AC}}$.

Between times 3 and 4, $\bar{\lambda}_{\text{CD}}$ and $\bar{\lambda}_{\text{BA}}$ trend down, but changes in their relative magnitude cause the antenna state to switch back and forth. Immediately before time 4, $\hat{S}_{\text{BA}}$ increases almost invisibly. Recall that node C is *not computing* S_{BA}, so a change in $\hat{S}_{\text{BA}}$ reflects the incorporation of a value broadcast by node B. This change is sufficient to cause an SINR constraint violation, driving $\bar{\lambda}_{\text{BA}}$ up at time 4.

Note two changes with regard to the gains: First $\hat{D}_{\text{AB}}\hat{D}_{\text{BA}}$ increases dramatically, reflecting a change in antenna configuration by node B. Second, the change in $\bar{\lambda}_{\text{BA}}$ causes node C to change its antenna configuration to diminish $\bar{D}_{\text{CA}}$, at the cost of also reducing $\bar{D}_{\text{CD}}$. This new configuration can accommodate both links, and $\hat{S}_{\text{BA}}$ tends toward 1 as node C receives updates from other nodes. At this point the RMP can schedule the two links concurrently with the configuration given. Note that $\bar{\lambda}$ continues to vary but this variation does not effect the primal estimates.

The prototype implementation contains significant inefficiencies. Every subproblem solution involves writing several files, synchronizing processes, and running a stand-alone NLP solver. Additionally, the STDMA driver schedules very large time slots, so freshly computed values wait as much as a second to be sent. The post-decomposition subproblems are simple enough for the results to be easily calculated directly without the machinery of a general-purpose solver. In preliminary experiments, this accelerates the computation process by a factor of 10^5. We expect that the communication process can similarly be improved, though not by as dramatic a factor.

VI. Conclusion

This paper presents a price-coupled decomposition structure for jointly optimizing *which users* access spectrum *when* and *how*. Using this structure, we solve the joint beam steering and scheduling problem. Optimal spatial reuse TDMA scheduling is known to be NP-hard, and the addition of antenna configuration increases the state space *exponentially* in the number of nodes. In solving this problem, we provide the *first implementation of wireless scheduling based on dual decomposition of signal constraints*. These algorithms are computationally efficient—they find the optimal solutions within hundreds of iterations, each of which requires only minimal computation. Despite the NP-hard nature of the underlying problem, our running time appears linear in the problem size in practice. The algorithm makes very few assumptions about the patterns of the antennas' directionality or the environment's path loss. These algorithms *directly* subsume power control and rate selection, and the same decomposition structure can form a basis for other algorithms and protocols, including ones addressing different physical parameters or making different optimality-overhead tradeoffs.

Though subject to some convexity requirements, the pattern of *signal quality price decomposition* generalizes to a wide range of joint optimization problems involving interacting radio-frequency systems [35]. We firmly believe that optimization decomposition is a paradigm that will drive next-generation wireless networks and offer our work here as an important step towards realizing the theoretical gains of this approach in real systems.

References

[1] E. Anderson, C. Philips, G. Yee, D. Sicker, and D. Grunwald, "Challenges in deploying steerable wireless testbeds," in *TridentCom*, 2010.

[2] R. L. Freeman, *Radio System Design for Telecommunications*, 2nd ed. Wiley InterScience, 1997.

[3] E. Arikan, "Some complexity results about packet radio networks," *IEEE Trans. on Information Theory*, vol. 30, 1984.

[4] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *Wireless Communications, IEEE Trans. on*, vol. 2, 2003.

[5] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, 1998.

[6] P. Björklund, P. Varbrand, and D. Yuan, "Resource optimization of spatial TDMA in ad hoc radio networks: a column generation approach," in *INFOCOM 2003*, vol. 2. IEEE, 2003.

[7] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *Communications, IEEE Trans. on*, vol. 52, no. 7, 2004.

[8] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *Selected Areas in Communications, IEEE Journal on*, vol. 23, 2005.

[9] ——, *Geometric Programming for Communication Systems*. Now Publishers Inc. Hannover, MA, 2005.

[10] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, 2007.

[11] C. W. Tan, D. Palomar, and M. Chiang, "Exploiting hidden convexity for flexible and robust resource allocation in cellular networks," in *INFOCOM 2007*, 2007.

[12] L. Yang, L. Cao, and H. H. Zheng, "Physical interference driven dynamic spectrum management," in *DySPAN 2010*, 2008.

[13] L. Cao, L. Yang, X. Zhou, Z. Zhang, and H. Zheng, "Optimus: Sinr-driven spectrum distribution via constraint transformation," in *DySPAN 2010*, 2010.

[14] J. B. Cain, T. Billhartz, L. Foore, E. Althouse, and J. Schlorff, "A link scheduling and ad hoc networking approach using directional antennas," in *Military Communications Conference, 2003*, vol. 1. IEEE, 2003.

[15] L. Bao and J. J. Garcia-Luna-Aceves, "Transmission scheduling in ad hoc networks with directional antennas," in *MOBICOM '02, Proceedings of*. ACM, 2002.

[16] A. Deopura and A. Ganz, "Provisioning link layer proportional service differentiation in wireless networks with smart antennas," *Wireless Networks*, vol. 13, 2007.

[17] K. Sundaresan, W. Wang, and S. Eidenbenz, "Algorithmic aspects of communication in ad-hoc networks with smart antennas," in *MobiHoc '06*. ACM, 2006.

[18] W.-D. Wirth, *Radar techniques using array antennas*. The Institution of Electrical Engineers, 2001.

[19] M. Sánchez and J. Zander, "Adaptive antennas in spatial TDMA multihop packet radio networks," in *RadioVetenskap och Kommunikation (RVK)*, 1999.

[20] K. Dyberg, F. Farman, L.and Eklof, J. Grönkvist, U. Sterner, and J. Rantakokko, "On the performance of antenna arrays in spatial reuse TDMA ad hoc networks," in *MILCOM 2002*, vol. 1, 2002.

[21] X. Liu, A. Sheth, M. Kaminski, K. Papagiannaki, S. Seshan, and P. Steenkiste, "DIRC: Increasing indoor wireless capacity using directional antennas," in *Proc. SIGCOMM 2009*. ACM, 2009.

[22] X. Liu, A. Sheth, K. Papagiannaki, S. Seshan, and P. Steenkiste, "Pushing the envelope of indoor wireless spatial reuse using directional access points and clients," in *MobiCom*, 2010.

[23] E. Jorswieck, P. Svedman, and B. Ottersten, "On the performance of TDMA and SDMA based opportunistic beamforming," *Wireless Communications, IEEE Trans. on*, 2007.

[24] P. Björklund, "Applications of resource optimization in wireless networks," Ph.D. dissertation, Linköping Universitet, Linköping, Sweden, 2006.

[25] L. Chen, S. H. Low, and J. C. Doyle, "Joint congestion control and media access control design for ad hoc wireless networks," in *IEEE INFOCOM*, vol. 3. IEEE, 2005.

[26] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *Wireless Communications, IEEE Trans. on*, vol. 5, 2006.

[27] V. Chvátal, *Linear Programming*. Freeman Press, 1980.

[28] X. Guan, P. B. Luh, and L. Zhang, "Nonlinear approximation method in lagrangian relaxation-based algorithms for hydrothermal scheduling," *IEEE Trans. on Power Systems*, vol. 10, 1995.

[29] N. Z. Shor, *Nondifferentiable Optimization and Polynomial Problems*, ser. Nonconvex Optimization and its Applications. Kluwer Academic Publishers, 1998.

[30] T. Larsson, M. Patriksson, and A.-B. Strömberg, "Ergodic, primal convergence in dual subgradient schemes for convex programming," *Mathematical Programming*, vol. 86, 1999.

[31] J. Desrosiers and M. E. Lübbecke, "A primer in column generation," in *Column Generation*, ser. Groupe d'études et de recheche en analyse des décisions (GERAD) 25th anniversary series, G. Desaulniers, J. Desrosiers, and M. M. Solomon, Eds. Springer, 2005, pp. 1 – 33.

[32] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," in *IEEE Trans. on Automatic Control*, vol. 31, 1986.

[33] A. Nedić, D. Berteskas, and V. Borkar, "Distributed asynchronous incremental subgradient methods," *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, 2001.

[34] D. Green and A. Obaidat, "An accurate line of sight propagation performance model for ad-hoc 802.11 wireless LAN (WLAN) devices," in *ICC 2002*, vol. 5, 2002.

[35] E. W. Anderson, "Integrated scheduling and beam steering for spatial reuse," Dissertation, University of Colorado, Department of Computer Science, 430 UCB Boulder, CO USA, 2010. [Online]. Available: http://www.ece.cmu.edu/~andersoe/