Optimal MIMO Precoding Under a Constraint on the Amplifier Power Consumption

Victor Cheng, Daniel Persson and Erik G Larsson

Tweet

LiU LINKÖPING UNIVERSITY

Optimal MIMO Precoding Under a Constraint on the Amplifier Power Consumption

Hei Victor Cheng *, Daniel Persson †, and Erik G. Larsson *
* Department of Electrical Engineering (ISY), Linköping University, Sweden
Email: {hei.cheng, erik.g.larsson}@liu.se
† Qamcom R&T, Gothenburg, Sweden
Email: persson15@yahoo.com

Abstract—The capacity of the MIMO channel taking into account both a limitation on total consumed power, and per-antenna radiated power constraints is considered. The total consumed power takes into account the traditionally used sum radiated power, and also the power dissipation in the amplifiers. For a fixed channel with full CSI at both the transmitter and the receiver, maximization of the mutual information is formulated as an optimization problem. Lower and upper bounds on the capacity are provided by numerical algorithms based on partitioning of the feasible region. Both bounds are shown to converge and give the exact capacity when number of regions increases. The bounds are also used to construct a monotonic optimization algorithm based on the branch-and-bound approach. An efficient suboptimal algorithm based on successive convex approximation performing close to the capacity is also presented. Numerical results show that the performance of the solution obtained from the suboptimal algorithm is close to that of the global optimal solution. Simulation results also show that in the low SNR regime, antenna selection provides performance that is close to the optimal scheme while at high SNR, uniform power allocation performs close to the optimal scheme.

Keywords—MIMO capacity, power amplifier, consumed power constraint, per-antenna power constraint

I. INTRODUCTION

Recently the use of smartphones and tablets has led to a tremendous increase in the demand for high data rates over the wireless networks. Meanwhile, wireless transmissions are required to be robust to fading and shadowing effects. The idea of using multiple antennas at both the transmitters and the receivers, i.e. multiple-input-multiple-output (MIMO) technology, can be a solution. Compared to single-antenna systems, MIMO offers higher data rates thanks to the multiplexing gain, as well as robustness against fading owing to a diversity gain, for the same amount of time and frequency resources. When full knowledge of the channel is available at the transmitter, the transmission can be optimized according to certain different criteria. Telatar showed in [2], that with circularly symmetric complex Gaussian noise and full channel state information (CSI) at the transmitter, the capacity of multiple-input-multiple-output (MIMO) channels under a sum

radiated power constraint over all antennas is achieved by eigen-beamforming together with water-filling power allocation over different eigen-modes.

The sum power constraint considered in [2] does not take into account the maximum output constraint of the individual power amplifiers (PAs). A more realistic model is obtained by including per-antenna power constraints, which were first considered in [3]. An algorithm was derived in [4] to obtain the optimal input distribution for a given channel under per-antenna power constraints. Later these results were extended to multi-hop communications systems in [5] using matrix monotone functions. In [6] the authors provide a general framework for covariance matrix optimization with different types of radiated power constraints. However in practice the power consumed by the PAs includes both the actual radiated power and the power dissipated as heat in the hardware. The capacity optimization under per-antenna power constraints does not tell us how the power should be optimally allocated to the antennas given a specific channel, total power consumption in each PA, and a total power budget for all antennas. Therefore a carefully modeled consumed power, which takes into account both the sum radiated power and power dissipation is needed.

According to a report, PAs are estimated to consume 57% of the energy in macro base stations [7, Figure 12]. Cooling contributes another 10%, which can also be reduced if power dissipation from the PA is reduced. Hence, it is essential to take the PA into account when designing the transmitters in the base stations. The importance of carefully taking the hardware into consideration when designing communication systems has recently been put forward by others, e.g. in [8] and [9]. Despite the importance, few efforts are reported on MIMO capacity optimization with a total consumed power budget. Sub-optimal algorithms with a diagonal input co-variance matrix were proposed in [10]. In [11] the authors derived the multiple-input single-input (MISO) capacity with full CSI and provided results on ergodic MISO capacity with per-antenna and total consumed power constraints. In [12] a special case of the MIMO channels with per-antenna and total consumed power constraints was solved. In [13] the authors applied the same consumed power model for PAs to model the circuit power consumption in multi-user MIMO scenarios for energy efficiency aximization. Similar energy efficiency optimization problem are studied in [14] for two-way relay channels. In the conference version of this work

[1], a preliminary study of MIMO capacity with per-antenna and consumed power constraints was done. References [10] and [12] provided preliminary studies on this topic, with focus on the special case when the MIMO channels have a diagonal structure. Until now the capacity of the general MIMO channel with per-antenna and total consumed power constraints is unknown and this work will find tight upper and lower bounds, and approximations, to this capacity. Specifically, we provide a complete analysis of the MIMO case where we exploit global optimization techniques. Extensions to multi-carrier and multi-user scenarios are also presented.

A. Technical contributions of this work

- We derive the capacity of flat fading point-to-point MIMO channels with full CSI, under per-antenna and total consumed power constraints. Finding the capacity involves solving a non-convex optimization problem, thus lower and upper bounds are derived.
- We propose a successive convex optimization approach and a monotonic optimization algorithm to find the capacity of MIMO channels under per-antenna and total consumed power constraints up to arbitrary accuracy.
- We extend the analysis to frequency-selective fading MIMO channels, and to multi-user MIMO beamforming problems, where the goal is to minimize the total consumed power under SINR constraints.

The paper is organized as follows: In Section II we describe the system model and establish capacity results under various power constraints, and formulate the pertinent optimization problems. In Section III we derive upper and lower bounds on capacity, and provide an efficient algorithm to find a local optimal point of the optimization problem which serves as a tight lower bound. In Section IV we utilize the derived upper and lower bounds to develop a global optimization algorithm based on monotonic optimization theory and compare it to the low-complexity algorithm. In Section V we extend the analysis to multi-carrier MIMO systems, and show that the same analysis and methods can be applied with slight modifications. In Section VI we extend our investigation to multi-user MIMO systems, and show that the power minimization with SINR constraints can be solved in special cases, corresponding to the use of class B and class D PAs in the base stations (BS). In Section VII we present numerical results and discuss a connection between our work and the field of compressed sensing.

In the following, scalars are denoted by lower-case letters, vectors are denoted by bold-face lower-case letters, and matrices are denoted by bold-face upper-case letters. A vector $\boldsymbol{x}$ of length N has components x_i, where $i = 1, ..., N$, and a matrix $\boldsymbol{X}$ of size $M \times N$ has elements $X_{m,n}$, with $m = 1, ..., M$ and $n = 1, ..., N$. The determinant of a matrix is denoted as $|\boldsymbol{X}|$. The symbol $\mathrm{diag}(\boldsymbol{X})$ represents a vector formed by the elements on the main diagonal of $\boldsymbol{X}$. Furthermore, $(\cdot)^{\mathrm{T}}$ is the transpose, $(\cdot)^*$ is the complex conjugate, $(\cdot)^H$ is the conjugate transpose, $\Re(\cdot)$ denotes the real part, and $\Im(\cdot)$ denotes the imaginary part. The relation $\boldsymbol{X} \succeq \boldsymbol{0}$ means that the matrix $\boldsymbol{X}$ is positive semi-definite.

II. CAPACITY OF POINT TO POINT MIMO CHANNELS

For the MIMO transmission, we assume a flat fading baseband channel model, given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{C}^{N_R \times 1}$ is the received signal, $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$ is the channel, and $\boldsymbol{x} \in \mathbb{C}^{N_T \times 1}$ is the transmitted symbol vector, where N_T is the number of transmit antennas. The noise vector $\boldsymbol{n} \in \mathbb{C}^{N_R \times 1}$ has independent identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian elements, i.e. $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I})$. In this paper we consider the case with full CSI at the transmitter and at the receiver for a fixed channel $\boldsymbol{H}$. Since the noise is Gaussian, it holds that for any average power constraint, the signal input giving the largest mutual information (MI) is zero-mean circularly symmetric complex Gaussian.

The signal $\boldsymbol{x}$ in (1) is obtained by precoding of independent information streams destined to the receiver, using a linear precoder $\boldsymbol{P}$. Let $\boldsymbol{Q} = E[\boldsymbol{x}\boldsymbol{x}^H] = \boldsymbol{P}\boldsymbol{P}^H$ be the covariance matrix of the precoded signal (this matrix is positive semi-definite by definition [2]). The MI is then

$$R(\boldsymbol{Q}) = \log_2 \left| \boldsymbol{I} + \frac{1}{\sigma^2} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H \right| \tag{2}$$

bit/s/Hz, cf. [2]. To find the capacity, the input distribution needs to be optimized subject to all power constraints. That is, we want to find

$$\sup_{\boldsymbol{Q} \in \mathcal{Q}} R(\boldsymbol{Q}), \tag{3}$$

where $\mathcal{Q}$ is the set of power constraints of interest. The radiated power constraint for antenna i is

$$E[|x_i|^2] = Q_{i,i} \leq P_{\max}, \tag{4}$$

where $P_{\max}$ is the maximum possible power radiated by antenna i. For the consumed power, following the discussions about amplifier modeling in [12], we set

$$\frac{Q_{i,i}}{P_{\mathrm{cons},i}} = \eta_{\max} \left(\frac{Q_{i,i}}{P_{\max}} \right)^{1-\epsilon}, \tag{5}$$

where ϵ is a parameter with $\epsilon \in [0, 0.5]$, $P_{\mathrm{cons},i}$ is the consumed power on antenna i, and $\eta_{\max} \in [0, 1]$ is the maximum possible power efficiency. This maximum efficiency is obtained when $Q_{i,i} = P_{\max}$. The maximum power efficiency is a fixed parameter, common for all the PAs at the transmitter. From the above, we can write

$$P_{\mathrm{cons},i} = \frac{1}{\eta_{\max}} Q_{i,i}^\epsilon P_{\max}^{1-\epsilon}. \tag{6}$$

The consumed power $P_{\mathrm{cons},i}$ is thus proportional to the ϵ-th power of the radiated power and it will always be greater than the radiated power.

We are now able to state the rate maximization problem

corresponding to (3):

$$\underset{\boldsymbol{Q}}{\text{maximize}} \quad R(\boldsymbol{Q})$$

$$\text{subject to} \quad \frac{P_{\max}^{1-\epsilon}}{\eta_{\max}} \sum_{i=1}^{N_{\mathrm{T}}} Q_{i,i}^{\epsilon} \leq \tilde{P}_{\mathrm{tot}} \tag{7}$$

$$0 \leq Q_{i,i} \leq P_{\max}, \ \ i = 1, ..., N_{\mathrm{T}}$$

$$\boldsymbol{Q} \succeq \boldsymbol{0},$$

where $\tilde{P}_{\mathrm{tot}}$ is the total consumed power limitation (in sum over all transmit antennas). This means that we aim to find the capacity, subject to both the consumed power constraint and the radiated power constraints. Writing

$$P_{\mathrm{tot}} = \frac{\tilde{P}_{\mathrm{tot}} \eta_{\max}}{P_{\max}^{1-\epsilon}}, \tag{8}$$

we can express the optimization problem as follows:

$$\underset{\boldsymbol{Q}}{\text{maximize}} \quad R(\boldsymbol{Q})$$

$$\text{subject to} \quad \sum_{i=1}^{N_{\mathrm{T}}} Q_{i,i}^{\epsilon} \leq P_{\mathrm{tot}} \tag{9}$$

$$0 \leq Q_{i,i} \leq P_{\max}, \ \ i = 1, ..., N_{\mathrm{T}}$$

$$\boldsymbol{Q} \succeq \boldsymbol{0}.$$

Problem (9) is non-convex due to the non-convex constraint on the consumed power. This makes the problem different from the traditional channel capacity maximization problem with a total radiated power constraint. In that case the radiated power is invariant to a multiplication by a unitary matrix on the left and on the right, and therefore the problem can be reduced to a simple water-filling problem. However the consumed power constraint in (9) does not have this property and therefore optimization of the whole input covariance matrix is needed.

For some values of the system parameters N_{T}, $P_{\max}$ and P_{tot}, the consumed power constraint can become inactive. Specifically, the consumed power constraint is inactive when

$$N_{\mathrm{T}} P_{\max}^{\epsilon} \leq P_{\mathrm{tot}}, \tag{10}$$

where the maximum consumed power is upper bounded by the per-antenna radiated power constraints. In this case, the optimization problem (9) reduces to the problem with only per-antenna constraints and can be solved efficiently as it is a convex problem.

The problem of interest in practice would be to consider a given requested user throughput, and then find the precoder that can deliver this throughput with the smallest possible consumed (that is, radiated plus dissipated as heat) power. This is desirable, for example for an operator, to minimize operational expenses. In the optimization above, we consider the equivalent problem of maximizing the throughput subject to a constraint on the consumed power.

Importantly, the efficiency of the amplifiers degrades when they are not operated at their maximum power. However, this does not mean that the optimal precoder for a pre-determined throughput target would operate the amplifiers at maximum permitted power. Moreover, we consider a predistorted output

signal here which is distortion-free. The power consumption model that we use is valid also for a signal that has undergone digital pre-distortion [15].

III. Upper Bound and Lower Bound on Capacity

Finding the globally optimal solution of the non-convex problem (9) is hard. However we can obtain bounds on the capacity as follows. Consider the constraint

$$f(\boldsymbol{Q}) \triangleq \sum_{i=1}^{N_{\mathrm{T}}} Q_{i,i}^{\epsilon} \leq P_{\mathrm{tot}}. \tag{11}$$

By providing bounds on the function $f(\boldsymbol{Q})$ in different quantized regions, we can obtain a lower bound and an upper bound on capacity. The bounds should become tighter as we increase the number of quantization steps. Here we introduce specific approaches that produce lower bounds and upper bounds.

A. Lower Bound on Capacity

By upper bounding each $Q_{i,i}^{\epsilon}$ with a linear function we obtain a lower bound on the capacity. Denote the sum of the upper bounds on each $Q_{i,i}^{\epsilon}$ by $f^{U}(\boldsymbol{Q})$. Then we are shrinking the feasible set since $f^{U}(\boldsymbol{Q}) \leq P_{\mathrm{tot}}$ implies $f(\boldsymbol{Q}) \leq P_{\mathrm{tot}}$ but not vice versa. Maximizing the same objective function with a smaller feasible set will give us a lower bound on the optimal value, thus this tightened problem provides a lower bound on the capacity.

As $Q_{i,i}^{\epsilon}$ is a concave function, the first-order Taylor expansion will give an upper bound on $Q_{i,i}^{\epsilon}$. By dividing the region $[0, P_{\max}]$ into K different parts

$$[0, P_{\max}/K], \ldots, [P_{\max}(K-1)/K, P_{\max}],$$

we can get an upper bound on $Q_{i,i}^{\epsilon}$ in each interval by picking any point in the sub-region and apply a first-order Taylor approximation around that point. Herein we choose the point as the mid-point of each sub-region, i.e. $P_{\max}/2K, 3P_{\max}/2K, \ldots, (2K-1)P_{\max}/2K$. Then in each sub-region, the optimal $\boldsymbol{Q}$ can be obtained by solving the following set of optimization problems:

$$\underset{\boldsymbol{Q}}{\text{maximize}} \quad R(\boldsymbol{Q})$$

$$\text{subject to} \quad \sum_{i=1}^{N_{\mathrm{T}}} a_i^j Q_{i,i} + b_i^j \leq P_{\mathrm{tot}}$$

$$\frac{c_i^j - 1}{K} P_{\max} \leq Q_{i,i} \leq \frac{c_i^j}{K} P_{\max}, \ \ i = 1, ..., N_{\mathrm{T}}$$

$$\boldsymbol{Q} \succeq \boldsymbol{0}, \tag{12}$$

where $c_i^j = 1, \ldots, K$ indicates the sub-region that the power radiated from each antenna i is in, and

$$a_i^j \triangleq \epsilon \left(\frac{2c_i^j - 1}{2K} P_{\max} \right)^{\epsilon-1}, \qquad b_i^j \triangleq (1-\epsilon) \left(\frac{2c_i^j - 1}{2K} P_{\max} \right)^{\epsilon} \tag{13}$$

specifies the first-order Taylor approximation. To get the lower bound, we choose the $\boldsymbol{Q}$ corresponding to the maximum rate

obtained by solving the K^{N_T} problems associated with the different regions.

B. Upper Bound on Capacity

By lower bounding each $Q_{i,i}^\epsilon$ by a linear function we obtain an upper bound on the capacity. Specifically, denoting the sum of the lower bounds on each $Q_{i,i}^\epsilon$ by $f^L(\boldsymbol{Q})$, we are enlarging the feasible set since $f(\boldsymbol{Q}) \leq P_{\text{tot}}$ implies $f^L(\boldsymbol{Q}) \leq P_{\text{tot}}$ but not vice versa. The optimal value cannot decrease when performing this relaxation, yielding an upper bound on the capacity.

Again, from the fact that $Q_{i,i}^\epsilon$ is concave, upper bounds can be constructed using the quantized regions

$$[0, P_{\max}/K], \ldots, [P_{\max}(K-1)/K, P_{\max}].$$

Specifically we join all the neighboring vertices of each boundary point, $0, P_{\max}/K, \ldots, P_{\max}(K-1)/K, P_{\max}$, by straight lines. Hence, the upper bound on capacity can be obtained by solving (12) with the variables in the first constraint defined as

$$
\begin{aligned}
a_i^j &\triangleq \left[\left(\frac{c_i^j}{K} P_{\max} \right)^\epsilon - \left(\frac{c_i^j - 1}{K} P_{\max} \right)^\epsilon \right] / (P_{\max}/K), \\
b_i^j &\triangleq c_i^j \left(\frac{c_i^j - 1}{K} P_{\max} \right)^\epsilon - (c_i^j - 1) \left(\frac{c_i^j}{K} P_{\max} \right)^\epsilon.
\end{aligned}
\tag{14}
$$

To get the upper bound, we choose the $\boldsymbol{Q}$ corresponding to the maximum rate obtained by solving the K^{N_T} problems associated with the different regions.

The upper and lower bounds will approach each other if the number K of quantization regions of $Q_{i,i}$ increases.

C. Efficient Algorithm for Computation of a Tight Lower Bound

The computational complexity of obtaining the upper and lower bounds given above is polynomial in K and exponential in N_T. If K and N_T are large, evaluation of the bounds may be infeasible. Here we present a practical algorithm for the computation of a lower bound, which can perform well and has acceptable complexity when N_T is large. When the suboptimal $\boldsymbol{Q}$ is obtained, a linear pre-coding scheme can be applied by using i.i.d. input with the pre-coder $\boldsymbol{P}$ satisfying $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{P}^H$. Note that any feasible solution to problem (9) is a lower bound on the capacity. The resulting capacity bound has the interpretation of an achievable rate with this signaling.

Our lower-bounding strategy is based on successive convex approximations. The successive convex approximation is an algorithm framework [16] traditionally used in power control applications to handle non-convex constraints [17, Sec IV. A]. The idea is to solve a series of approximated problems where the non-convex constraint is approximated with a convex function in each iteration. For our problem (9), we approximate the non-convex function $f(\boldsymbol{Q})$ with a convex function $f_k(\boldsymbol{Q})$ in the k-th iteration.

The convex optimization problem to be solved in the k-th iteration is

$$
\begin{aligned}
\underset{\boldsymbol{Q}}{\text{maximize}} \quad & R(\boldsymbol{Q}) \\
\text{subject to} \quad & f_k(\boldsymbol{Q}) \leq P_{\text{tot}} \\
& 0 \leq Q_{i,i} \leq P_{\max}, \quad i = 1, ..., N_{\text{T}} \\
& \boldsymbol{Q} \succeq \boldsymbol{0}.
\end{aligned}
\tag{15}
$$

If we construct a family of functions $f_k(\mathbf{Q})$ in each iteration k satisfying the conditions

1) $f(\mathbf{Q}) \leq f_k(\mathbf{Q})$, $\forall \mathbf{Q}$ in the feasible set,
2) $f(\mathbf{Q}^{(k-1)}) = f_k(\mathbf{Q}^{(k-1)})$, where $\mathbf{Q}^{(k-1)}$ is the solution from the previous iteration,
3) $\nabla f(\mathbf{Q}^{(k-1)}) = \nabla f_k(\mathbf{Q}^{(k-1)})$,

the algorithm will give a solution satisfying the Karush-Kuhn-Tucker (KKT) conditions for the original problem [16]. The first condition ensures that the solution is feasible for the original problem. The second condition guarantees that the solution from the previous iteration is feasible for the current iteration. As a result, the objective function increases after every iteration, since the solution from the previous iteration is a feasible solution to the problem in the current iteration. The second and third conditions together guarantee that the KKT conditions for the original problem will be satisfied at convergence. As $R(\boldsymbol{Q})$ is bounded from above and monotonically increasing in each iteration, convergence is guaranteed. Since each $\mathbf{Q}_{i,i}^\epsilon$ is concave in $\mathbf{Q}$, we can easily verify that approximating them with the first-order Taylor expansion will give a function satisfying all the above conditions. Therefore we choose the $f_k(\mathbf{Q})$ as follows:

$$
f_k(\mathbf{Q}) = \sum_{i=1}^{N_T} \epsilon (Q_{i,i}^{(k-1)})^{\epsilon-1}(Q_{i,i} - Q_{i,i}^{(k-1)}) + (Q_{i,i}^{(k-1)})^\epsilon. \tag{16}
$$

When one of the $Q_{ii}^{(k-1)}$ approaches 0, the algorithm proceeds by setting the corresponding row and column of $\boldsymbol{Q}$ to zero and removes it from the optimization. To conclude, we obtain a suboptimal solution to (9). The procedure is summarized in Algorithm 1.

Algorithm 1 Successive convex optimization approach to the problem (9)

1: choose $\mathbf{Q}^{(0)}$ as a scaled identity matrix satisfying the constraints, and initialize $k = 1$
2: **repeat**
3:     form the k-th approximated problem (15) of (9) by approximating $f(\mathbf{Q})$ with $f_k(\mathbf{Q})$ around $\mathbf{Q}^{(k-1)}$ as in (16),
4:     solve the k-th approximated problem to obtain $\mathbf{Q}^{(k)}$,
5:     $k \leftarrow k + 1$
6: **until** convergence
7: **return** $\mathbf{Q}^{(k)}$

The proposed upper bound, lower bound and Algorithm 1 all involve the solution of a sequence of convex optimization problems. The number of problem instances that need to be

solved for the upper bound and the lower bound is K^{N_T}. For Algorithm 1, the number of problems depends on the number of iterations required for convergence.[1] The subproblems that are solved in each iteration in the computation of the upper bounds, lower bounds and in Algorithm 1 are all convex and can be solved using standard interior point methods. The number of variables is $N_\mathrm{T}(N_\mathrm{T}+1)/2$ and the number of constraints is of $\mathcal{O}(N_\mathrm{T})$. Therefore the computational complexity of solving each convex subproblem is roughly $\mathcal{O}(N_\mathrm{T}^6)$ [18].

To illustrate the importance of the consumed power constraint, and to motivate the need of choosing the right precoder under this constraint, we compare the energy efficiency of the precoding schemes with and without the total consumed power constraint. The energy efficiency is defined as the rate divided by the total consumed power, i.e.

$$EE \triangleq \frac{R(\boldsymbol{Q})}{\tilde{P}_\mathrm{tot}} = \frac{\log_2 \left| \boldsymbol{I} + \frac{1}{\sigma^2} \boldsymbol{H} \boldsymbol{Q} \boldsymbol{H}^H \right|}{\frac{P_\mathrm{max}^{1-\epsilon}}{\eta_\mathrm{max}} \sum_{i=1}^{N_\mathrm{T}} Q_{i,i}^\epsilon} \quad \text{(b/Hz/J)}. \tag{17}$$

Figure 1 shows a comparison of energy efficiencies between the proposed Algorithm 1 and the benchmark precoder that does not take the consumed power into account [4]. The simulation setup is a 4×4 MIMO system with $P_\mathrm{max} = 1$, $\eta = 0.55$, $\epsilon = 0.5$, $P_\mathrm{tot} = 1$ for Algorithm 1, and the channels are generated randomly from an uncorrelated Rayleigh distribution. From Figure 1 we observe that uniformly over SNR, Algorithm 1 provides a much higher energy efficiency than the benchmark. This shows that it is important to take both the radiated power and the power dissipation into account when designing the precoder.

IV. Monotonic Optimization Approach for Global Optimal Solution

Based on the upper and lower bounds derived in Sections III, we develop a monotonic optimization algorithm that obtains the globally optimal solution. The algorithm is based on the branch-and-bound philosophy.

One can easily verify that all constraints in problem (9) are normal sets, and the objective function is monotonically increasing in the diagonal elements of $\boldsymbol{Q}$. As a result, the monotonic optimization framework can be applied to problem (9) for finding the global optimum up to any pre-determined accuracy in finite time. The run time required to find the globally optimal solution can be very long, as the worst-case complexity grows exponentially with N_T. Nevertheless, the globally optimal solution serves as benchmark for the suboptimal Algorithm 1.

A. Branch and Bound Algorithm

To apply monotonic optimization techniques, we need to construct boxes that restrict the optimization variables. How-

[1]In practice, Algorithm 1 typically converges in less than 10 iterations; in some extreme cases it can take some 20 iterations. The exact number of iterations depend on the system parameters, the prescribed accuracy and the dimensions of the problem. In our simulations, Algorithm 1 takes about 7 iterations on average to converge and the maximum required number of iterations that we have experienced is 35.
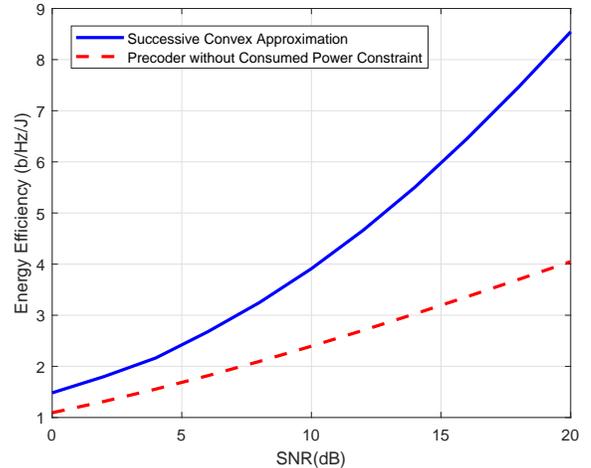

Fig. 1. Energy efficiency of the precoder obtained by Algorithm 1, compared to the benchmark precoder that does not take the consumed power into account. The reported energy efficiency is an average over 100 Monte-Carlo trials, for a 4×4 MIMO system with $P_\mathrm{max} = 1$, $\eta = 0.55$, $\epsilon = 0.5$, and $P_\mathrm{tot} = 1$.

ever, the optimization variable $\boldsymbol{Q}$ in (9) is restricted to be a positive semi-definite matrix. This makes the construction of box constraints hard. Nevertheless, we observe that the per-antenna constraints directly yield the box constraints

$$B_0: \quad \boldsymbol{0} \leq \mathrm{diag}(\boldsymbol{Q}) \leq P_\mathrm{max}. \tag{18}$$

Here the inequality sign $\leq$ is component-wise. We observe that the only non-convex constraint in the original problem is expressed only in terms of the values of the diagonal elements of $\boldsymbol{Q}$. Therefore, the upper and lower bounds on the capacity that we developed are also based on upper and lower bounding the values of the diagonal elements of $\boldsymbol{Q}$. Therefore when the diagonal elements of $\boldsymbol{Q}$ are determined (when the box shrinks to a point), the optimal input covariance matrix is determined as well. Thus performing branch-and-bound on the diagonal elements is sufficient.

The branch-and-bound algorithm starts with the initial box constraint B_0, and then computes the upper bound and lower bound on the capacity in the set B_0 as $U_0 = U(B_0)$ and $L_0 = L(B_0)$ where $U(\cdot)$ and $L(\cdot)$ are functions that give the upper respectively lower bounds on capacity. Then we compute the relative gap between the upper and lower bound as $(U_0 - L_0)/U_0$.

Next, the algorithm proceeds with a branching step: partition B_0 into two smaller boxes B_1 and B_2 with $B_0 = B_1 \cup B_2$, and delete the box B_0 that we partitioned. Then, subsequently, compute the corresponding upper respectively lower bounds on these smaller boxes and set $U_1 = \max(U(B_1), U(B_2))$ and $L_1 = \max(L(B_1), L(B_2))$. Since B_1 and B_2 are smaller boxes, the upper and lower bounds become tighter, i.e.

$$\begin{aligned} U_1 &\leq U_0, \\ L_1 &\geq L_0. \end{aligned} \tag{19}$$

This procedure is repeated and in each iteration we select the box with the largest upper bound and split it into two boxes along the longest edge (defined by the per-antenna power constraints), and the box that we partitioned is deleted. Then the upper and lower bounds are refined by taking the maximum over all the partitioned boxes. The algorithm terminates when the relative gap is smaller than the required accuracy δ. The above branch-and-bound algorithm can be viewed as a way of selecting the number of regions K that give upper respectively lower capacity bounds adaptively. From monotonic optimization theory, convergence to the global optimum is guaranteed [19]. The optimization yields both the optimal input covariance matrix, and the corresponding capacity up to a prescribed accuracy.

B. Pruning to Speedup

In this subsection we discuss two different pruning rules in implementing the branch-and-bound method specialized to our problem. First we observe that if two matrices $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ only differ in some entries on the main diagonal such that $Q_{kk} < \tilde{Q}_{kk}$, then we have

$$\det(I + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H) < \det(I + \frac{1}{\sigma^2}\boldsymbol{H}\tilde{\boldsymbol{Q}}\boldsymbol{H}^H). \quad (20)$$

Based on (20) we develop two rules for pruning the boxes generated in the branch-and-bound process. This pruning reduces the total computational complexity. The pruning rules are as follows:

- Prune all boxes with

$$\max_{\boldsymbol{Q} \in B_j} P_{\text{cons}} = f(Q) < P_{tot}$$

. From (20) we conclude there exists a box other than B_j that gives a higher rate. This condition can be easily tested by evaluating the consumed power in the upper corner of the box B_j, i.e.

$$\max_{Q_{ii} \in B_j} Q_{ii}, \forall i. \quad (21)$$

- Prune all boxes with

$$\min_{\boldsymbol{Q} \in B_j} P_{\text{cons}} = f(Q) > P_{tot}$$

. This implies that all points in the box B_j are infeasible due to the consumed power constraint. This condition can be easily tested by evaluating the consumed power in the lower corner of the box B_j, i.e.

$$\min_{Q_{ii} \in B_j} Q_{ii}, \forall i. \quad (22)$$

C. Numerical Example of Global Optimization

Figure 2 verifies the proposed global optimization algorithm and shows its convergence to the global optimum. The simulation setup is a 4×4 MIMO system with $P_{\max} = P_{\text{tot}} = 1$, $\delta = 0.001$ and $\sigma^2 = 0.1$, and the channels are generated randomly from an uncorrelated Rayleigh distribution. The rate is measured in bits per channel use (bpcu). In the simulations
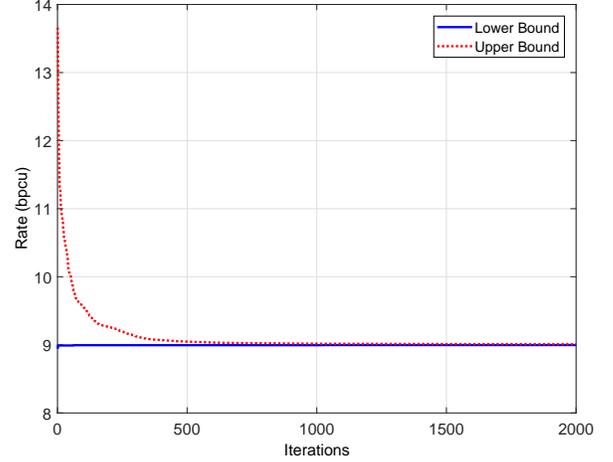

Fig. 2. An example showing the convergence of the global optimization (branch-and-bound) algorithm for a randomly generated $\boldsymbol{H}$. The upper and lower bounds on capacity are plotted against the number of iterations with $N_\text{T} = N_\text{R} = 4$, $P_{\max} = P_{\text{tot}} = 1$, $\delta = 0.001$ and $\sigma^2 = 0.1$.

we use Algorithm 1 to compute the lower bound for each box, by incorporating the box constraint into it. (Again, note that any feasible solution gives a lower bound on capacity.) From Figure 2 we see that the lower bound converges to the δ-optimal value in a few iterations, while the upper bound takes about 1000 iterations to converge.[2] This suggests that Algorithm 1 is performing close to the optimal solution, and using the result from Algorithm 1 is a good approximation for the actual capacity.

V. EXTENSION TO FREQUENCY-SELECTIVE POINT-TO-POINT MIMO

In a wideband system, the flat fading assumption does not hold in general as the channels are frequency selective. However, using multi-carrier techniques such as orthogonal frequency division multiplexing (OFDM) we can divide the whole frequency band into L different subcarriers. For each subcarrier l, we have a flat fading channel model:

$$\boldsymbol{y}_l = \boldsymbol{H}_l\boldsymbol{x}_l + \boldsymbol{n}_l, \quad l = 1, \dots, L, \quad (23)$$

where $\boldsymbol{y}_l \in \mathbb{C}^{N_R \times 1}$ is the received signal on the lth subcarrier, $\boldsymbol{H}_l \in \mathbb{C}^{N_R \times N_T}$ is the corresponding channel matrix, and $\boldsymbol{x}_l \in \mathbb{C}^{N_T \times 1}$ is the transmitted symbol vector. The noise vectors $\boldsymbol{n}_l \in \mathbb{C}^{N_R \times 1}$ are independent identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian, i.e., $\boldsymbol{n}_l \sim \mathcal{CN}(0, \sigma_l^2\boldsymbol{I})$. The model (23) assumes IFFT processing prior to transmission, however, the transmitted power is invariant under this transformation.

In a multi-carrier system, all subcarriers share the same set of antennas, and thereby the same set of amplifiers. Denote by

---

[2]In the numerical experiments that we have done, we have not observed any dependence of the convergence speed on a particular parameter. The convergence speed depends on all the parameters jointly and it is hard to estimate it in general.

$\boldsymbol{Q}^l \triangleq E[\boldsymbol{x}_l \boldsymbol{x}_l^H]$ the covariance matrix of the signal transmitted on the lth subcarrier. The radiated power on antenna i is

$$E\left[\sum_{l=1}^{L} |x_{li}|^2\right] = \sum_{l=1}^{L} Q_{i,i}^l. \tag{24}$$

The relation between the consumed power and the radiated power on antenna i is thus

$$\frac{\sum_{l=1}^{L} Q_{i,i}^l}{P_{\mathrm{cons},i}} = \eta_{\max}\left(\frac{\sum_{l=1}^{L} Q_{i,i}^l}{P_{\max}}\right)^{1-\epsilon}, \quad i = 1,\ldots,N_{\mathrm{T}} \tag{25}$$

with $\epsilon \in [0, 0.5]$ for different PAs.

As a result, the consumed power on antenna i can be written as

$$P_{\mathrm{cons},i} = \frac{1}{\eta_{\max}}\left(\sum_l Q_{i,i}^l\right)^{\epsilon} P_{\max}^{1-\epsilon}. \tag{26}$$

The total capacity of this multi-carrier system is now the sum of the capacities for each subcarrier:

$$R(\boldsymbol{Q}^1,\ldots,\boldsymbol{Q}^L) = \max_{\{\boldsymbol{Q}^l\}\in\mathcal{Q}} \sum_{l=1}^{L} \log_2 \left| \boldsymbol{I} + \frac{1}{\sigma_l^2}\boldsymbol{H}_l\boldsymbol{Q}^l\boldsymbol{H}_l^H\right|,$$

where $\mathcal{Q}$ specifies the set of power constraints, consisting of the consumed power constraints and per antenna power constraints. The constraint on the total consumed power is:

$$\frac{P_{\max}^{1-\epsilon}}{\eta_{\max}} \sum_{i=1}^{N_{\mathrm{T}}} \left(\sum_{l=1}^{L} Q_{i,i}^l\right)^{\epsilon} \leq \tilde{P}_{\mathrm{tot}}, \tag{27}$$

and per-antenna power constraint for antenna i is

$$\sum_{l=1}^{L} Q_{i,i}^l \leq P_{\max}, \quad i = 1,\ldots,N_{\mathrm{T}}. \tag{28}$$

Finally we formulate the optimization problem for finding the capacity under both the consumed power constraint and the per-antenna power constraints as follows:

$$\begin{aligned}
\underset{\{\boldsymbol{Q}^l\}}{\text{maximize}} \quad & R(\boldsymbol{Q}^1,\ldots,\boldsymbol{Q}^L) \\
\text{subject to} \quad & \sum_{i=1}^{N_{\mathrm{T}}} \left(\sum_{l=1}^{L} Q_{i,i}^l\right)^{\epsilon} \leq P_{\mathrm{tot}} \\
& 0 \leq \sum_{l=1}^{L} Q_{i,i}^l \leq P_{\max}, \ i = 1,\ldots,N_{\mathrm{T}} \\
& \boldsymbol{Q}^l \succeq \boldsymbol{0}, \quad l = 1,\ldots,L.
\end{aligned} \tag{29}$$

We observe that the mathematical structure of the problem (29) is similar to (9) in the flat fading case. The main difference is that the we have L input covariance matrices to optimize over. By adding an auxiliary variable $\boldsymbol{Q}^{\mathrm{sum}} = \sum_{l=1}^{L} \boldsymbol{Q}^l$, we rewrite problem (29) as

$$\begin{aligned}
\underset{\{\boldsymbol{Q}^l\},\boldsymbol{Q}^{\mathrm{sum}}}{\text{maximize}} \quad & R(\boldsymbol{Q}^1,\ldots,\boldsymbol{Q}^L) \\
\text{subject to} \quad & \sum_{i=1}^{N_{\mathrm{T}}} \left(\boldsymbol{Q}_{i,i}^{\mathrm{sum}}\right)^{\epsilon} \leq P_{\mathrm{tot}} \\
& 0 \leq \boldsymbol{Q}_{i,i}^{\mathrm{sum}} \leq P_{\max}, \ i = 1,\ldots,N_{\mathrm{T}} \\
& \boldsymbol{Q}^{\mathrm{sum}} \succeq \boldsymbol{0}, \ \boldsymbol{Q}^l \succeq \boldsymbol{0}, \quad l = 1,\ldots,L \\
& \boldsymbol{Q}^{\mathrm{sum}} = \sum_{l=1}^{L} \boldsymbol{Q}^l
\end{aligned} \tag{30}$$

Now the only non-convex constraint is from the consumed power constraint

$$\sum_{i=1}^{N_{\mathrm{T}}} \left(\boldsymbol{Q}_{i,i}^{\mathrm{sum}}\right)^{\epsilon} \leq P_{\mathrm{tot}},$$

which is of the same form as (11) in the flat fading case. The techniques for bounding, finding a local optimum and the global optimization can all be applied similarly to the problem (30) with straightforward modifications, and by replacing $\boldsymbol{Q}$ with $\boldsymbol{Q}^{\mathrm{sum}}$.

Note that while OFDM processing block-diagonalizes the channel such that inter-symbol interference resulting from time dispersion disappears; however, MIMO crosstalk interference is still present. Specifically, if $\boldsymbol{H}[t]$ is the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ matrix-valued impulse response of the channel, then $\boldsymbol{H}_l$ in (23) is given by the element-wise time-discrete Fourier transform, $\sum_t \boldsymbol{H}[t]e^{-j2\pi t/L}$, which is a non-diagonal matrix.

VI. EXTENSION TO MULTIUSER MIMO

In this section, we extend the analysis to a multiuser scenario, where a transmitter with N_{T} antennas is sending information to N_{R} different receiving antennas where these receiving antennas can either be cooperating or not cooperating with each other. We further assume that there are N_U users in the system where the N_{R} antennas are distributed among these N_U users.

Denote the input covariance matrices by $\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_{N_U}$. These matrices satisfy the per-antenna radiated power constraints,

$$\sum_{j=1}^{N_U} Q_{i,i}^j \leq P_{\max}, \quad i = 1,\ldots,N_{\mathrm{T}}, \tag{31}$$

and the total consumed power constraint:

$$\frac{P_{\max}^{1-\epsilon}}{\eta_{\max}} \sum_{i=1}^{N_{\mathrm{T}}} \left(\sum_{l=1}^{N_U} Q_{i,i}^l\right)^{\epsilon} \leq \tilde{P}_{\mathrm{tot}}. \tag{32}$$

The power constraints in the multiuser case are similar to in the multi-carrier case. The difference is that the sum is now over the users, rather than over the sub-carriers. Moreover, in multiuser MIMO, the concept of capacity does not exist. Instead one has to consider the entire capacity region, and the coding strategies required to achieve points in that region. In

the general case, the capacity region for the vector broadcast channel is the convex hull of the rate tuples:

$$R_{\phi(i)} \leq \log_2 \frac{|\sum_{j \geq i} \boldsymbol{H}_{\phi(j)} \boldsymbol{Q}^{\phi(j)} \boldsymbol{H}_{\phi(j)}^H + \boldsymbol{I}|}{|\sum_{j > i} \boldsymbol{H}_{\phi(j)} \boldsymbol{Q}^{\phi(j)} \boldsymbol{H}_{\phi(j)}^H + \boldsymbol{I}|}, \qquad (33)$$

where $R_{\phi(i)}$ is the rate achieved by user $\phi(i)$, and ϕ denotes a permutation on $[1 : N^U]$. This set of rates can be achieved by applying the dirty paper coding strategy [20] with the order $\phi(1) \to \ldots \to \phi(N^U)$. However, dirty paper coding has very high computational complexity which prohibits its use in practical systems. Therefore linear precoding strategies are preferred. Moreover, while in principle one could consider any operating point in the capacity region, we study the case when there are pre-determined rate requirements for the users, and minimize the consumed power required to achieve those requirements.

The general problem for different number of antennas at the users is non-convex, in the following we consider a special case in which the problem becomes convex and thus can be solved efficiently. Specifically, consider the case when the N_R receiving antennas do not cooperate. This corresponds to the case with N_R single antenna receivers. In this multiuser MIMO setup, the goal is to meet a predetermined quality-of-service (QOS) target for each user, and simultaneously minimize the total consumed power at the transmitter.

Under a flat fading channel model, the received signal for each user k can be expressed as

$$y_k = \boldsymbol{h}_k^H \boldsymbol{W} \boldsymbol{s} + n_k, \quad k = 1, \ldots, N_R, \qquad (34)$$

where $\boldsymbol{s} = [s_1, \ldots, s_{N_R}] \in \mathbb{C}^{N_R \times 1}$, s_k denotes the information symbol intended for user k, $n_k \sim \mathcal{CN}(0, \sigma^2)$ represents additive white Gaussian noise, and $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_R}]$ is the beamforming matrix where $\boldsymbol{w}_k = [w_{1k}, \ldots, w_{N_T k}]^T \in \mathbb{C}^{N_T \times 1}$ is the beamforming vector associated with user k. Note that all large-scale fading coefficients are included in the definition of the channel responses, $\boldsymbol{h}_k$.

With perfect CSI at the BS and at the users, an instantaneous achievable rate for user k can be written as

$$R_k = \log_2(1 + \gamma_k), \quad k = 1, \ldots, N_R, \qquad (35)$$

where γ_k represents the signal-to-interference-plus-noise-ratio (SINR) at user k, given by

$$\gamma_k = \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{i=1, i \neq k}^{N_R} |\boldsymbol{h}_k^H \boldsymbol{w}_i|^2 + \sigma^2}. \qquad (36)$$

Let $\boldsymbol{e}_i$ be the i-th column of the $N_T \times N_T$ identity matrix. The radiated power on antenna i can then be written as

$$\sum_{k=1}^{N_R} |w_{ik}|^2 = \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^2. \qquad (37)$$

The problem of minimizing the consumed power P_{tot}, under a constraint on the users SINRs is as follows:

$$\begin{aligned}
\underset{\{\boldsymbol{w}_k\}}{\text{minimize}} \quad & P_{\text{tot}} \\
\text{subject to} \quad & \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{i=1, i \neq k}^{N_R} |\boldsymbol{h}_k^H \boldsymbol{w}_i|^2 + \sigma^2} \geq \gamma_k, \quad \forall k, \\
& \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^2 \leq P_{\max}, \quad i = 1, \ldots, N_T
\end{aligned} \qquad (38)$$

To proceed, we express the consumed power in terms of the beamforming vectors $\{\boldsymbol{w}_k\}$. The relation between the consumed power and the radiated power on antenna i is

$$\begin{aligned}
& \frac{\|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^2}{P_{\text{cons},i}} = \\
& \eta_{\max} \left(\frac{\|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^2}{P_{\max}} \right)^{1-\epsilon}, \quad i = 1, \ldots, N_T
\end{aligned} \qquad (39)$$

with $\epsilon \in [0, 0.5]$ for different PAs. Furthermore, the consumed power on antenna i can be written as

$$P_{\text{cons},i} = \frac{1}{\eta_{\max}} \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^{2\epsilon} P_{\max}^{1-\epsilon}. \qquad (40)$$

By omitting irrelevant constants, the consumed-power minimization problem finally takes on the form

$$\begin{aligned}
\underset{\{\boldsymbol{w}_k\}}{\text{minimize}} \quad & \sum_i^{N_T} \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^{2\epsilon} \\
\text{subject to} \quad & \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{i=1, i \neq k}^{N_R} |\boldsymbol{h}_k^H \boldsymbol{w}_i|^2 + \sigma^2} \geq \gamma_k, \quad \forall k, \\
& \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_R}\|^2 \leq P_{\max}, \quad i = 1, \ldots, N_T
\end{aligned} \qquad (41)$$

Problem (41) is hard to solve in general as the objective is in general non-convex for $\epsilon \in [0, 0.5]$. However, we note that in the special case when $\epsilon = 0.5$, the problem can be solved efficiently. This special case was shown to be an accurate model for many classes of amplifiers when the output power is low [15, Equation (6.93)]. This was also confirmed by authors for class B amplifiers [10] and verified by experiments for class D amplifiers [21]. Then the objective function is a sum of vector norms, which is a convex function. Now the only non-convex constraints are the SINR constraints. However, we observe that the objective function and all constraints are invariant to a phase rotation in the following sense: If all elements of any of the beamforming vectors $\boldsymbol{w}_k$ are multiplied by the complex number $e^{j\theta_k}$, then the objective function and the constraints are unaffected. In particular, $\boldsymbol{w}_k$ may be phase-rotated such that $\Im(\boldsymbol{h}_k^H \boldsymbol{w}_k) = 0$. By exploiting this observation we can rewrite the SINR constraints as follows:

$$\sqrt{1 + \frac{1}{\gamma_k}} \Re(\boldsymbol{h}_k^H \boldsymbol{w}_k) \geq \left\| \begin{array}{c} \boldsymbol{h}_k^H \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{h}_k^H \boldsymbol{w}_{N_R} \\ \sigma \end{array} \right\|, \quad k = 1, \ldots, N_R, \qquad (42)$$

$$\Im(\boldsymbol{h}_k^H \boldsymbol{w}_k) = 0, \quad k = 1, \ldots, N_R,$$

which are now second-order cone constraints, respectively

linear constraints in the variables $\{\boldsymbol{w}_k\}$. The trick with the phase rotation that we exploit here was first introduced in a related context in [22].

After this transformation, we obtain the following second-order cone program:

$$\underset{\{\boldsymbol{w}_k\}}{\text{minimize}} \quad \sum_{i}^{N_{\text{T}}} \|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_{\text{R}}}\|$$

$$\text{subject to} \quad \sqrt{1 + \frac{1}{\gamma_k}} \Re(\boldsymbol{h}_k^H \boldsymbol{w}_k) \geq \left\| \begin{array}{c} \boldsymbol{h}_k^H \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{h}_k^H \boldsymbol{w}_{N_{\text{R}}} \\ \sigma \end{array} \right\|, \quad k = 1, \ldots, N_{\text{R}},$$

$$\Im(\boldsymbol{h}_k^H \boldsymbol{w}_k) = 0, \quad k = 1, \ldots, N_{\text{R}},$$

$$\|\boldsymbol{e}_i^T \boldsymbol{w}_1, \ldots, \boldsymbol{e}_i^T \boldsymbol{w}_{N_{\text{R}}}\|^2 \leq P_{\max}, \quad i = 1, \ldots, N_{\text{T}}. \tag{43}$$

As (43) is a convex problem, standard interior-point methods can be used to find the globally optimal solution.

VII. NUMERICAL EXPERIMENTS

We choose the parameters in the numerical experiments as follows. The maximum efficiency is selected to be $\eta_{\max} = 0.55$ which is a realistic value, cf. [23, eq. (6)]. The parameter ϵ is taken to be $\epsilon = 0.5$, which models a class B [10] or class D amplifier [21]. We want to investigate the operation of the system in the regime where the constraint on the consumed power is active. (In case the constraint on the consumed power were not active, our solution and the solution in [4] would be the same.) Hence we choose $P_{\max} = 1$, $\tilde{P}_{\text{tot}} = 1.818$ such that $P_{\text{tot}} = 1$. The normalization to unit power is made only for simplicity of the exposition, and the ratio between $P_{\max}$ and P_{tot} is constant, while the ratio between P_{tot} and σ^2 is varied. We define the signal-to-noise ratio (SNR) to be P_{tot}/σ^2, and rate is measured in bpcu. Note that the effect of the large-scale fading is embodied into the definition of the SNR.

A. Schemes Included in the Comparison

In order to evaluate our proposed methods in Section III-C, we consider the following baseline power allocation schemes:

1) Uniform Power Allocation: The power is uniformly allocated among all the transmit antennas, i.e. $Q_{i,i} = (P_{\text{tot}}/N_T)^{1/\epsilon}$, for $i = 1, \ldots, N_T$.

2) Antenna Selection: We use our algorithm in Section III-C combined with antenna selection. The motivation of using antenna selection is to avoid operating some antennas at an output power close to zero, as in this operating regime the efficiency is very poor. In addition, antenna selection can greatly reduce the complexity of the optimization procedure since we reduce the number of dimensions.
   We select the active antennas to be those for which the norms of the corresponding columns of $\boldsymbol{H}$ have the largest Euclidean norms. Thereafter, we run the algorithm in Section III-C only among these selected antennas.

3) Random Antenna Selection: We randomly choose a subset of the antennas and allocate all power among them using the algorithm in Section III-C.
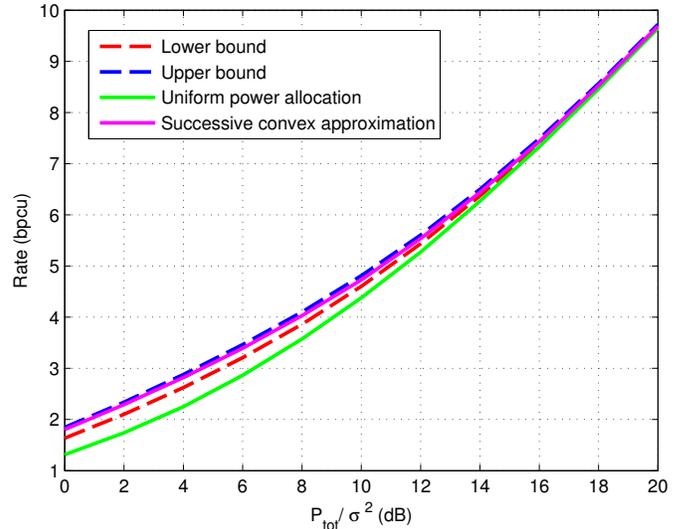

Fig. 3. Rate against SNR for a 2×2 MIMO system averaged over 100 channel realizations, $P_{\max} = 1$, and $P_{\text{tot}} = 1$. The number of sub-regions K in the generation of the upper and lower bounds is 10.

B. Simulation Results

Fig. 3 shows the rate as function of P_{tot}/σ^2 in a 2×2 MIMO system, averaged over 100 channel realizations. The number of sub-regions K for the generation of the upper and lower bounds is 10. We observe that at low SNR, the same rate is achieved with an SNR gain of 2.5 dB from optimizing the input distribution compared with the uniform power allocation scheme. We note that at low SNR, the optimal strategy is to choose a single antenna to which all power is allocated. At high SNR, all 4 curves converge, which suggests that uniform power allocation is close to optimal in that regime. This is analogous to the water-filling results in traditional MIMO with a sum radiated power constraint. The difference is that in our case the power is allocated among the antennas, not among the channel eigen-modes. Note that $\boldsymbol{Q}$ is positive semidefinite. hence if some of its diagonal elements is zero, say, the k-th element, then all elements in the k-th row and the k-th column also need to be zero.[3] The interpretation is that the k-th antenna is turned off. This suggest that we can turn off the hardware associated with some of the antennas and save power [24], [25].

Remark. *In conventional MIMO with total power constraints, we perform waterfilling over the eigenmodes. Then, uniform power allocation onto different eigenmodes is close to optimal at high SNR, and conversely, selecting the strongest eigenmode for transmission is close to optimal at low SNR. With the power constraint used in our paper, similar results hold, but with antennas taking the role of the eigenmodes instead.*

Next we look at the effect of having different total consumed power constraints in Figure 4. The simulation setup is a 2×2 MIMO system with $P_{\max} = 1$ and $\sigma^2 = 1$. We observe that as

---

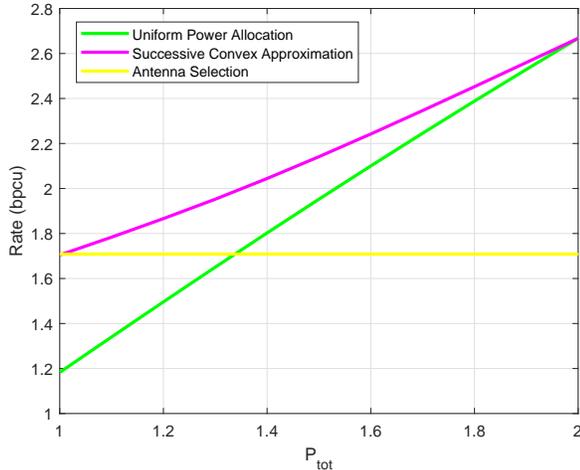[3]A proof of this fact is provided in the appendix.

Fig. 4. Rate as function of P_{tot} for a 2×2 MIMO system averaged over 100 channel realizations, for $P_{\max} = 1$ and $\sigma^2 = 1$.
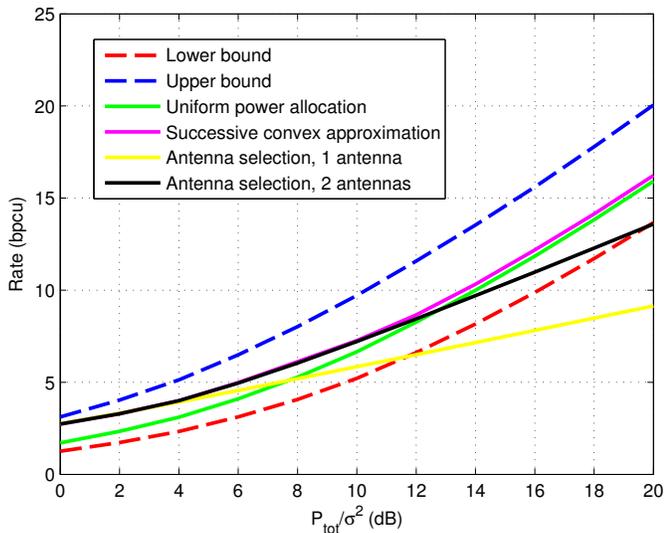

Fig. 5. Rate against SNR for a 4×4 MIMO system averaged over 100 channel realizations, $P_{\max} = 1$, and $P_{\text{tot}} = 1$. The number of sub-regions K for generating the upper and lower bounds is 3.

Fig. 6. Rate against SNR for a 10×2 MIMO system averaged over 100 channel realizations, $P_{\max} = 1$, and $P_{\text{tot}} = 1$.

allocation is close to optimal at high SNR, but sub-optimal at low SNR. The proposed algorithm in Section III-C performs best over the whole range of investigated SNR.

In a scenario when there are more transmit antennas than receive antennas, which is a practical case in downlink cellular communication, there is always a gain from optimizing the input distribution compared to uniform power allocation, as shown in Fig. 6. Moreover we observe that in this case using antenna selection scheme with 2 antennas is close to optimal in most cases. This suggests that carefully performing antenna selection such that the resulting effective MIMO system is of square dimension, does not lead to a significant rate loss. We also compare to the random antenna selection scheme with 1 antenna, which performs poorly in all regimes. Choosing the best antennas for antenna selection has a modest computational complexity of $O(N_T \log(N_T))$.

Next, we look at a larger MIMO system with $N_T = N_R = 10$. Fig. 7 shows the rate in this system for different P_{tot}/σ^2 and averaging over 100 channel realizations. We can see that in this case using a small number of antennas pays off more compared to the case with 2 receive antennas. There are still gains of the proposed antenna selection over random antenna selection in all SNR regimes, but the gap is getting smaller with increasing number of receive antennas. The gains reduce from around 1 bpcu in a 10×2 system to around 0.6 bpcu in a 10×10 system.

Finally we look at a multi-carrier system with $N_T = N_R = 4$ in Fig. 8. The number of sub-regions K for generating the upper and lower bounds is 3. We compare to the random antenna selection schemes suggested above with 1 antenna. The same conclusions as in the flat fading channels hold, i.e. the proposed antenna selection scheme is close to optimal at low SNR while uniform power allocation is close to optimal at high SNR. An important observation is that although random

P_{tot} is increased, the performance gap between uniform power allocation and Algorithm 1 becomes smaller and eventually it becomes 0 as the consumed power constraint is inactive when $P_{\text{tot}} = 2$. Meanwhile, the antenna selection scheme gives the same performance as Algorithm 1 when $P_{\max} = P_{\text{tot}}$, but it stays the same for all values of P_{tot} as it cannot benefit from the extra consumed power budget by using only one antenna.

Similar results are observed in a 4×4 MIMO system as illustrated in Fig. 5. The number of sub-regions K for generating the upper and lower bounds is 3. We compare to the antenna selection schemes suggested above with 1 and 2 antennas, which are close to optimal at low SNR, but non-optimal at high SNR. On the other hand, uniform power
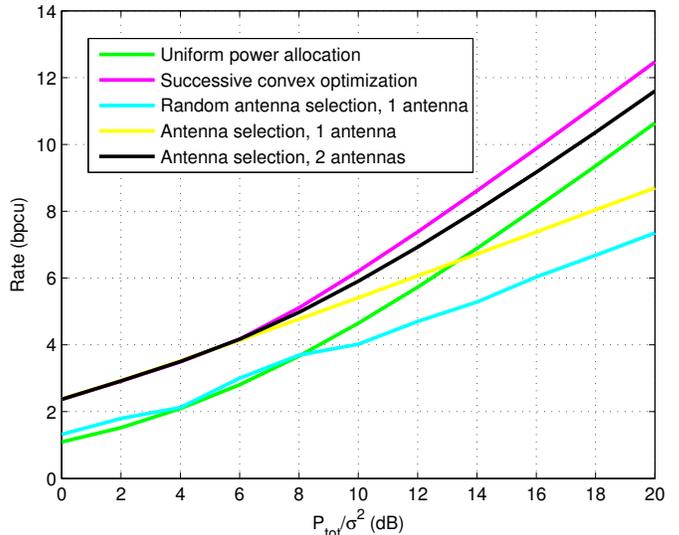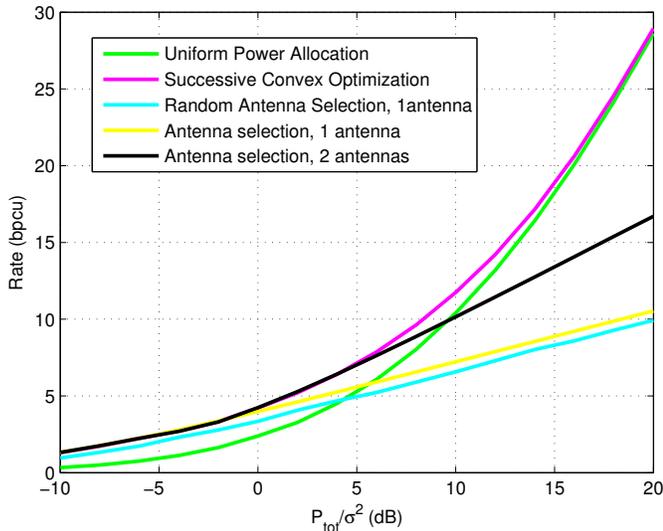
Fig. 7. Rate against SNR for a 10×10 MIMO system averaged over 100 channel realizations, $P_{\max} = 1$, and $P_{\text{tot}} = 1$.
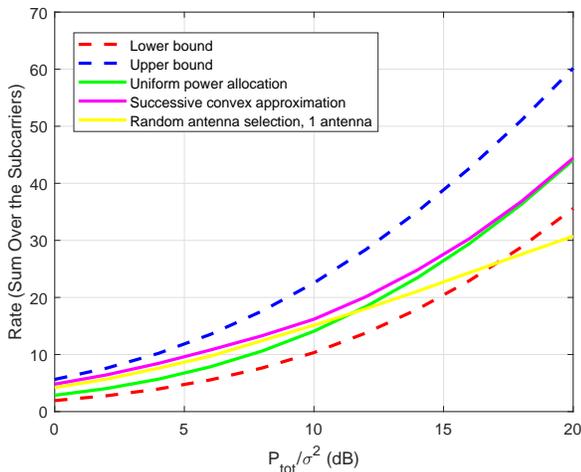

Fig. 8. Rate against SNR for a 4×4 MIMO multi-carrier system averaged over 100 channel realizations, $P_{\max} = 1$, and $P_{\text{tot}} = 1$ and the number of sub-carriers is 5. The number of sub-regions K for generating the upper and lower bounds is 3.

antenna selection is performing poorly in flat fading channels, it is close to optimal when we have more sub-carriers. The reason is that as we have more subcarriers, the strength of the channels seen by each antenna will be averaged out. Therefore which antenna to use does not make a big difference, which explains the good performance of the random antenna selection scheme in Fig. 8. As the number of subcarriers increases, we expect that the gap between the random antenna selection and Algorithm 1 will decrease.

C. Discussion

In all cases investigated, antenna selection results are showing up under the sum-consumed power constraint at low SNR, but not under a sum- radiated power constraint. This phenomenon can be intuitively explained by that the consumed power can be expressed as a constant times the ϵ-th power of the radiated power. Low radiated power will be penalized more compared to high radiated power, i.e. a lot of consumed power gives very little radiated power in this regime. This effect promotes allocating either a lot of power, or no power, to an antenna. Furthermore we observe that the constraint (11) is equivalent to

$$||\text{diag}(\boldsymbol{Q})||_\epsilon = \left(\sum_{i=1}^{N_t} Q_{i,i}^\epsilon\right)^{\frac{1}{\epsilon}} \leq P_{\text{tot}}^{1/\epsilon} \qquad (44)$$

This provides a link to the results in the compressed sensing literature. It is known that the ℓ_p-norm with $0 < p < 1$ enhances sparsity, and therefore it is used to replace the ℓ_0 norm, c.f. [26] and [27].

For a smaller ϵ, the sparsity enhancement effect is more significant, as we observed in preliminary simulations. Values of ϵ close to 0 are important for some classes of PAs, for example, [10] states that class A PAs have ϵ close to 0. We also note that there are other ways of performing antenna selection than by the maximum column norm described in Section VII-A, i.e. antenna selection may be an effective solution more often than what is observed in Fig. 3 to Fig. 7.

VIII. CONCLUSION

In this work we considered the capacity of the MIMO channel taking into account both a limitation of total consumed power and per-antenna radiated power constraints. For a fixed channel with full CSI at both the transmitter and the receiver, maximization of the mutual information was formulated as an optimization problem. Lower and upper bounds on the capacity were provided by numerical algorithms based on partitioning of the feasible region. Both bounds were shown to converge and give the exact capacity when number of regions increases. Later the bounds were used to construct a monotonic optimization algorithm based on the branch-and-bound approach. An efficient suboptimal algorithm based on successive convex approximation performing close to the capacity was also presented. Numerical results showed that the performance of the solution obtained from suboptimal algorithm is close the global optimal solution.

Simulation results showed that in the low SNR regime, antenna selection was the optimal scheme while at high SNR uniform power allocation was close to optimal. The sparsity of antennas being used increased when the parameter ϵ was decreased further. The connection with compressed sensing was established and our results indicated that antenna selection may play an important role in some future wireless systems. Although proposing antenna selection is not new, our results suggest that even only consumed power in the RF part is considered, it is better to use only a subset of antennas in some occasions.

APPENDIX A
PROOFS

Lemma 1. *If $Q \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{T}}$ is a positive semi-definite matrix and $Q_{kk} = 0$, then all elements in the k-th row and the k-th column of Q are also equal to 0.*

Proof: We prove this by contradiction. Assume there exists some nonzero element in the k-th row or column, $Q_{jk} = Q_{kj}^* \neq 0$, where without loss of generality $j > k$. Then consider the sub-matrix

$$A = \begin{pmatrix} 0 & Q_{kj} \\ Q_{jk} & Q_{jj} \end{pmatrix}. \tag{45}$$

We observe that $|A| = -|Q_{kj}|^2 < 0$; hence A is not positive semi-definite. Therefore, there exists a 2-dimensional vector $v = [v_1, v_2]^T$ such that

$$v^H A v < 0. \tag{46}$$

Now construct $x \in \mathbb{C}^{N_\mathrm{T}}$ to be the vector with its k-th element equal to v_1 and its j-th element equal to v_2, and all other elements equal to 0. Then we have

$$x^H Q x = v^H A v < 0, \tag{47}$$

which contradicts the assumption that Q is positive semi-definite. This concludes the proof. ∎

REFERENCES

[1] H. V. Cheng, D. Persson, and E. G. Larsson, "MIMO capacity under power amplifiers consumed power and per-antenna radiated power constraints," in *IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014, pp. 179–183.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] X. Zheng, Y. Xie, J. Li, and P. Stoica, "MIMO transmit beamforming under uniform elemental power constraint," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5395–5406, Nov. 2007.

[4] M. Vu, "The capacity of MIMO channels with per-antenna power constraint," *IEEE Transactions on Information Theory, submitted*, Jun. 2011. [Online]. Available: http://arxiv.org/abs/1106.5039

[5] C. Xing, Y. Ma, Y. Zhou, and F. Gao, "Transceiver optimization for multi-hop communications with per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1519–1534, Mar. 2016.

[6] C. Xing, Y. Jing, S. Wang, J. Wang, and J. An, "A general framework for covariance matrix optimization in MIMO systems," *CoRR*, vol. abs/1711.04449, 2017. [Online]. Available: http://arxiv.org/abs/1711.04449

[7] "EARTH - energy aware radio and network technologies deliverable d2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," 2012. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/d71252/EARTH_WP2_D2.3_v2.pdf

[8] M. Dohler, R. Heath, A. Lozano, C. Papadias, and R. Valenzuela, "Is the PHY layer dead?" *IEEE Communications Magazine*, vol. 49, no. 4, pp. 159 –165, Apr. 2011.

[9] G. Fettweis, M. Lohning, D. Petrovic, M. Windisch, P. Zillmann, and W. Rave, "Dirty RF: a new paradigm," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sep. 2005, pp. 2347–2355.

[10] A. He, S. Srikanteswara, K. K. Bae, T. Newman, J. Reed, W. Tranter, M. Sajadieh, and M. Verhelst, "Power consumption minimization for MIMO systems - a cognitive radio approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 469–479, 2011.

[11] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input single-output capacity," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 913–919, March 2014.

[12] D. Persson, T. Eriksson, and E. Larsson, "Amplifier-aware multiple-input multiple-output power allocation," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1112–1115, Jun. 2013.

[13] Y. Dong, Y. Huang, and L. Qiu, "Energy-efficient sparse beamforming for multiuser MIMO systems with nonideal power amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 134–145, Jan. 2017.

[14] Q. Cui, T. Yuan, and W. Ni, "Energy-efficient two-way relaying under non-ideal power amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1257–1270, Feb. 2017.

[15] A. Grebennikov, *RF and Microwave Power Amplifier Design, 1st ed.* New York, NY, USA: McGraw-Hill, 2005.

[16] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[17] M. Chiang, C. wei Tan, D. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[18] L. Vandenberghe, V. R. Balakrishnan, R. Wallin, A. Hansson, and T. Roh, *Interior-Point Algorithms for Semidefinite Programming Problems Derived from the KYP Lemma.* Springer Berlin Heidelberg, 2005, pp. 195–238.

[19] R. Horst, "A general class of branch-and-bound methods in global optimization with some new approaches for concave minimization," *Journal of Optimization Theory and Applications*, vol. 51, no. 2, pp. 271–291, Nov. 1986.

[20] A. E. Gamal and Y. H. Kim, *Network Information Theory.* Cambridge, U. K.: Cambridge University Press, 2012.

[21] H. Nemati, C. Fager, and H. Zirath, "High efficiency LDMOS current mode class-D power amplifier at 1 GHz," in *IEEE European Microwave Conference*, Sep. 2006, pp. 176 –179.

[22] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook of Antennas in Wireless Communications*, pp. 568–600, 2001.

[23] G. Tsouri and D. Wulich, "Impact of linear power amplifier efficiency on capacity of OFDM systems with clipping," in *IEEE Convention of Electrical and Electronics Engineers in Israel*, Dec. 2008, pp. 134–136.

[24] C. Yu and A. Zhu, "A single envelope modulator-based envelope-tracking structure for multiple-input and multiple-output wireless transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 10, pp. 3317–3327, Oct. 2012.

[25] "Texas Instruments RF front end power solutions for 2G, 3G, and 4G enabled devices," 2013. [Online]. Available: http://www.ti.com/ww/en/analog/power_management/envelope_tracking.htm

[26] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[27] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395 – 407, 2009.