Joint Multicast Beamforming and Antenna Selection

Omar Mehanna, *Student Member, IEEE*, Nicholas D. Sidiropoulos, *Fellow, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Multicast beamforming exploits subscriber channel state information at the base station to steer the transmission power towards the subscribers, while minimizing interference to other users and systems. Such functionality has been provisioned in the long-term evolution (LTE) enhanced multimedia broadcast multicast service (EMBMS). As antennas become smaller and cheaper relative to up-conversion chains, transmit antenna selection at the base station becomes increasingly appealing in this context. This paper addresses the problem of joint multicast beamforming and antenna selection for multiple co-channel multicast groups. Whereas this problem (and even plain multicast beamforming) is NP-hard, it is shown that the mixed $\ell_{1,\infty}$-norm *squared* is a prudent group-sparsity inducing convex regularization, in that it naturally yields a suitable semidefinite relaxation, which is further shown to be the Lagrange bi-dual of the original NP-hard problem. Careful simulations indicate that the proposed algorithm significantly reduces the number of antennas required to meet prescribed service levels, at relatively small excess transmission power. Furthermore, its performance is close to that attained by exhaustive search, at far lower complexity. Extensions to max-min-fair, robust, and capacity-achieving designs are also considered.**

***Index Terms*—Antenna selection, capacity, complexity, multicasting, NP-hard, relaxation, semidefinite programming, sparsity, transmit beamforming.**

I. Introduction

CONSIDER a base station (BS) transmitter using an antenna array to broadcast common information to multiple radio subscribers. Instead of broadcasting isotropically, the BS can exploit subscriber channel state information (CSI) to select different weights for each antenna in order to steer power in the directions of the subscribers while limiting interference to other users. This type of *multicast beamforming* is provisioned under the enhanced multimedia broadcast multicast service (EMBMS) of the long term evolution (LTE) standard. After considerable market-related delays, EMBMS is scheduled for initial roll-out in 2012. EMBMS can markedly boost spectral efficiency and reduce energy and infrastructure costs per bit when the same content must be delivered wirelessly to multiple subscribers.

Manuscript received March 21, 2012; revised August 27, 2012, December 03, 2012, and February 08, 2013; accepted February 12, 2013. Date of publication March 11, 2013; date of current version April 26, 2013. Supported in part by NSF CCF grant 0747332. A conference version [13] of part of this work was presented at the Thirteenth IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cesme, Turkey, June 17–20, 2012.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail meha0006@umn.edu; nikos@umn.edu; georgios@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2252167

In practice, a BS may have more antennas than expensive radio transmission chains, and it is desired to automatically switch the available chains to the most appropriate subset of antennas in an adaptive fashion. Each radio transmission chain includes a digital-to-analog (D/A) converter, a mixer, and a power amplifier. Antenna elements, on the other hand, are becoming smaller and cheaper; thus, *antenna selection* strategies are becoming increasingly desirable.

The multicast beamforming problem under minimum received signal-to-noise ratio (SNR) constraints was initially studied in [18]. The problem was shown to be NP-hard, however a computationally efficient approximate solution was developed based on semidefinite relaxation. This formulation was later extended to multiple co-channel multicast groups in [8], cognitive underlay scenarios [16], and joint multicast beamforming and admission control [12]. However, antenna selection has not been considered in any of these papers. On the other hand, antenna subset selection has been initially considered for point-to-point multiple-input multiple-output (MIMO) links using various techniques [6], [17], [25]. For the multicast scenario, an antenna selection scheme has been proposed in [15], where the antenna subset is chosen to maximize the minimum SNR across all users, assuming that the BS transmits mutually uncorrelated signals of equal power from the different antennas (across the transmission chains). In this case, maximizing the minimum SNR also maximizes the multicast rate under the constraint of spatially white transmission. A limitation is that attaining this rate requires complex multi-stream Shannon encoding and decoding at long block lengths, also implying long decoding delay that is not suitable for streaming media multicast. While using a spatially white transmit covariance does not require CSI at the transmitter, the antenna selection strategy in [15] requires knowledge of all channel gains at the transmitter. But if CSI is known at the transmitter, then it is possible to choose the transmit covariance accordingly, thus attaining higher rate. Beamforming, on the other hand, requires far simpler encoding and decoding with CSI at the transmitter, and is often close to attaining multicast capacity [7], [18]. It is also worth noting that the optimal higher-rank transmit covariance is obtained as a by-product of [18]. Another significant difference between the work reported here and [15] is that the latter requires exhaustively searching through all antenna subset possibilities, whereas the present paper's computationally efficient algorithm performs the antenna selection and beamforming design tasks jointly.

Convex sparsity-inducing regularizers have been widely used in various applications (cf. [1] and references therein). The most commonly used regularizer is the ℓ_1-norm, which has been used in recent works for receive beamforming antenna selection [5], [14]. Beampattern synthesis with antenna selection was pursued in [14], using a convex optimization formulation that controls the mainlobe and sidelobes while minimizing the sparsity-inducing ℓ_1-norm to produce a sparse beamforming weight vector involving fewer antennas. The setup in [14] only applies to uniform linear antenna-array (ULA) far-field scenarios, whereas the present paper's approach works for arbitrary channel (or *steering*) vectors. Another important difference is that [14] restricts the beamforming weights to be conjugate symmetric in order to turn the non-convex lower bound constraints on the beampattern into affine ones. This gives up half of the problem's design variables (degrees of freedom), thereby yielding suboptimal solutions when only the magnitude of the beampattern is important, as in transmit beamforming. No such restriction is placed on the beamforming weight vectors here. In a similar vain, [5] considered using the ℓ_1-norm to obtain sparse solutions to convex beampattern synthesis problems. Whereas [5] does not restrict the weight vector to be conjugate symmetric, it does unnecessarily constrain the phase of the beampattern; without such a constraint on the phase, the problem is non-convex, and thus more challenging.

In this paper, the joint problem of transmit beamforming and antenna selection is considered for multiple co-channel multicast groups. Whereas this problem (and even plain multicast beamforming) is NP-hard, we show that using the mixed $\ell_{1,\infty}$-norm *squared* as a *group-sparsity* inducing convex regularization yields a natural semidefinite programming (SDP) relaxation. Sparse beamforming vectors can be obtained from the resulting sparse solution, implying antenna selection. In order to further enhance sparsity, an iterative re-weighting scheme similar to the one used in [4] is employed. Moreover, we show that the same approach can be used to obtain a tight lower bound on the multicast channel capacity with antenna selection. More generally, the proposed novel algorithm can easily be extended and applied to obtain sparse solutions for a wide class of non-convex quadratically constrained quadratic programming (QCQP) problems for which SDP relaxation is relevant (cf. [11] and references therein). Simulations indicate that the proposed algorithm considerably reduces the number of antennas required to meet prescribed service levels, at a small cost in excess transmission power. Furthermore, its performance is close to that attained by exhaustively trying all antenna subsets, at far lower complexity.

Relative to the conference submission [13], this journal version i) treats the general case of *multi-group* multicasting with *group sparsity* of the matrix of beamforming vectors, instead of the single group case with plain sparsity of the beamforming vector; ii) proves that the proposed relaxation admits a Lagrange bi-dual interpretation, which is interesting because the native (group) sparsity-inducing formulation is not a QCQP; iii) includes a discussion of relevant extensions, from max-min to robust and capacity-achieving designs; and iv) fleshed-out numerical results and comparisons.

The algorithms presented here employ general-purpose SDP solvers, which can effectively deal with up to a moderate number of antennas and users (order of 100 when using a typical personal computer as of this writing). They are not customized to handle many hundreds or even thousands of transmit antennas, as in some recent proposals for *Massive MIMO* systems [26]. Developing custom algorithms for joint multicast beamforming and antenna selection for Massive MIMO is certainly of interest, but striking the right balance between performance and complexity for such systems requires a very different approach. We have preliminary results in this direction, which will be reported in follow-up work. Here we focus on up to moderate-size systems, which are the norm as of this writing.

Notation: Boldface uppercase letters denote matrices, whereas boldface lowercase letters denote column vectors. The superscripts $(\cdot)^T$ and $(\cdot)^{\mathcal{H}}$ denote transpose and Hermitian (conjugate) transpose operators, respectively. $\mathrm{tr}(\cdot)$, $\mathrm{rank}(\cdot)$, $\|\cdot\|_2$, $|\cdot|$, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the trace, the rank, the Euclidean norm, the absolute value (element-wise absolute if used with a matrix), the real, and the imaginary operators, respectively; $x(i)$ denotes the i-th entry of $\mathbf{x}$ and $X(i,j)$ the (i,j)-th entry of $\mathbf{X}$. MATLAB notation $\mathbf{X}(i_1 : i_2, j_1 : j_2)$ stands for the submatrix of $\mathbf{X}$ obtained by deleting all rows and columns whose indices do not fall in the range $i_1 : i_2$ and $j_1 : j_2$, respectively; $\mathbf{X} \geq \mathbf{Y}$ denotes an element-wise inequality, whereas $\mathbf{X} \succeq 0$ denotes that $\mathbf{X}$ is a Hermitian positive-semidefinite matrix. Finally, $\mathbf{I}_N$, $\mathbf{1}_N$, $\mathbf{1}_{N\times M}$, and $\mathbf{0}_{N\times M}$ denote the $N \times N$ identity matrix, the $N \times N$ matrix with all one entries, the $N \times M$ all ones matrix, and the $N \times M$ all zeros matrix, respectively.

II. Problem Formulation

A. Basic Model

The system model is similar to [8], comprising a single BS transmitter with N antennas and M single-antenna receivers. We assume there are K multicast groups ($1 \leq K \leq M$), and each receiver listens to a single multicast. The set of receivers participating in multicast group $k \in \{1,\ldots,K\}$ is denoted by G_k, and $\sum_{k=1}^{K} |G_k| = M$. The BS broadcasts a common message to the receivers of each multicast group. Vector $\mathbf{w}_k \in \mathbb{C}^N$ is formed by the beamforming weights applied to the N transmit-antenna elements for transmission to multicast group k. The temporal information-bearing waveform intended for multicast group k is denoted by $s_k(t)$. The transmitted signal vector is $\sum_{k=1}^{K} \mathbf{w}_k^{\mathcal{H}} s_k(t)$. Assuming that $\{s_k(t)\}_{k=1}^{K}$ are temporally white, zero-mean, unit variance, and mutually uncorrelated, the total transmission power is $\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2$. The complex vector that models the propagation loss and the frequency-flat quasi-static channel from each transmit antenna to the receive antenna of user m is denoted by $\mathbf{h}_m$, $m \in \{1,\ldots,M\}$. The noise at receiver m is assumed zero-mean white, with variance σ_m^2. The signal-to-interference-plus-noise ratio (SINR) at receiver $m \in G_k$ is then given by

$$\mathrm{SINR}_{m,k} = \frac{|\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m|^2}{\sum\limits_{l=1,l\neq k}^{K} |\mathbf{w}_l^{\mathcal{H}}\mathbf{h}_m|^2 + \sigma_m^2}.$$

It is assumed that the BS has acquired $\{\mathbf{h}_m\}_{m=1}^{M}$ and $\{\sigma_m^2\}_{m=1}^{M}$. The design problem is to minimize the total transmit-power, subject to prescribed receive-SINR thresholds γ_m at each user; that is

$$\boxed{\begin{array}{ll} \min\limits_{\{\mathbf{w}_k \in \mathbb{C}^N\}_{k=1}^K} & \sum\limits_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \\ \text{s.t.:} & \dfrac{|\mathbf{w}_k^{\mathcal{H}} \mathbf{h}_m|^2}{\sum\limits_{l \neq k} |\mathbf{w}_l^{\mathcal{H}} \mathbf{h}_m|^2 + \sigma_m^2} \geq \gamma_m, \\ & \forall m \in G_k, \quad \forall k, l \in \{1, \ldots, K\}. \end{array}} \tag{1}$$

The quadratic constraints in (1) are *non-convex*; therefore, (1) is a non-convex optimization problem. In fact, (1) is NP-hard for general channel vectors, even for a single multicast group ($K = 1$) [18]. Problem (1) has been studied in [8], where a convex approximate SDP reformulation was developed yielding an efficient near-optimal solution. For the special case $K = M$, that is, when each user receives an independent message with no multicasting, (1) can be reformulated as a convex, second-order cone programming problem [2]. It is also worth noting that if the channel vectors are confined to those resulting from a transmit ULA in the far-field, line-of-sight scenario (Vandermonde channels $\{\mathbf{h}_m\}_{m=1}^M$), problem (1) can be recast as a convex problem, and thus it can be solved efficiently [9].

B. Antenna Selection

Suppose now that only $L \leq N$ RF transmission chains are available, and thus only L antennas can be transmitting simultaneously. The goal is to jointly select the*best* L out of N antennas, and find the corresponding beamforming vectors $\{\mathbf{w}_k\}_{k=1}^K$ so that the transmission power is minimized, subject to receive-SINR constraints per subscriber. Both objectives must be jointly considered, because the constituent *selection* and *beamforming* problems are tightly coupled.

Define the $K \times 1$ vector $\bar{\mathbf{w}}_n := [w_1(n), \ldots, w_K(n)]^T$, where $w_k(n)$ is the n-th component of $\mathbf{w}_k$. Vector $\bar{\mathbf{w}}_n$ collects all multicast group weights applied to the n-th antenna. Define also the $NK \times 1$ concatenated beamforming vector $\mathbf{w} := [\mathbf{w}_1^T \ldots \mathbf{w}_K^T]^T$, and the $N \times 1$ vector $\tilde{\mathbf{w}} := [\|\bar{\mathbf{w}}_1\|_2, \ldots, \|\bar{\mathbf{w}}_N\|_2]^T$. For an antenna to be excluded from transmission, vector $\bar{\mathbf{w}}_n$ must be set to zero. This means that the n-th entry of each $\mathbf{w}_k$, for all K multicast groups, must be set to zero simultaneously. Hence, the joint antenna selection and transmit-power minimization problem can be expressed as

$$\begin{array}{ll} \min\limits_{\{\mathbf{w}_k \in \mathbb{C}^N\}_{k=1}^K} & \sum\limits_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \\ \text{s.t.:} & \dfrac{|\mathbf{w}_k^{\mathcal{H}} \mathbf{h}_m|^2}{\sum\limits_{l \neq k} |\mathbf{w}_l^{\mathcal{H}} \mathbf{h}_m|^2 + \sigma_m^2} \geq \gamma_m, \\ & \|\tilde{\mathbf{w}}\|_0 \leq L \quad \forall m \in G_k, \ \forall k, l \in \{1, \ldots, K\} \end{array} \tag{2}$$

where the ℓ_0-(quasi)norm is the number of nonzero entries of $\tilde{\mathbf{w}}$; i.e., $\|\tilde{\mathbf{w}}\|_0 := |\{n : \|\bar{\mathbf{w}}_n\|_2 \neq 0\}|$. Instead of the hard sparsity constraint, an ℓ_0 penalty can be employed to promote sparsity, leading to

$$\boxed{\begin{array}{ll} \min\limits_{\{\mathbf{w}_k \in \mathbb{C}^N\}_{k=1}^K} & \sum\limits_{k=1}^{K} \|\mathbf{w}_k\|_2^2 + \lambda \|\tilde{\mathbf{w}}\|_0 \\ \text{s.t.:} & \dfrac{|\mathbf{w}_k^{\mathcal{H}} \mathbf{h}_m|^2}{\sum\limits_{l \neq k} |\mathbf{w}_l^{\mathcal{H}} \mathbf{h}_m|^2 + \sigma_m^2} \geq \gamma_m, \\ & \forall m \in G_k, \ \forall k, l \in \{1, \ldots, K\} \end{array}} \tag{3}$$

where λ is a positive real tuning parameter that controls the sparsity of the solution, and thus the number of selected antennas. Problem (3) strikes a balance between minimizing the transmission power and minimizing the number of selected antennas, where a larger λ implies a sparser solution. Note that for any λ, there is a corresponding L for which problems (3) and (2) yield the same sparse solution, and thus focus is placed on (3) only.

Whereas the SINR constraints can be satisfied in the single multicast group case with only one antenna ($L = 1$) transmitting at sufficiently high power (assuming no channel coefficient is identically zero), the situation is not the same for multiple multicast groups. Problems (2) and (3) can be infeasible due to strong interference, stringent SINR constraints, high correlation between channels of users belonging to different multicast groups, and/or insufficient number of transmit-antennas used.

Unfortunately, due to the ℓ_0-(quasi)norm, solving (3) requires an exhaustive combinatorial search over all $\binom{N}{L}$ possible sparsity patterns of $\tilde{\mathbf{w}}$, where the NP-hard problem (1) must be solved (or closely approximated using the algorithm in [18]) for each of these patterns. This motivates the pursuit of computationally efficient, near-optimal solutions. The ensuing section introduces a convex sparsity-inducing approximation to the ℓ_0-norm, which is then used in obtaining a convex relaxation to (3).

III. RELAXATION

A. Group-sparsity Inducing Norms

For the special case of a single multicast group ($K = 1$), the ℓ_1-norm (defined as $\|\mathbf{w}\|_1 := \sum_{n=1}^N |w(n)|$) is known to offer the closest convex approximation to the ℓ_0-norm, albeit a weaker and indirect measure of sparsity [4]. However, for general $K \in \{1, \ldots, M\}$, directly applying the ℓ_1-norm per $\mathbf{w}_k$ does not imply antenna selection. Indeed, replacing the non convex ℓ_0-norm in the objective function of (3) with $\sum_{k=1}^K \|\mathbf{w}_k\|_1$ would result in a sparse solution for each $\mathbf{w}_k$, but the zero entries of each $\mathbf{w}_k$ will not necessarily align to the same antenna(s) n to be omitted. Therefore, it is crucial to utilize a regularization norm that explicitly promotes sparsity for all the entries of $\bar{\mathbf{w}}_n$ simultaneously.

The widely used *group-sparsity* promoting regularization, which was first introduced in the context of the group *least-absolute selection and shrinkage operator* (group *Lasso*) [24], is the mixed $\ell_{1,2}$-norm, defined as

$$\|\mathbf{w}\|_{1,2} := \sum_{n=1}^{N} \|\bar{\mathbf{w}}_n\|_2.$$

Note that $\|\mathbf{w}\|_{1,2} = \|\tilde{\mathbf{w}}\|_1$. The $\ell_{1,2}$-norm behaves as the ℓ_1-norm on $\tilde{\mathbf{w}}$, which implies that each $\|\bar{\mathbf{w}}_n\|_2$ (or equivalently $\bar{\mathbf{w}}_n$) is encouraged to be set to zero, therefore inducing group-sparsity. More generally, it has been shown that mixed $\ell_{1,q}$-norms, defined as

$$\|\mathbf{w}\|_{1,q} := \sum_{n=1}^{N}\left(\sum_{k=1}^{K}|w_k(n)|^q\right)^{1/q}$$

induce group sparsity for $q > 1$ [1]. Setting $q = 1$ yields $\sum_{k=1}^{K}\|\mathbf{w}_k\|_1$, which does not induce group sparsity.

Next, we argue that it is possible to replace any sparsity-inducing norm regularization with the squared norm without changing the regularization properties of the problem. Define the convex function $f(\mathbf{w}) := \sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2$, and define $\Omega(\mathbf{w}) := \|\mathbf{w}\|_{1,q}$ for $q > 1$ as any convex sparsity-inducing norm that replaces the ℓ_0-(quasi)norm in (3). Problem (3) can thus be generically written as

$$\min_{\mathbf{w}\in\mathcal{F}} f(\mathbf{w}) + \lambda\Omega(\mathbf{w}) \tag{4}$$

where

$$\mathcal{F} := \left\{\{\mathbf{w}_k \in \mathbb{C}^N\}_{k=1}^{K} : \frac{|\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m|^2}{\sum\limits_{l\neq k}|\mathbf{w}_l^{\mathcal{H}}\mathbf{h}_m|^2 + \sigma_m^2} \geq \gamma_m, \right.$$
$$\left. \forall m \in G_k,\ \forall k,l \in \{1,\ldots,K\}\right\}.$$

Problem (4) is equivalent to

$$\min_{\mathbf{w}\in\mathcal{F}} f(\mathbf{w}) \qquad \text{s.t.} \quad \Omega(\mathbf{w}) \leq \tau \tag{5}$$

since for any λ, one can find a τ such that the both problems yield the same optimum sparse solution. By squaring both sides of the constraint, problem (5) can be written as

$$\min_{\mathbf{w}\in\mathcal{F}} f(\mathbf{w}) \qquad \text{s.t.} \quad \Omega^2(\mathbf{w}) \leq \bar{\tau} \tag{6}$$

where $\bar{\tau} := \tau^2$. If the Pareto boundary is convex, then there exists a $\bar{\lambda} \geq 0$ such that problem

$$\min_{\mathbf{w}\in\mathcal{F}} f(\mathbf{w}) + \bar{\lambda}\Omega^2(\mathbf{w}) \tag{7}$$

is equivalent to (6) [3, Section 2.6.3], i.e., (7) is just a re-parametrization of (4). This is always true for convex problems, e.g., the Lasso[1] [20], suggesting that $\Omega^2(\mathbf{w})$ can be used as a sparsity-inducing regularization. In our case $\mathcal{F}$ is non-convex, hence convexity of the Pareto boundary is not guaranteed. Still, the above discussion motivates using $\Omega^2(\mathbf{w})$ as a sparsity-inducing penalty in place of the ℓ_0 penalty in (3).

For our purposes, we will use the convex $\ell_{1,\infty}$-norm squared as a group-sparsity inducing regularization to replace the non-convex ℓ_0-norm in (3). The $\ell_{1,\infty}$-norm is defined as

$$\|\mathbf{w}\|_{1,\infty} := \sum_{n=1}^{N}\max_{k=1,\ldots,K}|w_k(n)|.$$

[1]It is also easy to check that the soft thresholding (shrinkage) property of the Lasso holds when the ℓ_1-norm squared is used instead of the ℓ_1-norm to induce sparsity, albeit with a different scaling for the threshold.

The reason why the $\ell_{1,\infty}$-norm squared is used in particular will become clear in the next subsection. Note that if $K = 1$ where no group-sparsity is required, the $\ell_{1,\infty}$-norm reduces to the ℓ_1-norm. The group-sparsity promoting properties of the $\ell_{1,\infty}$-norm were studied in [21]. The joint antenna selection and transmit-power minimization problem (3) can thus be relaxed to

$$\boxed{\begin{aligned} \min_{\{\mathbf{w}_k\in\mathbb{C}^N\}_{k=1}^{K}} \quad & \sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 + \lambda\left(\sum_{n=1}^{N}\max_k|w_k(n)|\right)^2 \\ \text{s.t.:} \quad & \frac{|\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m|^2}{\sum\limits_{l\neq k}|\mathbf{w}_l^{\mathcal{H}}\mathbf{h}_m|^2 + \sigma_m^2} \geq \gamma_m, \\ & \forall m \in G_k, \quad \forall k,l \in \{1,\ldots,K\}. \end{aligned}} \tag{8}$$

Using the mixed $\ell_{1,\infty}$-norm (or equivalently the $\ell_{1,\infty}$-norm squared) as a convex surrogate of the ℓ_0-norm in (3) results in a solution that is no longer necessarily the minimum power solution. This limitation is due to the properties of the ℓ_1- and ℓ_∞-norms. One shortcoming is that the ℓ_1-norm is size-sensitive, whereas the ℓ_0-norm counts the number of nonzero entries without regard to their size. Another issue is that ℓ_∞-norms may have the undesired effect to favor solutions with many components of equal magnitude. The solution of the relaxed problem (8) compromises between minimizing the $\ell_{1,\infty}$- and ℓ_2-norms. This implies that after obtaining an approximate solution to (3), one should solve a reduced-size ℓ_2-norm minimization problem of type (1) as a last step, omitting the antennas corresponding to the zero entries of the sparse approximate solution.

B. Semidefinite Program Formulation

After replacing the ℓ_0-norm in (3) by the $\ell_{1,\infty}$-norm squared, the resulting problem (8) is still NP-hard since it contains (1). In this subsection we show that (8) can be relaxed to a convex semidefinite program (SDP) [22]. SDP problems can be efficiently solved (in polynomial time) using interior point methods. Define $\mathbf{Q}_m := \mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}$, $\mathbf{X} := \mathbf{w}\mathbf{w}^{\mathcal{H}}$ (where $\mathbf{w} := [\mathbf{w}_1^T \ldots \mathbf{w}_K^T]^T$), and $\mathbf{X}_{ij} := \mathbf{w}_i\mathbf{w}_j^{\mathcal{H}}$ for $i,j \in \{1,\ldots,K\}$, such that

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \ldots & \mathbf{X}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{K1} & \ldots & \mathbf{X}_{KK} \end{bmatrix}.$$

Then, the optimization variables can be changed from $\{\mathbf{w}_k\}_{k=1}^{K}$ to $\mathbf{X}$ using the following transformations:

$$\begin{aligned} \|\mathbf{w}_k\|_2^2 &= \mathrm{tr}(\mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}) = \mathrm{tr}(\mathbf{X}_{kk}) \\ |\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m|^2 &= \mathbf{h}_m^{\mathcal{H}}\mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m = \mathrm{tr}(\mathbf{h}_m^{\mathcal{H}}\mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}\mathbf{h}_m) \\ &= \mathrm{tr}(\mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}\mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}) = \mathrm{tr}(\mathbf{Q}_m\mathbf{X}_{kk}). \end{aligned}$$

The $\ell_{1,\infty}$-norm squared is also transformed as follows:

$$\begin{aligned} &\left(\sum_{n=1}^{N}\max_k|w_k(n)|\right)^2 \\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\left((\max_k|w_k(n_1)|)\,.\,(\max_k|w_k(n_2)|)\right) \\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\max_{i,j\in\{1,\ldots,K\}}|X_{ij}(n_1,n_2)|. \end{aligned}$$

Note that $\mathbf{X} = \mathbf{w}\mathbf{w}^{\mathcal{H}}$ if and only if $\mathbf{X} \succeq 0$ and $\mathrm{rank}(\mathbf{X}) = 1$. By dropping the non-convex $\mathrm{rank}(\mathbf{X}) = 1$ constraint, problem (8) can be relaxed to the SDP:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{C}^{NK\times NK}} \quad & \sum_{k=1}^{K}\mathrm{tr}(\mathbf{X}_{kk}) + \lambda\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\max_{i,j}|X_{ij}(n_1,n_2)| \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}_{kk}\mathbf{Q}_m) \geq \gamma_m\sum_{l\neq k}\mathrm{tr}(\mathbf{X}_{ll}\mathbf{Q}_m) + \gamma_m\sigma_m^2, \\ & \mathbf{X}\succeq 0, \quad \forall m\in G_k, \quad \forall k,l\in\{1,\ldots,K\} \\ & \mathbf{X}_{ij} = \mathbf{X}((i-1)N+1{:}iN,(j-1)N+1{:}jN), \\ & \forall i,j\in\{1,\ldots,K\}. \end{aligned} \tag{9}$$

Due to rank relaxation, the off-diagonal matrices $\mathbf{X}_{ij}, \forall i \neq j$, do not appear in the constraints of (9); thus, in light of the cost in (9), they can be set to zero. Hence, using $\mathbf{X}_k = \mathbf{X}_{kk}$ for brevity, and defining $\tilde{\mathbf{X}} := \max_k |\mathbf{X}_k|$ as the element-wise absolute maximum among all $\{\mathbf{X}_k\}_{k=1}^K$, a simplified expression for the $\ell_{1,\infty}$-norm squared is:

$$\begin{aligned} &\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\max_{i,j\in\{1,\ldots,K\}}|X_{ij}(n_1,n_2)| \\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\max_{k\in\{1,\ldots,K\}}|X_k(n_1,n_2)| \\ &= \mathrm{tr}(\mathbf{1}_N\tilde{\mathbf{X}}). \end{aligned} \tag{10}$$

Therefore, the rank-relaxed SDP problem (9) can be re-written as

$$\boxed{\begin{aligned} \min_{\{\mathbf{X}_k\}_{k=1}^K,\tilde{\mathbf{X}}\in\mathbb{R}^{N\times N}} \quad & \sum_{k=1}^{K}\mathrm{tr}(\mathbf{X}_k) + \lambda\mathrm{tr}(\mathbf{1}_N\tilde{\mathbf{X}}) \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}_k\mathbf{Q}_m) \geq \gamma_m\sum_{l\neq k}\mathrm{tr}(\mathbf{X}_l\mathbf{Q}_m) + \gamma_m\sigma_m^2 \\ & \mathbf{X}_k\succeq 0, \quad \tilde{\mathbf{X}}\geq|\mathbf{X}_k|, \quad \forall m\in G_k, \\ & \forall k,l\in\{1,\ldots,K\} \end{aligned}} \tag{11}$$

where the element-wise inequality $\tilde{\mathbf{X}} \geq |\mathbf{X}_k|, \forall k$, can be sustained using positive semidefinite constraints as shown in the Appendix. For the single multicast group case, problem (11) simplifies to

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{X}) + \lambda\mathrm{tr}(\mathbf{1}_N|\mathbf{X}|) \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq \gamma_m\sigma_m^2, \\ & m = 1,\ldots,M, \quad \mathbf{X}\succeq 0. \end{aligned} \tag{12}$$

On the other hand, the power minimization problem (1), without antenna selection, can be relaxed to the SDP:

$$\begin{aligned} \min_{\{\mathbf{X}_k\in\mathbb{C}^{N\times N}\}_{k=1}^K} \quad & \sum_{k=1}^{K}\mathrm{tr}(\mathbf{X}_k) \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}_k\mathbf{Q}_m) \geq \gamma_m\sum_{l\neq k}\mathrm{tr}(\mathbf{X}_l\mathbf{Q}_m) + \gamma_m\sigma_m^2 \\ & \mathbf{X}_k\succeq 0, \quad \forall m\in G_k, \quad \forall k,l\in\{1,\ldots,K\}. \end{aligned} \tag{13}$$

Insights From Duality. To gain some insight on the relationship between (11) and the NP-hard problem (8), we shall invoke duality. The Lagrangian dual problem of (8), which is by definition a convex problem, and the SDP relaxation (11), both provide lower bounds on the optimal value of the NP-hard problem (8). The following result shows that these two lower bounds in fact coincide.

Proposition 1: Problem (11) is the Lagrange bi-dual of problem (8).

Proof: Refer to the Appendix for the complete proof. ∎

Proposition 1 implies that the SDP relaxation (11) yields the same lower bound on the optimal solution of (8) as that obtained from the Lagrangian dual problem, which is the tightest lower bound attainable via duality. The main element of the proof is to reformulate (8) as a QCQP. The dual of a QCQP is an SDP [23, pp. 403-404], which is relatively easy to find. It then follows readily that the dual of this SDP dual problem is the SDP relaxation (11).

To extract the minimum power beamforming vectors corresponding to the selected antennas after solving (11), we use the following procedure. Let $\tilde{\mathbf{X}}^{(s)}$ denote the sparse solution $\tilde{\mathbf{X}}$ of (11). Its zero diagonal entries correspond to the antennas that should be left out, whereas the nonzero ones correspond to the selected antennas. Note that if an entry of $\tilde{\mathbf{X}}^{(s)}$ is zero, then the corresponding entry in all $\{\mathbf{X}_k\}_{k=1}^K$ must be zero. Suppose that the number of nonzero diagonal entries of $\tilde{\mathbf{X}}^{(s)}$ is $\hat{N} \leq N$, and let $S \subseteq \{1,\ldots,N\}$ denote the corresponding subset of antennas that should be utilized, where the cardinality of S is $\hat{N}$. Due to the influence of the mixed $\ell_{1,\infty}$-norm squared minimization, the minimum power beamforming vector cannot be directly extracted from $\tilde{\mathbf{X}}^{(s)}$. Thus, to find the minimum power solution, (13) is solved for the reduced size problem, namely $\{\mathbf{X}_k \in \mathbb{C}^{\hat{N}\times\hat{N}}\}_{k=1}^K$, where $\mathbf{Q}_m$ in this case is an $\hat{N}\times\hat{N}$ matrix obtained after omitting the channel entries corresponding to the left-out antennas. Due to the rank relaxation, the solution to (13), denoted by $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$, might not comprise only rank-one matrices in general; hence, the optimum beamforming vectors cannot be directly extracted from the obtained $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$. However, it is possible to adopt the approach of [8], where an approximate solution to the original problem (1) can be found using a Gaussian randomization technique to generate candidate beamforming vectors from $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$ and choose the ones yielding a feasible solution of minimum power. If $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$ are all rank-one matrices, then their respective principal components, suitably scaled, will be the optimal beamforming vectors for problem (1). Scaling these principal components is a multicast power control problem, which can be optimally solved by linear programming [8].

The sparsest solution (meaning the one with the minimum number of antennas) that can be obtained using this approach corresponds to using $\lambda \rightarrow \infty$ in (11), or equivalently

$$\begin{aligned} \min_{\{\mathbf{X}_k\in\mathbb{C}^{N\times N}\}_{k=1}^K,\tilde{\mathbf{X}}\in\mathbb{R}^{N\times N}} \quad & \mathrm{tr}(\mathbf{1}_N\tilde{\mathbf{X}}) \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}_k\mathbf{Q}_m) \geq \gamma_m\sum_{l\neq k}\mathrm{tr}(\mathbf{X}_l\mathbf{Q}_m) + \gamma_m\sigma_m^2 \\ & \mathbf{X}_k\succeq 0, \quad \tilde{\mathbf{X}}\geq|\mathbf{X}_k|, \\ & \forall m\in G_k, \quad \forall k,l\in\{1,\ldots,K\}. \end{aligned} \tag{14}$$

The use of the *size-sensitive* ℓ_1-norm (or $\ell_{1,\infty}$-norm squared), however, often precludes very sparse solutions, simply because they cost too much in terms of ℓ_1 cost. This motivates adapting the sparsity-enhancing iteratively re-weighted ℓ_1-norm idea, originally proposed in the context of (linear) compressive sampling problems [4], to the present context.

C. Enhancing Sparsity: Iterative Algorithm

To further increase the group-sparsity of $\{\mathbf{w}_k\}_{k=1}^K$, the iteratively re-weighted ℓ_1-norm penalty in [4] is adapted to suit our problem. Consider the weight vector $\mathbf{u}$, where $u(1), u(2), \ldots, u(N)$ are positive weights, and define the weight matrix $\mathbf{U} := \mathbf{u}\mathbf{u}^T$. Using $\mathbf{X}_k := \mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}$ and $\tilde{\mathbf{X}} := \max_k |\mathbf{X}_k|$ as before, and invoking again the implication of rank-relaxation that was previously used to obtain the $\ell_{1,\infty}$-norm squared expression (10), the weighted $\ell_{1,\infty}$-norm squared can be written as

$$\begin{aligned}&\left(\sum_{n=1}^{N} u(n) \max_k |w_k(n)|\right)^2\\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} u(n_1)u(n_2)\max_k |X_k(n_1,n_2)|\\ &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} U(n_1,n_2)\tilde{X}(n_1,n_2)\\ &= \mathrm{tr}(\mathbf{U}\tilde{\mathbf{X}}).\end{aligned}$$

The iterative algorithm that enhances group-sparsity can then be described as follows:

1) *Initialize* the iteration count to $r = 0$, and the weight matrix to $\mathbf{U}^{(0)} = \mathbf{1}_N$.
2) *Solve* the weighted $\ell_{1,\infty}$-norm squared minimization SDP problem

$$\begin{aligned}\min_{\{\mathbf{X}_k^{(r)}\}_{k=1}^K, \tilde{\mathbf{X}}^{(r)}} \quad & \mathrm{tr}(\mathbf{U}^{(r)}\tilde{\mathbf{X}}^{(r)})\\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}_k^{(r)}\mathbf{Q}_m) \geq \gamma_m \sum_{l\neq k}\mathrm{tr}(\mathbf{X}_l^{(r)}\mathbf{Q}_m) + \gamma_m\sigma_m^2\\ & \mathbf{X}_k^{(r)} \succeq 0, \quad \tilde{\mathbf{X}}^{(r)} \geq |\mathbf{X}_k^{(r)}|,\\ & \forall m \in G_k, \quad \forall k,l \in \{1,\ldots,K\}\end{aligned} \tag{15}$$

to obtain the optimum $\{\mathbf{X}_k^{(r)}\}_{k=1}^K$ at the r-th iteration.
3) *Update* the weight matrix entries to be used in the next iteration as

$$U^{(r+1)}(n_1,n_2) = \frac{1}{\tilde{X}^{(r)}(n_1,n_2)+\epsilon}, \ \forall n_1,n_2 \in \{1,\ldots,N\}.$$

4) Terminate on convergence, or, when a certain maximum number of iterations for r is reached. Otherwise, increment r, and go to step 2.

The weight matrix updates force small entries of $\tilde{\mathbf{X}}$ (and thus the corresponding entries of $\{|\mathbf{X}_k|\}_{k=1}^K$) to zero, and avoid unduly restraining large entries. The small parameter ϵ provides stability, and ensures that a zero-valued entry of $\tilde{\mathbf{X}}^{(r)}$ does not strictly prohibit a nonzero estimate at the next step. In the initial step of the iterative algorithm, problem (14) is solved for initialization. Convergence of this algorithm is very fast ($\sim 5-15$ iterations), as observed in the simulations. It is worth reminding the reader that this iterative algorithm is not guaranteed to find the minimum number of antennas that yield a feasible solution of (1); finding such minimum-antenna solution is NP-hard.

IV. Proposed Algorithm

The proposed algorithm that jointly selects $L \leq N$ antennas and finds the beamforming vector for each multicast group such that the transmit-power is minimized, subject to receive-SINR constraints for each user, can be summarized as follows:

- *Step 1:* Run the weighted $\ell_{1,\infty}$-norm iterative algorithm described in Section III-C. Terminate the weighted $\ell_{1,\infty}$-norm iterative algorithm 'prematurely' if a solution comprising L or fewer antennas is encountered during outer iterations. Record the resulting sparse solution $\tilde{\mathbf{X}}^*$ and the corresponding weight matrix $\mathbf{U}^*$. Let $\hat{N}^*$ denote the number of nonzero diagonal entries in $\tilde{\mathbf{X}}^*$. If $\hat{N}^* > L$ after the iterative algorithm terminates, then the proposed algorithm fails to provide a sparse-enough solution. Brute-force enumeration can be used in this case to find a solution, if the problem is feasible. If $\hat{N}^* = L$, then pick S to contain the antennas corresponding to the nonzero diagonal entries of $\tilde{\mathbf{X}}^*$ and skip to step 3. Otherwise, continue.
- *Step 2:* Solve the SDP problem

$$\boxed{\begin{aligned}&\min_{\{\mathbf{X}_k\}_{k=1}^K, \tilde{\mathbf{X}}\in\mathbb{R}} \textstyle\sum_{k=1}^K \mathrm{tr}(\mathbf{X}_k) + \lambda\mathrm{tr}(\mathbf{U}^*\tilde{\mathbf{X}})\\ &\text{s.t.:} \quad \mathrm{tr}(\mathbf{X}_k\mathbf{Q}_m) \geq \gamma_m \sum_{l\neq k}\mathrm{tr}(\mathbf{X}_l\mathbf{Q}_m) + \gamma_m\sigma_m^2\\ &\qquad \mathbf{X}_k \succeq 0, \quad \tilde{\mathbf{X}} \geq |\mathbf{X}_k|,\\ &\qquad \forall m \in G_k, \quad \forall k,l \in \{1,\ldots,K\}\end{aligned}} \tag{16}$$

  using the obtained weights $\mathbf{U}^*$, which is problem (11) with $\mathbf{1}_N$ replaced by $\mathbf{U}^*$, and use binary search to find λ that gives the required number of antennas L. The binary search procedure works as follows. For a given upper bound λ_{UB} and lower bound λ_{LB}, set $\lambda = (\lambda_{UB} - \lambda_{LB})/2 + \lambda_{LB}$ and solve the SDP problem. Let $\tilde{\mathbf{X}}^{(s)}$ denote the solution of (16) having $\hat{N}$ nonzero diagonal entries. If $\hat{N} = L$, then find the subset of selected antennas S corresponding to the nonzero diagonal entries of $\tilde{\mathbf{X}}^{(s)}$, and move to the next step. Otherwise, if $\hat{N} > L$ then set $\lambda_{LB} = \lambda$ while if $\hat{N} < L$ then set $\lambda_{UB} = \lambda$, and repeat this step until $\hat{N} = L$.
- *Step 3:* Now that L antennas have been selected, (13) is solved for the reduced-size problem, namely $\{\mathbf{X}_k \in \mathbb{C}^{L\times L}\}_{k=1}^K$, to find the minimum power beamforming vector. If the solution, denoted as $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$, contains only rank-one matrices, then the (suitably scaled [8]) principal component of each $\mathbf{X}_k^{(o)}$ is the optimal beamforming vector for group k. Otherwise, use the randomization technique of [8] to generate candidate sets of beamforming vectors from $\{\mathbf{X}_k^{(o)}\}_{k=1}^K$, and choose the set that yields a minimum power solution among all feasible ones.

Note that early termination of the binary search when a solution with fewer than the desired L antennas has been obtained will result in higher transmission power. Since λ is non-negative, λ_{LB} can simply be set to zero. Suitable λ_{UB} can be obtained empirically, depending primarily on ϵ (since the value of entries of $\mathbf{U}^*$ that correspond to zero entries of $\tilde{\mathbf{X}}^*$ is $1/\epsilon$ as a result of the updating step of the iterative sparsity-enhancing algorithm), in addition to the network parameters N, M, K, and the channel statistics.

Although the binary search over λ may require solving (16) more than once for different values of λ until the appropriate one is found, an important advantage over the exhaustive search method is that the number of iterations is independent of N and L, unlike exhaustive search, which requires solving $\binom{N}{L}$ problems of type (13). The solution obtained using the novel algorithm occasionally coincides with that obtained using exhaustive search, while the transmission power increase for the other cases is insignificant, as demonstrated in the simulations of Section VI.

Complexity analysis. Following [8], the worst-case complexity of solving the SDP problem (13) using interior point methods is $O(\sqrt{KN}\log(1/\varepsilon))$ iterations, where ε represents the accuracy of the solution at the algorithm's termination, and each iteration requires at most $O(K^3N^6 + MKN^2)$ arithmetic operations. The actual runtime complexity scales much slower with K, N, M than this worst-case bound predicts. The SDP problem (16) includes an additional $N \times N$ auxiliary matrix and KN^2 positive semidefinite constraints (as shown in the appendix), that increase the actual runtime of (16) as compared to that of (13). However, the worst-case complexity order remains the same.

Let $O(R)$ denote the runtime complexity of problem (16) (same as (11) and (15)), where R is a function of K, N, M, and consider the complexity analysis of each of the three steps of the proposed algorithm. In step 1, the weighted $\ell_{1,\infty}$-norm iterative algorithm typically terminates within less than 15 iterations, irrespective of the problem size. An SDP of type (15) is solved in each iteration. Thus the total complexity of this step is $O(R)$. In step 2, the binary search can be considered of constant complexity order. The number of binary searches is typically very small with the proper choice of λ_{UB}, as shown in the simulations of Section VI. In each iteration, an SDP of type (16) is solved. Hence the total complexity of this step is also $O(R)$. In step 3, one SDP of type (13) is solved (replacing N with L), with a runtime complexity that is less than $O(R)$. Finally, the randomization technique that may be used to obtain the beamforming vectors has been analyzed in [8], where it is shown that an ε-optimal solution can be obtained in $O(\sqrt{K}\log(1/\varepsilon))$ iterations, each requiring at most $O(K^3 + MK)$ arithmetic operations. Thus, the overall worst-case complexity of the proposed 3-step algorithm is $O\left((K^{3.5}N^{6.5} + MK^{1.5}N^{2.5})\log(1/\varepsilon)\right)$.

V. Relevant Extensions

The proposed novel algorithm can easily be extended and applied to obtain sparse solutions for a wide class of non-convex QCQP problems, where SDP relaxation is relevant. MIMO detection and sensor network localization are two such applications. For further details on applications where the SDP relaxation is used, the reader is referred to [11] and references therein. In this section we discuss two important variations to the multicast beamforming problem, where our proposed approach can also be applied.

A. Limiting Inter-Cell or Primary User Interference

Suppose there is only one multicast group ($K = 1$), and consider joint antenna selection and beamformer design to minimize the transmit-power, subject to prescribed receive-SNR constraints γ_m for each user. In addition, consider that the interference induced to J other users must not exceed a given threshold η. The channel vector from the transmit antennas to the receive antenna of user j is denoted by $\mathbf{g}_j$, $j \in \{1, \ldots, J\}$, and is assumed known at the transmitter BS. The joint problem is expressed as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^N} \quad & \|\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_0 \\ \text{s.t.:} \quad & \frac{|\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2}{\sigma_m^2} \geq \gamma_m, \quad m = 1, \ldots, M \\ & |\mathbf{w}^{\mathcal{H}}\mathbf{g}_j|^2 \leq \eta, \quad j = 1, \ldots, J \end{aligned} \tag{17}$$

which is the same as (3) with the additional interference constraints for J users. Problem (17) appears in two main scenarios: inter-cell interference mitigation in a co-channel cellular multicast setting, and secondary multicasting in a cognitive underlay setting, where there is a need to limit interference inflicted to primary users. These scenarios have been considered in [16], without antenna selection. Similar to [16], our formulation can be suitably modified to handle cases where only imperfect channel state information is available at the BS, in the form of channel estimates with norm-bounded errors.

Returning to (17), upon replacing $\|\mathbf{w}\|_0$ by $\|\mathbf{w}\|_1^2$ and using the same semidefinite relaxations discussed in Section III, problem (17) can be relaxed to the SDP:

$$\boxed{\begin{aligned} \min_{\mathbf{X}\in\mathbb{C}^{N\times N}} \quad & \mathrm{tr}(\mathbf{X}) + \lambda\mathrm{tr}(\mathbf{1}_N|\mathbf{X}|) \\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m) \geq \sigma_m^2\gamma_m, \ m = 1, \ldots, M \\ & \mathrm{tr}(\mathbf{X}\tilde{\mathbf{Q}}_j) \leq \eta, \ j = 1, \ldots, J, \quad \mathbf{X} \succeq 0 \end{aligned}} \tag{18}$$

where $\tilde{\mathbf{Q}}_j := \mathbf{g}_j\mathbf{g}_j^{\mathcal{H}}$. To select $L \leq N$ antennas, the proposed algorithm in Section IV can be directly applied after adding the constraints $\mathrm{tr}(\mathbf{X}\tilde{\mathbf{Q}}_j) \leq \eta$ for $j = 1, \ldots, J$ to all the SDP problems solved. For the final step, the randomization algorithm proposed in [16] can be used to find the minimum power beamforming vector corresponding to the selected antennas.

B. Max-Min Fair Beamforming

We now consider the related joint problem of maximizing the minimum received SNR over all users together with antenna selection, subject to a bound P on the transmission power (assuming one multicast group for simplicity):

$$\begin{aligned} \max_{\mathbf{w}\in\mathbb{C}^N} \quad & \left(\min_m \left\{\frac{|\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2}{\sigma_m^2}\right\}_{m=1}^M\right) \\ \text{s.t.:} \quad & \|\mathbf{w}\|_2^2 \leq P, \quad \|\mathbf{w}\|_0 \leq L. \end{aligned} \tag{19}$$

Problem (19) is equivalent to maximizing the beamforming downlink achievable rate using L out of N antennas, since in the multicast scenario, the worst-user SNR determines the common (multicast) rate [18]. Problem (19) was studied in [18] without the $\|\mathbf{w}\|_0 \leq L$ constraint, and was shown to be NP-hard. Problem (19) can be equivalently re-written as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{C}^N, t\in\mathbb{R}} \quad & -t + \lambda\|\mathbf{w}\|_0 \\ \text{s.t.:} \quad & \frac{|\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2}{\sigma_m^2} \geq t,\ m = 1,\ldots,M \\ & \|\mathbf{w}\|_2^2 = P. \end{aligned} \tag{20}$$

Following the same approximation steps as in Section III, problem (20) can be relaxed to the SDP:

$$\boxed{\begin{aligned} \min_{\mathbf{X}\in\mathbb{C}^{N\times N}, t\in\mathbb{R}} \quad & -t + \lambda\text{tr}(\mathbf{1}_N|\mathbf{X}|) \\ \text{s.t.:} \quad & \text{tr}(\mathbf{X}\mathbf{Q}_m) \geq \sigma_m^2 t,\ m = 1,\ldots,M \\ & \text{tr}(\mathbf{X}) = P, \quad \mathbf{X} \succeq 0. \end{aligned}} \tag{21}$$

To select $L \leq N$ antennas, the proposed algorithm in Section IV can be applied by solving the appropriate SDPs of type (21), and using the randomization algorithm proposed in [18] in the final step to extract the beamforming vector.

In closing this section, two remarks are in order on the relations between maximizing the minimum received SNR (19), the capacity of the multicast channel [7], and the antenna selection with spatial multiplexing scheme in [15]:

Remark 1: Defining $\mathbf{X}$ as the covariance of the transmitted signal, the optimal solution $\mathbf{X}^*$ to the rank-relaxed SDP problem (21), without the sparsity inducing term ($\lambda\text{tr}(\mathbf{1}_N|\mathbf{X}|)$), is the optimal covariance that achieves the capacity of the multicast channel (maximum achievable common rate) for an N-antenna BS with full CSI at the transmitter [7]. Whereas exhaustive search is required to achieve capacity when only $L < N$ antennas are utilized, the proposed algorithm in Section IV can be used to obtain an approximate, less complex, solution (by solving the appropriate SDPs of type (21)). The only difference between the multicast beamforming rate maximization and the multicast channel capacity is that $\mathbf{X}$ is restricted to be rank one with beamforming (and the randomization algorithm proposed in [18] is needed to extract the beamforming vector from the optimal $\mathbf{X}$), whereas there is no such restriction (and no approximation) for the capacity-achieving transmit covariance. The role of the rank restriction and the use of the sparsity inducing ℓ_1-norm squared approximation are illustrated in Section VI-C.

Remark 2: In the absence of CSI at the transmitter, the alternative is to transmit using a spatially white covariance, i.e., $\mathbf{X} = \frac{P}{N}\mathbf{I}_N$, where P is the total transmission power and $\mathbf{X}$ denotes the covariance of the transmitted signal [7]. An antenna selection scheme has been proposed in [15] for maximizing the minimum received SNR based on this setup. When utilizing a subset of antennas $\mathcal{S}$ of size L, the transmission power is equally divided among all L antennas yielding an SNR for the m-th user $\text{SNR}_m = \frac{P}{L}\frac{\sum_{n\in\mathcal{S}}|h_m(n)|^2}{\sigma_m^2}$. From all possible antenna subsets $\mathcal{S}$ of size L, the selected subset $\mathcal{S}^*$ is the one maximizing the minimum SNR across all users, namely

$$\mathcal{S}^* = \arg\max_{\mathcal{S}\in\tilde{\mathcal{S}}} \min_{m\in\{1,\ldots,M\}} \frac{P}{L}\frac{\sum_{n\in\mathcal{S}}|h_m(n)|^2}{\sigma_m^2}$$

where $\tilde{\mathcal{S}}$ is the set of all $\binom{N}{L}$ possible antenna subsets of size L. This antenna selection scheme requires knowledge of the channel gain corresponding to each transmit antenna at the transmitter ($|h_m(n)|^2$) for each user, in addition to exhaustively searching through all $\binom{N}{L}$ different antenna subset $\mathcal{S}$ selections. The results of [7] imply that transmitting with spatially white covariance will outperform beamforming (in terms of spectral efficiency) when $M \gg L$, because every beamforming direction will likely be nearly orthogonal to at least one user's channel, whereas beamforming performs significantly better (very close to the multicast capacity) for relatively large L. Attaining this rate with spatially white covariance is a challenge since it requires complex multi-stream Shannon encoding and decoding at long block lengths, also implying long decoding delay that is not suitable for streaming media multicast. Beamforming, on the other hand, requires far simpler encoding and decoding. The performance of our proposed beamforming based algorithm is compared with that of [15] in Section VI-C.

VI. Simulated Tests

To test the proposed SDP-based algorithms, YALMIP was used. YALMIP is a modeling language for optimization problems that is implemented as a free toolbox for MATLAB [10], and uses SeDuMi, a MATLAB implementation of second-order interior-point methods, for the actual computations [19]. The novel algorithm was tested with two channel types; Rayleigh fading channels and Vandermonde channels corresponding to a far-field ULA setup. Throughout this section, the noise variance for all users was set to $\sigma^2 = 1$.

A. Single Multicast Group

We first consider a single multicast group, and set the minimum required SNR to $\gamma = 1$ at all users.

Rayleigh fading with $N = 8$ antennas. The first simulation setup included a BS with $N = 8$ transmit-antennas broadcasting a common message to $M = 16$ receivers. Independent identically distributed (i.i.d.) Rayleigh fading channel vectors $\{h_{\mathrm{m}}\}_{\mathrm{m}=1}^{M}$ were generated, each with i.i.d. entries circularly symmetric zero-mean complex Gaussian random variables of variance 1. To gain insight, detailed results are provided first for a single "typical" channel realization, which allows comparing the selected antenna subsets with the baseline exhaustive search solution. Running the weighted ℓ_1-norm iterative algorithm described in Section III-C results in the sparsest solution of $\hat{N}^* = 1$ antenna, which corresponds to selecting antenna number 5. This result is obtained when the iterative algorithm converges after 8 iterations. It is worth noting that after the initial step of the iterative weighted ℓ_1-norm algorithm (which is equivalent to solving problem (14)), the resulting sparse solution has $\hat{N} = 6$ antennas, many more than the single antenna

TABLE I
PERFORMANCE OF THE PROPOSED ALGORITHM AND THE EXHAUSTIVE SEARCH ALGORITHM FOR A PARTICULAR CHANNEL REALIZATION FOR DIFFERENT ANTENNA SELECTIONS L, FOR $N=8$ ANTENNA BS AND A SINGLE MULTICAST GROUP WITH $M=16$ USERS

L	Proposed Algorithm				Exhaustive Search			
	Selected antennas S	Power (dBm)	Power inc. (dB)	Total SDP's	Selected antennas S	Power (dBm)	Power inc. (dB)	Total SDP's
8	{1,...,8}	30.82	0	1	{1,...,8}	30.82	0	1
7	{1,2,3,4,5,7,8}	31.16	0.34	9	{1,2,3,4,5,7,8}	31.16	0.34	8
6	{2,3,4,5,7,8}	31.81	0.99	2	{1,2,4,5,7,8}	31.75	0.93	28
5	{2,3,4,5,7}	31.84	1.02	5	{2,4,5,7,8}	31.80	0.98	56
4	{2,4,5,7}	31.92	1.11	3	{2,4,5,7}	31.92	1.11	70
3	{2,5,7}	32.42	1.6	6	{2,5,7}	32.42	1.6	56
2	{5,7}	33.86	3.04	7	{5,7}	33.86	3.04	28
1	{5}	35.78	4.96	8	{5}	35.78	4.96	8

solution obtained after the iterative weighted ℓ_1-norm algorithm terminates.

Table I summarizes the results obtained using the novel algorithm and by exhaustively searching over possible antenna subset selections for this representative channel realization. The required number of antennas to be selected (or, the available number of RF chains) L is listed in column 1. The subset of selected antennas is given in columns 2 and 6 for the proposed algorithm and exhaustive search, respectively. The minimum transmit-power corresponding to each L is listed in columns 3 and 7 (in dBm units). The increase of transmission power (compared to the case of using all $N=8$ antennas) due to antenna selection is given in columns 4 and 8 (in dB units). Finally, the total number of SDP problems solved in order to obtain the required solution is shown in columns 5 and 9.

The results in Table I demonstrate that as the number of antennas selected for transmission decreases (as the solution becomes more sparse) the corresponding minimum transmission power increases, due to the decrease in degrees of freedom, as expected. Interestingly, the simulations suggest that the number of transmit antennas can be significantly reduced at only a small price in terms of excess transmission power. Halving the number of antennas from 8 to 4, for example, entails only 1.11 dB extra power. Comparing with the exhaustive search results, one can verify that exhaustive search slightly outperforms the proposed algorithm only for the cases of $L=5$ and $L=6$ antennas (by less than 0.1 dB), by selecting different antenna subsets. However, the number of SDP problems that must be solved for the exhaustive search is significantly larger. The maximum number of iterations required for the binary search process, namely step 2 in the proposed algorithm, is 7—these are needed to select $L=7$ antennas, where 1 SDP problem is solved for step 1, 7 for step 2, and 1 for the final step, yielding a worst-case total of 9 SDP problems. On the other hand, the exhaustive search algorithm requires solving $\binom{8}{4}=70$ SDP problems to select $L=4$ antennas.

Table II reports the average and maximum increase in transmission power (compared to the case of using all $N=8$ antennas) that correspond to selecting L antennas for the proposed algorithm, the exhaustive search, and the case where the number of available antennas is only L (not N) such that no antenna selection is performed (this is equivalent to randomly selecting the L antennas). In addition, the average and maximum number of

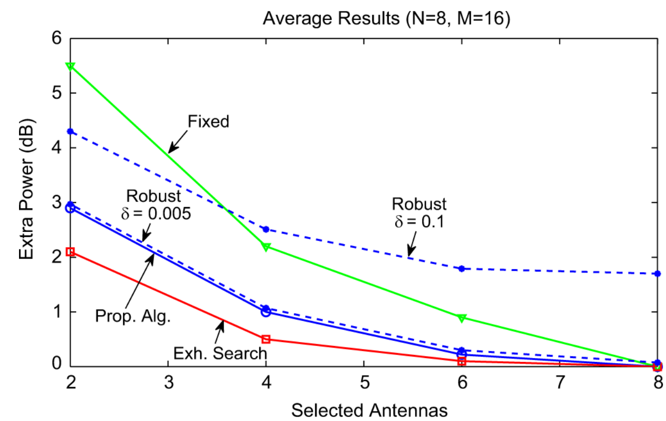

Fig. 1. The necessary extra power versus L for $N=8$ antennas and a single multicast group with $M=16$ users in a Rayleigh fading environment.

SDP problems solved for the proposed algorithm and exhaustive search are reported. For a better visual comparison, Fig. 1 plots the average increase in transmission power versus L for the compared schemes (corresponding to columns 2, 6 and 9 of Table II). The results are obtained for 100 different Rayleigh channel realizations. The main conclusions from Table II and Fig. 1 are summarized as follows:

1) The number of transmit antennas can be considerably reduced at a relatively small cost in terms of excess transmission power. If we halve the number of antennas, the transmission power increases by only 1 dB, on average, to satisfy the SNR constraints using the proposed algorithm.
2) Compared to the exhaustive search, the proposed algorithm incurs much lower complexity (measured in terms of the number of SDP problems solved) at a very small additional power cost. The difference in power is less than 1 dB, on average.
3) If only L RF transmission chains are available at the BS, increasing the number of transmit antennas N (from which only L are activated) results in a reduction in transmission power due to the additional diversity. For example, if only 4 RF chains and 4 antennas are available $(N=L=4)$, 1.2 dB more transmission power is required compared to having the option of selecting 4 out of 8 antennas using the proposed algorithm, on average.

Rayleigh fading with $N=16$ antennas. In Fig. 2, we consider $N=16$ antennas and $M=32$ users, again assuming

TABLE II
PERFORMANCE COMPARISON BETWEEN THE PROPOSED ALGORITHM, EXHAUSTIVE SEARCH AND NO ANTENNA SELECTION, FOR $N = 8$ ANTENNA BS AND A SINGLE MULTICAST GROUP WITH $M = 16$ USERS IN A RAYLEIGH FADING ENVIRONMENT

L	Proposed Algorithm				Exhaustive Search			Fixed Antennas	
	Power Inc. (dB)		Total SDP's		Power Inc. (dB)		Total SDP's	Power Inc. (dB)	
	Avg.	Max.	Avg.	Max.	Avg.	Max.		Avg.	Max.
6	0.22	2.7	7.5	17	0.1	1.12	28	0.9	3.1
4	1	3.2	4.3	11	0.5	2.3	70	2.2	4.8
2	2.9	6.7	4.4	9	2.1	6.7	28	5.5	11.8

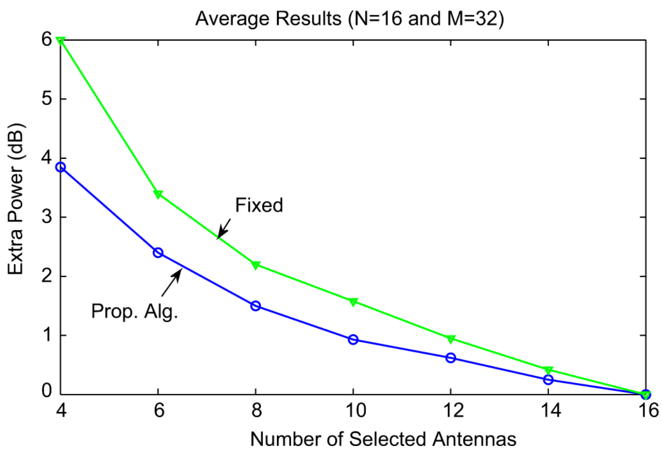

Fig. 2. The necessary extra power versus L for $N = 16$ antennas and a single multicast group with $M = 32$ users in a Rayleigh fading environment.

i.i.d. Rayleigh fading across antennas and users. The figure depicts the average increase in transmission power (compared to the case of using all $N = 16$ antennas) versus L. If we halve the number of selected antennas ($L = 8$), the transmission power increases by only 1.5 dB to satisfy the SNR constraints using the proposed algorithm, whereas if only 8 antennas were installed instead of 16 (i.e., no antenna selection), an additional 0.7 dB transmission power would be necessary (compared to the proposed algorithm), on average. The results for the exhaustive search algorithm are not included because of its prohibitive complexity. If, for example, it is required to select $L = 8$ antennas, exhaustive search requires solving $\binom{16}{8} = 12{,}870$ SDP problems per channel realization, which is clearly prohibitive. On the other hand, the proposed algorithm required solving less than 7 SDP problems for $L = 8$, on average.

Rayleigh fading with $N = 100$ antennas. In Fig. 3, we consider a scenario with a large number of antennas and users ($N = 100$, $M = 100$), again assuming i.i.d. Rayleigh fading. The figure shows the average additional transmit-power needed using the proposed algorithm, which is 1–2 dB less than the transmit-power needed when the first L antennas are blindly selected, for all values of L considered. Read in a different way, the proposed algorithm uses far fewer transmit antennas for the same transmit-power. Of course, it is computationally prohibitive to apply exhaustive search in this scenario. Note that the gains offered by the proposed algorithm are relatively small when the number of users is relatively large and the channel is i.i.d. across antennas and users—because the law of large numbers kicks in. The situation is different when M is small. For example, with $N = 100$ antennas to choose from, $M = 2$ users, and $L = 2$ antennas to be selected, the maximum transmit-power using our proposed algorithm, over 1000 Rayleigh channel realizations, was 34.7 dBm (23.9 dBm on average), whereas the maximum transmit-power when blindly selecting the first 2 antennas was 58.9 dBm (30.3 dBm on average). This means that the proposed algorithm can save up to approximately 24 dB in transmit-power compared to fixed antenna selection in this setting.

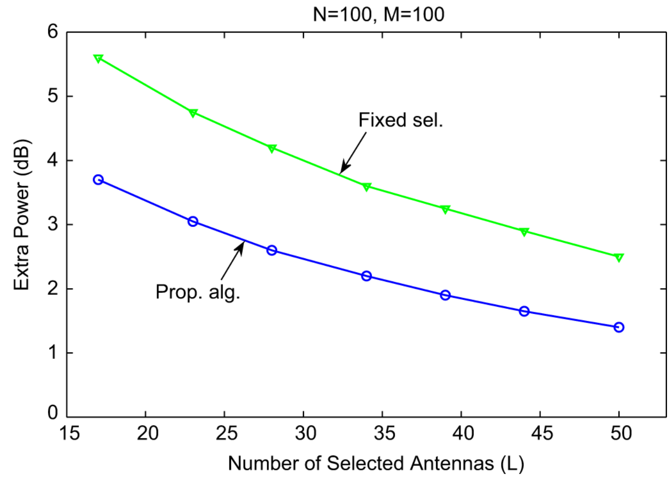

Fig. 3. The necessary extra power versus L with $N = 100$ and $M = 100$ in a Rayleigh fading environment.

Far-field beamforming with $N = 8$ ULA. Fig. 4 illustrates the beampatterns for a particular far-field multicasting scenario with $N = 8$ ULA antennas and $M = 33$ users. The $N \times 1$ complex channel vector for each user m is Vandermonde: $\mathbf{h}_m = [1, e^{j\theta_m}, e^{j2\theta_m}, \ldots, e^{j(N-1)\theta_m}]^T$, where the angles θ_m are given by $\theta_m = -2\pi\frac{d}{\lambda}\sin(\phi_m)$, with d denoting inter-element spacing between successive antennas, λ the carrier wavelength, and the angles ϕ_m define the directions of the receivers. We set $d/\lambda = 1/2$ and the $M = 33$ users were allocated such that the direction angles of the first 11 users ϕ_m, $m \in \{1, \ldots, 11\}$, were from 0^o to 10^o with 1^o spacing between each user, the direction angles to second 11 users ϕ_m, $m \in \{12, \ldots, 22\}$, were from 40^o to 50^o with 1^o spacing between each user, and the direction angles to last 11 users ϕ_m, $m \in \{23, \ldots, 33\}$, were from 80^o to 90^o with 1^o spacing between each user. Fig. 4 compares the beampatterns resulting from: (a) using all $N = 8$ antennas, (b) using the proposed algorithm to select $L = 4$ antennas, and (c) using exhaustive search to select the best $L = 4$ antennas. The proposed algorithm selects the antennas $S_{\text{Prop.}} = \{3, 4, 5, 6\}$ after solving 3 SDP problems and incurs additional transmit-power of 1.38 dB (compared to using all 8 antennas), whereas the exhaustive search selects the antennas $S_{\text{Exh.}} = \{1, 5, 7, 8\}$ after solving 70 SDP problems and incurs 0.8 dB extra power.

Robust beamforming. In case of imperfect CSI, it is possible to adopt robust beamforming designs such as those considered in [16], which rely on the notion of worst-case design. It is assumed in [16] that all channel vectors are known with certain errors $\mathbf{v}$, and that these errors are all norm-bounded $\|\mathbf{v}\|_2 \leq \delta$, where δ is known. The worst-case SINR constraint for user m in multicast group k can be expressed as:

$$\min_{\|\mathbf{v}_m\|_2 \leq \delta} \frac{|\mathbf{w}_k^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2}{\sum\limits_{l \neq k} |\mathbf{w}_l^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2 + \sigma_m^2} \geq \gamma_m.$$

TABLE III
PERFORMANCE COMPARISON BETWEEN THE PROPOSED ALGORITHM, EXHAUSTIVE SEARCH AND NO ANTENNA SELECTION, FOR $N = 12$ ANTENNA BS AND TWO MULTICAST GROUPS WITH $|G_1| = 5$ AND $|G_2| = 5$ IN A RAYLEIGH FADING ENVIRONMENT

L	Proposed Algorithm				Exhaustive Search			Fixed Antennas	
	Power Inc. (dB)		Total SDP's		Power Inc. (dB)		Total SDP's	Power Inc. (dB)	
	Avg.	Max.	Avg.	Max.	Avg.	Max.		Avg.	Max.
10	0.43	0.7	9	20	0.27	0.4	66	0.85	2
8	1	1.55	11.5	25	0.81	0.97	495	2.1	3.35
6	2.4	2.8	7.5	22	1.7	2.3	924	3.8	6.7
4	4.4	5.92	5.5	25	3.7	4.3	495	7.6	9.8

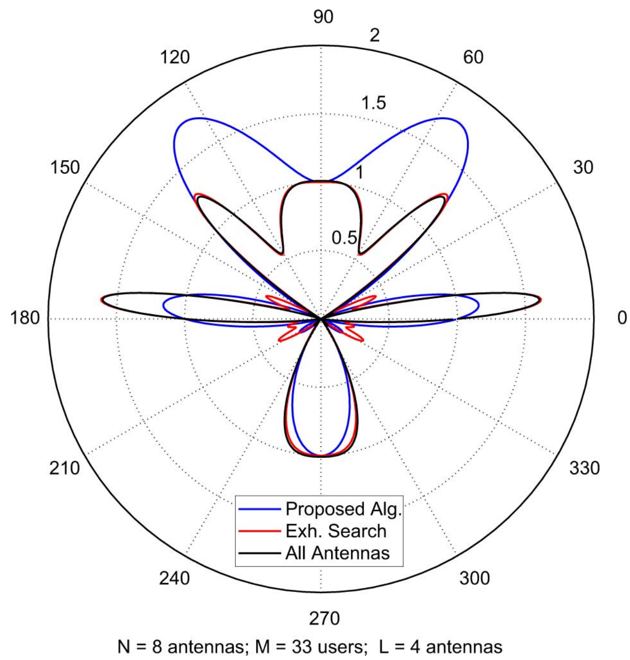

Fig. 4. Beampatterns for a far-field single-group multicasting scenario with $N = 8$ ULA and $M = 33$ users, comparing between (a) using all $N = 8$ antennas, (b) the proposed algorithm selecting $L = 4$ antennas, and (c) exhaustive search to select the best $L = 4$ antennas.

The worst-case SINR is lower bounded by

$$\min_{\|\mathbf{v}_m\|_2 \le \delta} \frac{|\mathbf{w}_k^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2}{\sum_{l \neq k} |\mathbf{w}_l^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2 + \sigma_m^2} \ge \frac{\min_{\|\mathbf{v}_m\|_2 \le \delta} |\mathbf{w}_k^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2}{\sum_{l \neq k} \max_{\|\mathbf{v}_m\|_2 \le \delta} |\mathbf{w}_l^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2 + \sigma_m^2}.$$

Using the bounds developed in [16],

$$\mathbf{w}_k^{\mathcal{H}} \mathbf{G}_m \mathbf{w}_k \le \min_{\|\mathbf{v}_m\|_2 \le \delta} |\mathbf{w}_k^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2$$
$$\mathbf{w}_k^{\mathcal{H}} \tilde{\mathbf{G}}_m \mathbf{w}_k \ge \max_{\|\mathbf{v}_m\|_2 \le \delta} |\mathbf{w}_k^{\mathcal{H}}(\mathbf{h}_m + \mathbf{v}_m)|^2$$

where $\mathbf{G}_m := \mathbf{h}_m \mathbf{h}_m^{\mathcal{H}} + \delta(\delta - 2\sqrt{\mathbf{h}_m^{\mathcal{H}} \mathbf{h}_m})\mathbf{I}$ and $\tilde{\mathbf{G}}_m := \mathbf{h}_m \mathbf{h}_m^{\mathcal{H}} + \delta(\delta + 2\sqrt{\mathbf{h}_m^{\mathcal{H}} \mathbf{h}_m})\mathbf{I}$, the robust SINR constraints can be approximated by

$$\frac{\mathbf{w}_k^{\mathcal{H}} \mathbf{G}_m \mathbf{w}_k}{\sum_{l \neq k} \mathbf{w}_l^{\mathcal{H}} \tilde{\mathbf{G}}_m \mathbf{w}_l + \sigma_m^2} \ge \gamma_m, \quad \forall\, m,\ k,\ l.$$

The robust beamforming with antenna selection algorithm then proceeds simply by replacing $\mathbf{Q}_m$ with $\mathbf{G}_m$ or $\tilde{\mathbf{G}}_m$ in the SDP problem formulations. This robust algorithm was applied to the setup of Fig. 1 yielding the average additional transmit-power represented by the dotted lines. To simulate imperfect CSI, two scenarios were considered where the error vectors $\{\mathbf{v}_m\}$ were uniformly and randomly generated in a sphere centered at zero with radii $\delta = 0.005$ and $\delta = 0.1$. With δ increasing, more transmit-power is needed to satisfy the SNR requirements, as expected.

B. Two Multicast Groups

We now switch to simulations for the multi-group case, with $K = 2$ groups for clarity of exposition.

Rayleigh fading with $N = 12$ antennas. In this setup, we consider a BS with $N = 12$ transmit-antennas transmitting to two multicast groups where each multicast group consists of 5 users. The minimum required receive SINR is assumed to be 1 dB for each user in each group, and Rayleigh fading channel vectors are generated. Table III reports the average and maximum increase in transmission power (compared to using all $N = 12$ antennas) that correspond to selecting L antennas, for the proposed algorithm, the exhaustive search, and the case where only L antennas are available such that no antenna selection is performed. In addition, the average and maximum number of SDP problems solved for the proposed algorithm and exhaustive search are reported. The results emphasize the conclusions obtained for the single multicast group. For example, if only $L = 6$ RF chains are available, using the proposed algorithm results in a 2.4 dB transmission power increase (compared to using all $N = 12$ antennas), and requires solving 7.5 SDP problems, on average. On the other hand, exhaustive search results in 0.7 dB lesser transmission power (on average), but requires solving 924 SDP problems. Finally, having the option of selecting 6 antennas out of 12 saved 1.4 dB in transmission power (using the proposed algorithm) compared to having only 6 antennas available.

To illustrate the effects of the minimum required receive SINR and the channel conditions on the transmit-power, some variations of the last setup were considered. For a minimum required receive-SINR = 3 dB per user in each group, the average transmit-power using all 12 antennas was 31.2 dBm, whereas the average transmit-power after selecting 6 antennas increased to 33.6 dBm. To simulate for better channel conditions, each user's channel was multiplied by a constant $c = 5$. As a result, the average transmit-power decreased to 17.2 dBm when all 12 antennas were utilized, and to 19.6 dBm when

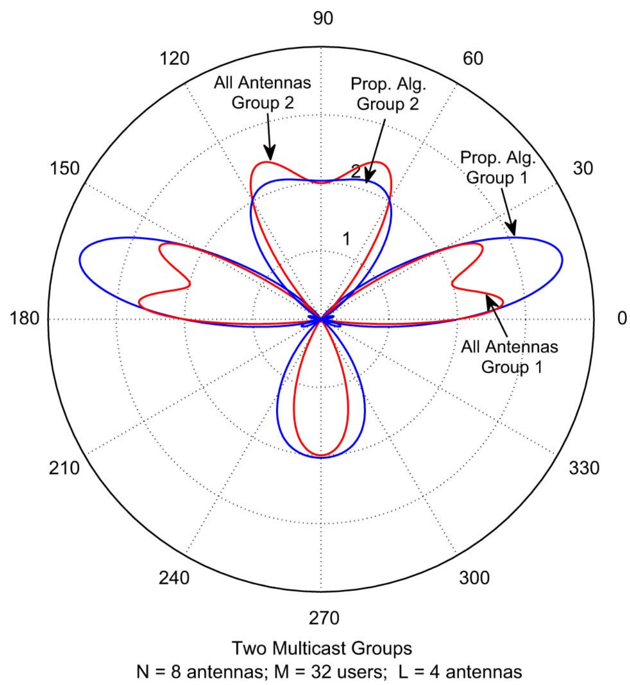

Fig. 5. Beampatterns for a far-field two-group multicasting scenario with $N = 8$ ULA and $M = 32$ users, comparing between using all $N = 8$ antennas and the proposed algorithm selecting $L = 4$ antennas.

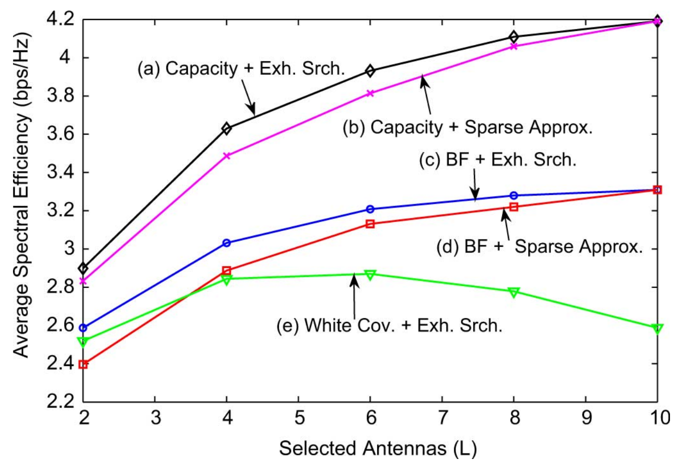

Fig. 6. Average spectral efficiency versus L for $N = 10$ antennas and $M = 16$ users.

6 antennas were selected. For a 12 dB minimum SINR, the average transmit-power was 41 dB when all antennas were used, and 44.4 dBm when 6 antennas were selected. When each user's channel was multiplied by $c = 5$, the average transmit-power decreased to 27.2 dBm when all 12 antennas were utilized, and to 30 dBm when 6 antennas were selected.

Far-field beamforming with $N = 8$ ULA. Fig. 5 illustrates the beampatterns for a particular far-field multicasting scenario with $N = 8$ ULA and $M = 32$ users. The users are equally divided into two multicast groups. The 16 users of the first multicast group G_1 have direction angles ϕ_m $(m \in G_1)$ from 0^o to 30^o with 2^o spacing between each user, while the 16 users of the second multicast group G_2 have direction angles ϕ_m $(m \in G_2)$ from 60^o to 90^o with 2^o spacing between each user. The minimum required receive-SINR was assumed to be 3 dB for each user, and we set $d/\lambda = 1/2$. Fig. 5 compares the beampattern resulting from using all $N = 8$ antennas with that resulting from using the proposed algorithm to select $L = 4$ antennas. For this setting, the proposed algorithm selects the same 4 antennas as the exhaustive search yielding the same beampattern. The proposed algorithm (and exhaustive search) incurs additional transmit-power of only 1.66 dB (compared to using all 8 antennas).

C. Max-Min-Fair Beamforming and Spectral Efficiency Considerations

Here, we consider the problem of maximizing the minimum received SNR over all users with antenna selection, which is described in Section V-B. In this setup, we considered a BS with $N = 10$ antennas, $M = 16$ users, Rayleigh fading channels and the transmission power was bounded below $P = 10$ dB. Fig. 6 compares the following schemes: (a) Capacity achieving transmit-covariance with exhaustive search antenna selection (which corresponds to the multicast channel capacity with antenna selection); (b) Capacity achieving transmit-covariance with sparsity-inducing ℓ_1-norm squared approximation; (c) Beamforming with exhaustive search antenna selection; (d) Beamforming with sparsity-inducing ℓ_1-norm squared approximation; and, (e) Spatially white transmit-covariance with exhaustive search antenna selection, as considered in [15]. Note that for (c) and (d), beamforming implies rank-one transmit-covariance. For each scheme, the (average) maximum achievable rate per unit bandwidth (which is the average spectral efficiency given by $\mathbb{E}_{\mathbf{h}}[\log_2(1+\tilde{\gamma}(\mathbf{h}))]$ in bps/Hz units, where $\tilde{\gamma}(\mathbf{h})$ is the minimum received SNR among all M users for each channel realization, and $\mathbb{E}_{\mathbf{h}}$ denotes Monte-Carlo expectation over all Rayleigh fading channel realizations) is plotted versus the number of selected antennas L.

Fig. 6 confirms that the previous conclusions for minimizing the transmission power with SNR constraints are also valid when antenna selection is jointly considered. For example, the average spectral efficiency with beamforming decreases by less than 0.5 bps/Hz when $L = 4$ antennas are selected compared to using all $N = 10$ antennas, which is an insignificant decrease compared to the reduction in RF chains. Moreover, the figure shows that (b) and (d) are within 0.25 bps/Hz less spectral efficiency than (a) and (c), respectively. On the other hand, (b) and (d) required solving less than 5 SDP problems, on average, while (a) and (c) required solving 210 SDP problems to select $L = 4$ or $L = 6$ antennas. This emphasizes the effectiveness of using the sparsity-inducing ℓ_1-norm squared approximation. Finally, we see that (c) outperforms (e), or in other words beamforming outperforms using spatially white transmit-covariance, since L is not very small compared to M. The performance of beamforming becomes significantly better as L increases, whereas this advantage vanishes for smaller L. The reason, as explained in [7], is that every beamforming direction will be nearly orthogonal to at least one user's channel with high probability when $M \gg L$.

VII. Conclusions

We studied the joint problem of multicast beamforming to multiple multicast groups with antenna selection. The objective is to select sparse beamforming vectors such that the transmission power is minimized, subject to the SINR constraints at all subscribers. Instead of using the ℓ_1-norm to promote sparsity, we argued that the mixed $\ell_{1,\infty}$-norm squared offers a more prudent group-sparsity inducing regularization for our purposes. The reason is that it naturally (and elegantly) yields a semidefinite relaxation that is similar in spirit to the corresponding one for the baseline multicasting problem without antenna selection, considered in [18]. One interesting result is that the number of transmit antennas can be considerably reduced with only minimal increase in the transmission power. We also showed that our proposed algorithm performs joint antenna selection and weight optimization at significantly lower complexity compared to using exhaustive search for antenna selection, and at negligible excess power. The novel algorithm can be combined with admission control [12], and can be easily modified to obtain sparse solutions for a wide class of non-convex QCQP problems, and applications where SDP relaxation is relevant.

Finally, developing custom algorithms for joint multicast beamforming and antenna selection for *Massive MIMO* systems [26] is of interest. We have preliminary work in this direction; but striking the right balance between performance and complexity in the large system regime requires a very different approach from the one presented herein. We will therefore report these findings in follow-up work.

Appendix

Proof of Proposition 1: Consider the single multicast group scenario. In this case, the $\ell_{1,\infty}$-norm reduces to the ℓ_1-norm and problem (8) is expressed as

$$\begin{aligned}\min_{\mathbf{w}\in\mathbb{C}^N} \quad & \|\mathbf{w}\|_2^2+\lambda\|\mathbf{w}\|_1^2\\ \text{s.t.:} \quad & |\mathbf{w}^{\mathcal{H}}\mathbf{h}_m|^2\geq 1, \quad m=1,\ldots,M \end{aligned}\tag{22}$$

where $\mathbf{h}_m$ is normalized by $\sqrt{\sigma_m^2\gamma_m}$. Using $\mathbf{Q}_m := \mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}$ and $\mathbf{X} := \mathbf{w}\mathbf{w}^{\mathcal{H}}$, problem (22) is equivalent to

$$\begin{aligned}\min_{\mathbf{X}\in\mathbb{C}^{N\times N},\mathbf{Y}\in\mathbb{R}^{N\times N}} \quad & \mathrm{tr}(\mathbf{X})+\lambda\mathrm{tr}(\mathbf{1}_N\mathbf{Y})\\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{X}\mathbf{Q}_m)\geq 1,\ m=1,\ldots,M\\ & |X(i,j)|\leq Y(i,j),\ \forall i,j\in\{1,\ldots,N\}\\ & \mathbf{X}\succeq 0, \quad \mathrm{rank}(\mathbf{X})=1. \end{aligned}\tag{23}$$

In order to transform (23) from the complex domain to the real domain, we define $\mathbf{w}_R := \Re\{\mathbf{w}\}$, $\mathbf{w}_I := \Im\{\mathbf{w}\}$, $\mathbf{z} := [\mathbf{w}_R^T\ \mathbf{w}_I^T]^T$, and $\mathbf{Z} := \mathbf{z}\mathbf{z}^T$ such that $\mathbf{Z}\in\mathbb{R}^{2N\times 2N}$. Now, it is easy to see that $\Re\{X(i,j)\} = Z(i,j)+Z(N+i,N+j)$ and $\Im\{X(i,j)\} = -Z(i,N+j)+Z(N+i,N)$. Thus, the constraints $|X(i,j)|\leq Y(i,j), \forall i,j$ are equivalent to

$$\begin{aligned}\|\,&(Z(i,j)+Z(N+i,N+j))\\ &+\sqrt{-1}\,(-Z(i,N+j)+Z(N+i,N))\,\|_2\leq Y(i,j).\end{aligned}$$

These constraints can be expressed as the positive semidefinite constraints (24) at the bottom of the page, $\forall i,j$, in the real domain.

The channel matrix $\mathbf{Q}_m$ can be transformed to the real domain by defining $\mathbf{g}_m := [\Re\{\mathbf{h}_m\}^T\ \ \Im\{\mathbf{h}_m\}^T]^T$, $\bar{\mathbf{g}}_m := [\Im\{\mathbf{h}_m\}^T\ -\Re\{\mathbf{h}_m\}^T]^T$, and the $2N\times 2N$ rank-2 matrix $\bar{\mathbf{Q}}_m := \mathbf{g}_m\mathbf{g}_m^T+\bar{\mathbf{g}}_m\bar{\mathbf{g}}_m^T$. Hence, problem (23) can be expressed in the real domain as

$$\begin{aligned}\min_{\mathbf{Z}\in\mathbb{R}^{2N\times 2N},\mathbf{Y}\in\mathbb{R}^{N\times N}} \quad & \mathrm{tr}(\mathbf{Z})+\lambda\mathrm{tr}(\mathbf{1}_N\mathbf{Y})\\ \text{s.t.:} \quad & \mathrm{tr}(\mathbf{Z}\bar{\mathbf{Q}}_m)\geq 1,\ m=1,\ldots,M\\ & \mathbf{C}_{ij}\succeq 0,\ \forall i,j\in\{1,\ldots,N\}\\ & \mathbf{Z}\succeq 0, \quad \mathrm{rank}(\mathbf{Z})=1. \end{aligned}\tag{25}$$

Finally, defining $\tilde{\mathbf{Z}} := \mathrm{diag}(\mathbf{Z}\ \ \mathbf{Y}\ \ \mathbf{C}_{11}\ \ \mathbf{C}_{12}\ \ \ldots\ \ \mathbf{C}_{1N}\ \ \mathbf{C}_{21}\ \ \ldots\ \ \mathbf{C}_{NN})$ such that $\tilde{\mathbf{Z}}\in\mathbb{R}^{\tilde{N}\times\tilde{N}}$, where $\tilde{N} := 3N+2N^2$,

$$\mathbf{A} := \begin{bmatrix}\mathbf{I}_{2N} & \mathbf{0}_{2N\times N} & \mathbf{0}_{2N\times 2N^2}\\ \mathbf{0}_{N\times 2N} & \lambda.\mathbf{1}_N & \mathbf{0}_{N\times 2N^2}\\ \mathbf{0}_{2N^2\times 2N} & \mathbf{0}_{2N^2\times N} & \mathbf{0}_{2N^2\times 2N^2}\end{bmatrix}$$

and

$$\tilde{\mathbf{Q}}_m := \begin{bmatrix}-\bar{\mathbf{Q}}_m & \mathbf{0}_{2N\times(N+2N^2)}\\ \mathbf{0}_{(N+2N^2)\times 2N} & \mathbf{0}_{(N+2N^2)\times(N+2N^2)}\end{bmatrix}$$

problem (25) can be expressed as

$$\begin{aligned}\min_{\tilde{\mathbf{Z}}\in\mathbb{R}^{\tilde{N}\times\tilde{N}}} \quad & \mathrm{tr}(\tilde{\mathbf{Z}}\mathbf{A})\\ \text{s.t.:} \quad & \mathrm{tr}(\tilde{\mathbf{Z}}\tilde{\mathbf{Q}}_m)+1\leq 0,\ m=1,\ldots,M\\ & \mathrm{tr}\left(\tilde{\mathbf{Z}}\mathbf{E}_{ij}^{(11)}\right)=0,\ \mathrm{tr}\left(\tilde{\mathbf{Z}}\mathbf{E}_{ij}^{(12)}\right)=0\\ & \mathrm{tr}\left(\tilde{\mathbf{Z}}\mathbf{E}_{ij}^{(21)}\right)=0,\ \mathrm{tr}\left(\tilde{\mathbf{Z}}\mathbf{E}_{ij}^{(22)}\right)=0,\\ & \forall i,j\in\{1,\ldots,N\}, \quad \tilde{\mathbf{Z}}\succeq 0, \quad \mathrm{rank}(\tilde{\mathbf{Z}})=1 \end{aligned}\tag{26}$$

where $\mathbf{E}_{ij}^{(11)}$, $\mathbf{E}_{ij}^{(22)}$, $\mathbf{E}_{ij}^{(12)}$ and $\mathbf{E}_{ij}^{(21)}$ provide the four entries of $\mathbf{C}_{ij}$, for each i and j, by ensuring that:

$$\begin{aligned}&\tilde{\mathbf{Z}}\,(3N+2(i-1)N+2j-1,3N+2(i-1)N+2j-1)\\ &\quad=\tilde{\mathbf{Z}}(2N+i,2N+j)-\tilde{\mathbf{Z}}(i,j)-\tilde{\mathbf{Z}}(N+i,N+j)\\ &\tilde{\mathbf{Z}}\,(3N+2(i-1)N+2j,3N+2(i-1)N+2j)\\ &\quad=\tilde{\mathbf{Z}}(2N+i,2N+j)+\tilde{\mathbf{Z}}(i,j)+\tilde{\mathbf{Z}}(N+i,N+j)\\ &\tilde{\mathbf{Z}}\,(3N+2(i-1)N+2j-1,3N+2(i-1)N+2j)\\ &\quad=-\tilde{\mathbf{Z}}(i,N+j)+\tilde{\mathbf{Z}}(N+i,j)\\ &\tilde{\mathbf{Z}}\,(3N+2(i-1)N+2j,3N+2(i-1)N+2j-1)\\ &\quad=-\tilde{\mathbf{Z}}(i,N+j)+\tilde{\mathbf{Z}}(N+i,j).\end{aligned}$$

$$\mathbf{C}_{ij} := \begin{bmatrix}Y(i,j)-Z(i,j)-Z(N+i,N+j) & -Z(i,N+j)+Z(N+i,j)\\ -Z(i,N+j)+Z(N+i,j) & Y(i,j)+Z(i,j)+Z(N+i,N+j)\end{bmatrix}\succeq 0.\tag{24}$$

By dropping the $\mathrm{rank}(\tilde{\mathbf{Z}}) = 1$ constraint, problem (26) is in the standard SDP form. Defining the $\tilde{N} \times 1$ vector $\mathbf{z} := [\mathbf{w}_R^T \ \mathbf{w}_I^T \ \mathbf{y}^T \ \mathbf{c}_{11}^T \ \mathbf{c}_{12}^T \ \ldots \ \mathbf{c}_{1N}^T \ \mathbf{c}_{21}^T \ \ldots \ \mathbf{c}_{NN}^T]^T$ where $\mathbf{y}$ is an $N \times 1$ auxiliary vector and $\mathbf{c}_{ij}$ is a 2×1 vector, problem (26), which is equivalent to the original problem (22), can be expressed in the following standard QCQP form:

$$\begin{aligned} &\min_{\mathbf{z}} \mathbf{z}^T \mathbf{A} \mathbf{z} \\ &\text{s.t.: } \mathbf{z}^T \tilde{\mathbf{Q}}_m \mathbf{z} + 1 \leq 0, \ m = 1, \ldots, M \\ &\qquad \mathbf{z}^T \mathbf{E}_{ij}^{(11)} \mathbf{z} = 0, \ \mathbf{z}^T \mathbf{E}_{ij}^{(12)} \mathbf{z} = 0 \\ &\qquad \mathbf{z}^T \mathbf{E}_{ij}^{(21)} \mathbf{z} = 0, \ \mathbf{z}^T \mathbf{E}_{ij}^{(22)} \mathbf{z} = 0, \ \forall i, j \in \{1, \ldots, N\}. \end{aligned} \tag{27}$$

Introducing the Lagrange multipliers $\boldsymbol{\mu}$ $(M \times 1)$, $\boldsymbol{\nu}$ $(4N^2 \times 1)$, and defining $l := 4(i-1)N + j$ and

$$\begin{aligned} \mathbf{P} := \mathbf{A} + \sum_{m=1}^{M} \mu_m \tilde{\mathbf{Q}}_m + \sum_{i=1}^{N} \sum_{j=1}^{N} \Big(&\nu_l \mathbf{E}_{ij}^{(11)} \\ &+ \nu_{l+N} \mathbf{E}_{ij}^{(12)} + \nu_{l+2N} \mathbf{E}_{ij}^{(21)} + \nu_{l+3N} \mathbf{E}_{ij}^{(22)} \Big), \end{aligned}$$

the Lagrangian of problem (27) is $\mathcal{L}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \mathbf{z}^T \mathbf{P} \mathbf{z} + \sum_{m=1}^{M} \mu_m$,and the dual problem is

$$\max_{\boldsymbol{\mu} \succeq 0, \boldsymbol{\nu}} \inf_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\nu}).$$

It is easy to see that

$$\inf_{\mathbf{z}} \mathbf{z}^T \mathbf{P} \mathbf{z} + \sum_{m=1}^{M} \mu_m = \begin{cases} \sum_{m=1}^{M} \mu_m & \text{if } \mathbf{P} \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem can thus be expressed as

$$\begin{aligned} &\max_{\boldsymbol{\mu}, \boldsymbol{\nu}} \quad \sum_{m=1}^{M} \mu_m \\ &\text{s.t.:} \quad \mathbf{P} \succeq 0, \quad \mu_m \geq 0, \quad m = 1, \ldots, M \end{aligned} \tag{28}$$

which is an easily solvable convex SDP. Finally, it is easy to see that the dual of the SDP (28), which is the bi-dual of (22), is problem (23) after dropping the $\mathrm{rank}(\mathbf{X}) = 1$ constraint [22]. Also, the dual of the rank-relaxed problem (23) is problem (28).

The duality results are easily extended to the multiple multicast groups scenario by extending the matrix $\mathbf{Z}$ to $\{\mathbf{Z}_k\}_{k=1}^K$ and adding K replicates for the positive semidefinite constraints $\mathbf{C}_{ij} \succeq 0, \forall i, j \in \{1, \ldots, N\}$, in (25) corresponding to each $\mathbf{Z}_k$, $\forall k \in \{1, \ldots, K\}$. The rest of the steps are a straightforward extension from the single multicast group case.

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," Tech. Rep., 2011 [Online]. Available: http://arxiv.org/pdf/1108.0775v2.pdf

[2] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Commun.*, L.C. Godara, Ed. Boca Raton, FL: CRC Press, 2001, ch. 18.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[5] J. de Andrade, Jr, M. Campos, and J. Apolinario, "Sparse solutions for antenna arrays," in *Proc. XXIX Simposio Brasileiro de Telecomunicacoes*, Curitiba, Brazil, Oct. 2–5, 2011.

[6] A. Dua, K. Medepalli, and A. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, Sep. 2006.

[7] N. Jindal and Z. Q. Luo, "Capacity limits of multiple antenna multicast," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, USA, Jul. 9–14, 2006, pp. 1841–1845.

[8] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min-fair transmit beamforming to multiple co-channel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[9] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4916–4927, Oct. 2007.

[10] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in MATLAB," presented at the Proc. CACSD, Taipei, Taiwan, Sep. 4, 2004.

[11] Z.-Q. Luo, W. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[12] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Efficient batch and adaptive approximation algorithms for joint multicast beamforming and admission control," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4882–4894, Dec. 2009.

[13] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Multicast beamforming with antenna selection," presented at the 13th Workshop Signal Process. Adv. Wireless Commun., Cesme, Turkey, Jun. 17–20, 2012.

[14] S. Nai, W. Ser, Z. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *IEEE Trans. Antennas Propag.*, vol. 58, no. 12, pp. 3923–3930, Dec. 2010.

[15] S. Park and D. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, Jun. 2008.

[16] K. Phan, S. Vorobyov, N. D. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QOS-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2323–2335, Jun. 2009.

[17] S. Sanayei and A. Nostratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, pp. 68–73, Oct. 2004.

[18] N. D. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pt. 1, pp. 2239–2251, Jun. 2006.

[19] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1, pp. 625–653, 1999.

[20] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.

[21] B. Turlach, W. N. Venables, and S. J. Wright, "Simultaneous variable selection," *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.

[22] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, pp. 49–95, 1996.

[23] H. Wolkowicz, "Relaxations of Q2p," in *Chapter 13.4 in Handbook of Semidefinite Programming: Theory, Algorithms, and Applicat*, H. Wolkowicz, R. Saigal, and L. Venberghe, Eds. Norwell, MA, USA: Kluwer, 2000.

[24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. B.*, vol. 68, no. 1, pp. 49–67, 2006.

[25] M. Yukawa and I. Yamada, "Minimal antenna-subset selection under capacity constraint for power-efficient MIMO systems: A relaxed ℓ_1 minimization approach," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 14–19, 2010, pp. 3058–3061.

[26] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

Omar Mehanna (S'05) received the B.Sc. degree in electrical engineering from Alexandria University, Egypt, in 2006 and the M.Sc. degree in electrical engineering from Nile University, Egypt, in 2009.

Since 2009, he has been working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. His current research focuses on signal processing for communications, ad-hoc networks, and cognitive radio.

Nicholas D. Sidiropoulos (F'09) received the Diploma degree in electrical engineering from the Aristotelian University of Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland—College Park, in 1988, 1990, and 1992, respectively.

He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota, Minneapolis (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the University of Minnesota—Minneapolis (2011-present). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement.

Dr. Sidiropoulos received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, and 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009), and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008).

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, 1981. From 1982 to 1986, he was with the University of Southern California (USC), where he received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

Since 1999 he has been a Professor with the University of Minnesota, Minneapolis, MN, where he now holds an ADC Chair in Wireless Telecommunications in the ECE Department and serves as director of the Digital Technology Center. His general interests span the areas of communications, networking and statistical signal processing—subjects on which he has published more than 340 journal papers, 570 conference papers, 20 book chapters, two edited books, and two research monographs (h-index 100). Current research focuses on sparsity in signals and systems, wireless cognitive radios, mobile ad hoc networks, wireless sensor, renewable energy, power grid, gene-regulatory, and social networks.

Prof. Giannakis is the (co-) inventor of 21 patents issued, and the (co-) recipient of eight best paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.