Reducing the Number of Signaling Points Keeping Capacity and Cutoff Rate High

Anke Schmeink #1, Rudolf Mathar *2, Haoqing Zhang #3

UMIC Research Centre, RWTH Aachen University
52056 Aachen, Germany
1 anke.schmeink@rwth-aachen.de
3 haoqing.zhang@rwth-aachen.de

* *Inst. for Theoretical Information Technology, RWTH Aachen University*
52056 Aachen, Germany
2 mathar@ti.rwth-aachen.de

***Abstract*—Imposing so called peak power constraints on the input of a channel in many cases leads to a discrete capacity-achieving distribution with finite support. Given a finite number of signaling points ab initio, in this paper we determine reduced subsets in order to simplify the receiver design. The objective is to still keep the channel quality high. Two criteria are maximized, channel capacity and cutoff rate. By considering a uniform distribution over all signaling points, a lower bound to the general problem of finding an optimum signaling constellation in a bounded set and simultaneously the optimum input distribution is obtained. By semidefinite programming we show that even if only a small number of signaling points is selected the capacity and cutoff rate of the channel can be kept high.**

I. Introduction

Shannon's classical result shows that the scalar additive Gaussian noise channel subject to average power constraints achieves capacity if the input distribution is Gaussian as well. Telatar extended the result to complex circularly symmetric Gaussian vector channels, see [1]. This general model particularly applies to multiple-input multiple-output (MIMO) transmission system. Because of the unlimited support of the normal distribution, this input, however, is not realizable in practice. In order to avoid unbounded power requirements for the transmitter, peak power constraints of different types have been imposed. Interestingly, the capacity achieving input distribution then becomes discrete with finite support, as was shown in [2] for the real and in [3] for the complex Gaussian channel. A good overview of previous research on this topic is given in [4]. Especially for Poissonian channels, for channels with quadrature Gaussian and additive vector Gaussian noise distributions it is shown that the capacity-achieving input distribution subject to average and peak power constraints is discrete. By considering conditionally Gaussian vector channels subject to bounded-input constraints by some bounded set $\mathcal{S} \in \mathbb{R}^n$ this reference and [5] generalize a number of previous papers on the subject. Under certain conditions on $\mathcal{S}$ the capacity achieving distribution is discrete, which includes the previously mentioned channels as special cases. Non-coherent additive white Gaussian noise channels are investigated in [6] and it is shown that the optimum distribution is discrete. The same conclusion was shown for general fading channels in [7] and for Rician fading channels in [8]. Related topics are discussed in the following two references. In [9] a characterization for the optimum number of mass points is given. Reference [10] investigates the optimum constellation of M equiprobable complex signals for an additive Gaussian channel under average power constraints such that the error probability is minimum.

Summarizing the above, for practical purposes it is sufficient to investigate signaling constellations of a maximum number M of mass points. In this context, the following general problem of uttermost interest arises. Starting from a closed and bounded subset $\mathcal{S} \in \mathbb{R}^n$ of possible signaling points, determine a discrete input distribution consisting of a maximum number M of support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M \in \mathcal{S}$ and probabilities $P(\boldsymbol{X} = \boldsymbol{x}_i) = p_i$, $1 \le i \le M$, which maximizes mutual information between channel input and output, thus is capacity-achieving in the set of discrete distributions over $\mathcal{S}$ with at most M support points. Note that the optimum solution may exhibit $p_i = 0$ for some $i \in \{1, \ldots, M\}$ such that the number of effectively used points may be less than M. For the special case of conditional Gaussian channels and a nonrestricted number of signaling points, a partial answer is given in [4]. However, in general this seems to be a hard problem.

In this paper, we confine ourselves to a large, finite constellation set and ask the question of how to select a small subset of prescribed cardinality such that the capacity and cutoff rate is highest. The main purpose of using only a small number of signaling points is to simplify the receiver design and corresponding decoding algorithms.

The material in this correspondence is organized as follows. First, we introduce the precise system model and the problem description in Section II. In Section III we show how the problem can be transformed into a semidefinite program and solved by using two different relaxation techniques. Further, the case of sum power constraints is also considered in this section. Numerical results are presented in Section IV. The paper concludes with a short summary and outlook on future research in Section V.

978-1-4244-6316-9/10/$26.00 © 2010 IEEE

II. System Model and Prerequisites

Consider a channel with discrete input and continuous output. Let random variable $\boldsymbol{X}$ denote the discrete channel input with finite input alphabet $\mathcal{X}$ of possible signaling points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M \in \mathbb{R}^n$. These signaling points are used by the transmitter according to a certain input distribution $\boldsymbol{p} = (p_1, \cdots, p_M)$. The channel output $\boldsymbol{Y}$ corresponds to randomly distorted channel input. The distribution of $\boldsymbol{Y}$ given input $\boldsymbol{X} = \boldsymbol{x}_i$ is assumed to have Lebesque density

$$f(\boldsymbol{y}\,|\boldsymbol{x}_i) = f_i(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^n.$$

An example for such a channel is the additive channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}$ with $f_i(\boldsymbol{y}) = g(\boldsymbol{y} - \boldsymbol{x}_i)$ where g denotes the noise density.

Now assume that a finite set of signaling points is given under the above system model. In general, it is still a challenging task to select a subset of prescribed size from these points and simultaneously determine a distribution such that the channel capacity is maximized. We confine ourselves to the simpler, but for practical purposes most relevant problem of selecting a subset of given cardinality from the initial signaling points while using the uniform as input distribution. Hence, the task is reduced to deciding which signaling points should be included in the subset and which are excluded. Once having found the optimum configuration, mutual information may be improved through determining the optimum input distribution by help of the following proposition, which was shown in [11].

Proposition 1: Input distribution $\boldsymbol{p}^*$ is capacity achieving if and only if

$$D(f_i \,\|\, \sum_{j=1}^{M} p_j^* f_j) = \text{const},$$

for all i such that $p_i > 0$, where $D(f \,\|\, g) = \int f \log \frac{f}{g}$ denotes the Kullback-Leibler divergence between densities f and g. Furthermore, if $H(f_i) = -\int f_i(\boldsymbol{y}) \log f_i(\boldsymbol{y})\, d\boldsymbol{y}$ is independent of i, then $\boldsymbol{p}^*$ maximizes the mutual information over the set if and only if

$$\int f_i(\boldsymbol{y}) \log(\sum_{j=1}^{M} p_j^* f_j(\boldsymbol{y}))\, d\boldsymbol{y} = \text{const}$$

for all i such that $p_i > 0$.

Being more precise, the task is to find the equiprobable constrained-input alphabet capacity of the channel with input-alphabet with cardinality at most K, for some fixed number $K < |\mathcal{X}|$.

In what follows, we consider two different criteria to find the best subset $\mathcal{X}'$ of given size K.

A. The capacity maximizing subset

The direct approach is to find the subset of equiprobable input symbols that that maximizes the mutual information

$$\max_{\mathcal{X}' \subset \mathcal{X}} \sum_{\boldsymbol{x}_i \in \mathcal{X}'} \int p_i f(\boldsymbol{y}\,|\,\boldsymbol{x}_i) \log \frac{f(\boldsymbol{y}\,|\,\boldsymbol{x}_i)}{\sum_{\boldsymbol{x}_j \in \mathcal{X}'} p_j f(\boldsymbol{y}\,|\,\boldsymbol{x}_j)}\, d\boldsymbol{y}.$$

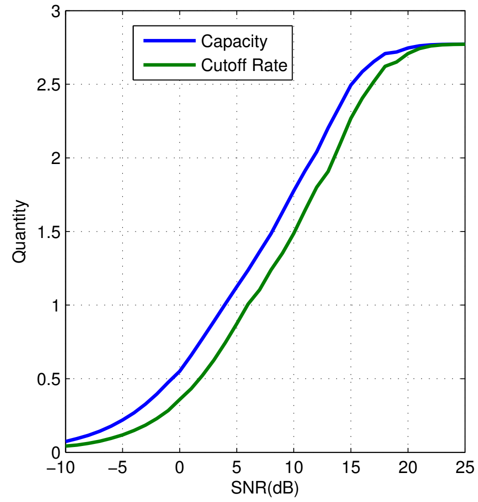

Fig. 1. Gap between the cutoff rate and the channel capacity

Using the assumption that $p_i = \frac{1}{K}$, equation (1) is equivalent to solving

$$\max_{\mathcal{X}' \subset \mathcal{X}} \sum_{\boldsymbol{x}_i \in \mathcal{X}'} \int f(\boldsymbol{y}\,|\,\boldsymbol{x}_i) \log \frac{f(\boldsymbol{y}\,|\,\boldsymbol{x}_i)}{\sum_{\boldsymbol{x}_j \in \mathcal{X}'} f(\boldsymbol{y}\,|\,\boldsymbol{x}_j)}\, d\boldsymbol{y}. \quad (1)$$

B. The cutoff rate maximizing subset

The second criterion we use is the maximization of the cutoff rate, which is a lower bound of the channel capacity [12],

$$\max_{\mathcal{X}' \subset \mathcal{X}} -\log \int \left[\sum_{\boldsymbol{x}_i \in \mathcal{X}'} p_i \sqrt{f(\boldsymbol{y}\,|\,\boldsymbol{x}_i)} \right]^2 d\boldsymbol{y}. \quad (2)$$

Figure 1 depicts the gap between the cutoff rate and the channel capacity using the setup given in Section IV. Using again that $p_i = \frac{1}{K}$ and that $\log$ is a monotone function, problem (2) transforms to

$$\min_{\mathcal{X}' \subset \mathcal{X}} \int \left[\sum_{\boldsymbol{x}_i \in \mathcal{X}'} \sqrt{f(\boldsymbol{y}\,|\,\boldsymbol{x}_i)} \right]^2 d\boldsymbol{y}. \quad (3)$$

This problem is easier to handle, as shown in the next section. We will use problem (3) to first generate a selection of subsets and then choose the one that maximizes mutual information. By doing so, we obtain a subset with reasonably high channel capacity.

III. The Subset Selection

In [13], a binary switching algorithm was introduced to increase the cutoff rate in discrete memoryless channels. The idea can be used to solve (3). As highest cutoff rate does not always lead to highest capacity, we use the cutoff rate approach to find distributions with relatively high capacity first, and then choose the best one among them. Problem (3) can be

Algorithm 1 Subset Selection Algorithm

1: **Initialization:**
$n = M+2$,
$$\boldsymbol{A} = \int \begin{bmatrix} \sqrt{f_1(\boldsymbol{y})} \\ \vdots \\ \sqrt{f_M(\boldsymbol{y})} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{f_1(\boldsymbol{y})} \\ \vdots \\ \sqrt{f_M(\boldsymbol{y})} \end{bmatrix}^T \mathrm{d}\boldsymbol{y},$$
$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0}_{2\times M} \\ \boldsymbol{0}_{M\times 2} & \boldsymbol{0}_{2\times 2} \end{bmatrix},$$
define an empty vector $\boldsymbol{r}$;

2: **Solve the semi-definite problem:**
$\hat{\boldsymbol{S}} = \mathrm{argmin}_{\boldsymbol{S}\geq \boldsymbol{0}}\ \mathrm{trace}(\boldsymbol{BS})$ s.t.
$S_{ii} = S_{in}, \quad \forall i,$
$S_{nn} = 1,$
$\sum_{i=1}^{M} S_{ni} = K,$
$\sum_{i=1}^{M} w_i S_{ni} + S_{n,M+1} = KW.$

3: **Adjustment:**
$n = M+1$,
delete the $(M+1)$th row and $(M+1)$th column of $\hat{\boldsymbol{S}}$,
$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0}_{1\times M} \\ \boldsymbol{0}_{M\times 1} & 0 \end{bmatrix}.$$

4: **Cholesky factorization:**
$\hat{\boldsymbol{S}} = \hat{\boldsymbol{V}}^T \hat{\boldsymbol{V}}$.

5: **Randomization:**

6: **for** $i = 1, \ldots, N_{rand}$ **do**

7: randomly generate a vector $\boldsymbol{u}^{(i)}$ uniformly distributed on a n-dimensional unit sphere;

8: computer $\tilde{\boldsymbol{s}}^{(i)} = \hat{V}^T \boldsymbol{u}^{(i)}, \quad \forall i;$

9: $\tilde{s}_n^{(i)} \leftarrow \mathrm{sign}(\tilde{s}_n^{(i)});$

10: $\tilde{\boldsymbol{s}}^{(i)} \leftarrow \tilde{s}_n^{(i)} \tilde{\boldsymbol{s}}^{(i)};$

11: quantize the K highest entries of $\left[\tilde{s}_1^{(i)}, \ldots, \tilde{s}_M^{(i)}\right]$ to 1 and the others to 0;

12: **if** $\sum_{\tilde{s}_j^{(i)}=1} w_j \leq KW$ **then**
$t = \tilde{\boldsymbol{s}}^{(i)T} \boldsymbol{B} \tilde{\boldsymbol{s}}^{(i)}$,

13: **else continue**

14: **end if**

15: **if** t is not yet in vector $\boldsymbol{r}$ **then**
$r_i = t$,
calculate channel capacity C_i based on $\tilde{\boldsymbol{s}}^{(i)}$.

16: **end if**

17: **end for**

18: **Choose** $\tilde{\boldsymbol{s}} = \mathrm{argmax}_{\tilde{\boldsymbol{s}}^{(i)}} C_i$.

19: **Take** $\boldsymbol{b} = [\tilde{s}_0, \ldots, \tilde{s}_M]^T$ as approximate solution.

reformulated as a constrained binary quadratic minimization problem (BQP)

$$\begin{array}{ll} \min_{\boldsymbol{b}\in\{0,1\}^M} & \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b} \\ s.t. & \boldsymbol{1}_M^T \boldsymbol{b} = K \end{array}$$

where the ij-th entry of $\boldsymbol{A}$ is $\int \sqrt{f_i(\boldsymbol{y}) f_j(\boldsymbol{y})}\, d\boldsymbol{y}$. Vector $\boldsymbol{b}$ is a binary vector where a one indicates that the corresponding symbol is included in the subset, a zero that the corresponding element is excluded.

We further assume that the power consumption of signaling point i is w_i. Introducing a total power constraint then yields the additional constraint $\boldsymbol{w}^T \boldsymbol{b} \leq W$ with $\boldsymbol{w} = (w_1, \cdots, w_M)^T$. By using slack variable $b_{M+1} \geq 0$, the inequality is replaced by equality $\boldsymbol{w}^T \boldsymbol{b} + b_{M+1} = W$. To make the optimization problem symmetrical, thus gaining access to a formulation as a convex problem, we introduce another vector $\boldsymbol{s} = (s_1, \cdots, s_n)^T$ with $n = M+2$ and the meaning

$$\boldsymbol{b} = s_n (s_1, \cdots, s_M)^T,$$
$$b_{M+1} = s_n s_{M+1}$$

where $s_n \in \{-1, 1\}$. Then, the above problem is equivalent to

$$\begin{array}{lll} \min\limits_{\boldsymbol{s}} & \boldsymbol{s}^T \boldsymbol{B} \boldsymbol{s} & (4) \\ \text{s.t.} & s_i s_n = s_i^2 \quad \forall i, & \\ & s_n^2 = 1, & \\ & s_n \boldsymbol{1}_M^T (s_1, \ldots, s_M)^T = K, & \\ & s_n \boldsymbol{w}^T (s_1, \ldots, s_M)^T + s_n s_{M+1} = KW, & \end{array}$$

with

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0}_{2\times M} \\ \boldsymbol{0}_{M\times 2} & \boldsymbol{0}_{2\times 2} \end{bmatrix}.$$

Using substitution $\boldsymbol{S} = \boldsymbol{s}\boldsymbol{s}^T$ it can be shown that the above optimization problem can be rewritten as

$$\hat{\boldsymbol{S}} = \mathrm{argmin}_{\boldsymbol{S}\geq \boldsymbol{0}}\, \mathrm{trace}(\boldsymbol{BS}) \quad (5)$$
$$s.t.\ S_{ii} = S_{in} \forall i, S_{nn} = 1, \textstyle\sum_{i=1}^{M} S_{ni} = K,$$
$$\textstyle\sum_{i=1}^{M} w_i S_{ni} + S_{n,M+1} = KW, \mathrm{rank}(S) = 1.$$

The semidefinite programming (SDP) relaxation of (5) can be efficiently solved in polynomial time [14]. If the resulting matrix $\hat{\boldsymbol{S}}$ has rank one, then the relaxation is tight. Otherwise, special techniques are required to convert the SDP relaxation solution back into an approximate BQP solution, see, e.g., [15], [16]. Using any of these relaxation techniques, we then obtain the set of estimators $\{\tilde{\boldsymbol{s}}\}$. Before remapping it back to vector $\boldsymbol{b}$, we have to examine whether the estimator fulfills the average power constraint (see the comment in the next paragraph) and find the one which maximizes the channel capacity among all those. The resulting vector $\tilde{\boldsymbol{s}}$ leads to the approximate solution of the original binary vector $\boldsymbol{b}$. The algorithm is summarized in Algorithm 1.

For clarification, we would like to make some remarks to Algorithm 1. Originally, $\tilde{\boldsymbol{s}}^{(i)}$ fulfills the average power constraint. However, after the quantization in Step 11, it is very probable that the average power overflows. This is the reason that rechecking in Step 12 is necessary even if the average power constraint has been already added before. The metric to choose $\tilde{s}$ differce from [13]. It is a traditional way to use

$$\tilde{\boldsymbol{s}} = \mathrm{argmin}\, \tilde{\boldsymbol{s}}^{(i)} \boldsymbol{s}^{(i)^T} \boldsymbol{B} \boldsymbol{s}^{(i)}$$

which directly maximizes the cutoff rate as expected. However, there exists a gap between cutoff rate and channel capacity. To keep always the best result among the randomizations alive, we

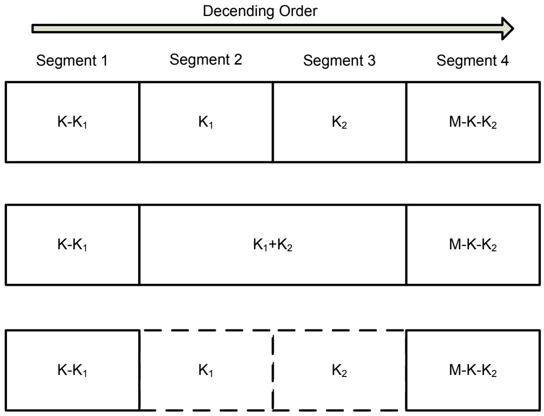

Fig. 2. Heuristic method to improve the subset selection.

directly calculate capacity for each candidate of $\tilde{s}$ in Step 15. Further, Step 15 helps to reduce the computational load since sometimes different randomization loops result in the same or equivalent selections due to a symmetric noise distribution. We do not need to include any selection which has already occured.

The solution can be further improved by applying the following heuristic method. For the sake of readability, we mention this sub-algorithm separately from Algorithm 1. Consider the not yet quantized vector $\tilde{\boldsymbol{s}}^{(i)}$ in Step 10 of Algorithm 1 and sort the first M entries in descending order. In Algorithm 1, the highest K entries are chosen as a single lot. Obviously this selection is not necessarily optimum. This holds as the order is based on the weight of each signaling point and thus a higher weight implies a higher probability to be a better point. As exhaustive search is too complex to apply even for moderate problem sizes, we use a simple heuristic method to reduce computational complexity of an optimal exhaustive search. In the method, the unquantized first M entries of $\tilde{\boldsymbol{s}}^{(i)}$ are kept in the buffer, adding $\tilde{\boldsymbol{s}}_{un}^{(i)} = [\tilde{s}_1^{(i)}, \cdots, \tilde{s}_M^{(i)}]$ between Step 10 and 11. In Step 18, we obtain $\tilde{s}$ and the pair $\tilde{s}_{un} = \tilde{\boldsymbol{s}}_{un}^{(i)}$, which we divide into four segments in descending order:

- the largest $K - K_1$ entries,
- the second largest K_1 entries,
- the third largest K_2 entries and
- the smallest $M - K - K_2$ entries

where $K_1, K_2 \in \mathbb{Z}$ and $0 \leq K_1 \leq K, 0 \leq K_2 \leq M - K$. The first segment is included in the final subset, while the fourth is excluded. From the second and third segment, we choose the best K_1 entries to obtain the new subset. Figure 2 depicts how this method works. The solution is already improved by rather small K_1 and K_2 such that the computational load of this heuristic method is negligible.

IV. Simulation Results

For our simulations, we use the following scenario . We aim to choose $K = 16$ signaling points from a M-QAM scenario with $M = 64$. 64-QAM points in the square $[-3, 3]^2$ are used as initial configuration. We consider 2-dimensional Gaussian noise with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with varying parameter ρ. As outlined above, using a uniform distribution over the selected subset provides a lower bound for the capacity-achieving non-uniform distribution.

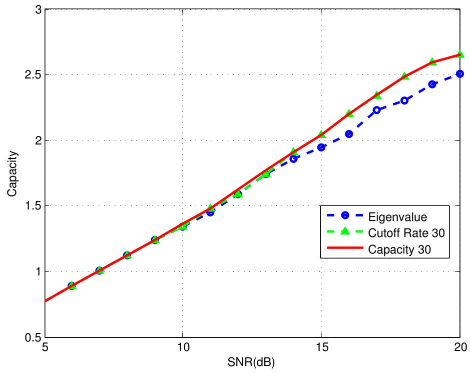

Fig. 3. SDP relaxation comparison by changing SNR

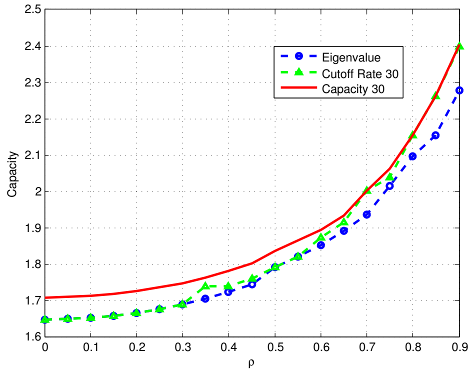

Fig. 4. SDP relaxation comparison by changing ρ

First, we compare different SDP relaxation techniques, see Figure 3. The blue curve is the channel capacity obtained by applying the dominant eigenvector approximation, see [16], to acquire the SDP relaxation. The green and red curve use randomization with 30 loops, i.e., an estimator of $\boldsymbol{s}$ is obtained after each randomization loop. The red one aims at maximizing the cutoff rate, while the green one directly maximizes the channel capacity. As can be seen in the figure and as expected, maximizing the channel capacity with randomization always performs best, though the difference is interestingly not very large.

Figure 4 shows the capacity for different values of ρ for the above two SDP techniques. Again, randomization with maximizing channel capacity performs best. This plot also tells us, that under certain average noise power, more correlated noise with $|\rho|$ large is better for finding a higher channel capacity.

Concerning the number of randomization loops, Figure 5 shows both cutoff rate and capacity oriented methods for 10, 30 ad 100 loops. In the right plot, the curves become quite smooth. This holds, because the a total power of the selected

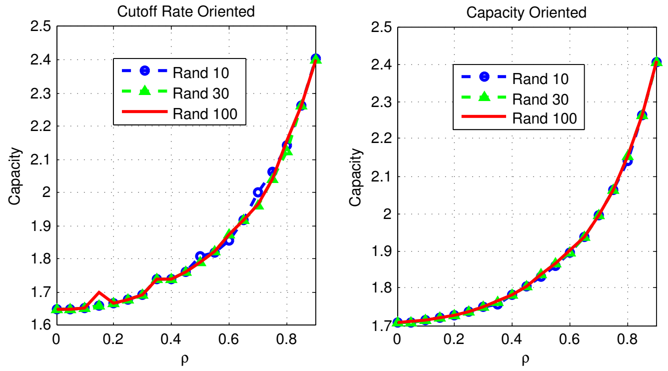

Fig. 5. Influence of the number of randomizations

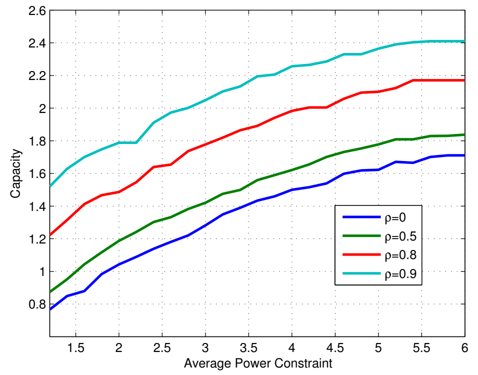

Fig. 6. Channel capacity with power constraints

signaling points depends on the specific selection and thus, a larger cutoff rate does not imply a larger channel capacity. We can conclude, that a reasonable number of randomizations, e.g., $N_{rand} = 30$, is close to maximize capacity in all considered cases.

The resulting signaling schemes will be changed when a total power constraint is applied. Figure 6 shows the channel capacity vs. the total power constraint for four different values of the correlation parameter ρ for a given total power constraint per dimension. The curves are monotone increasing when the total power becomes larger. This holds, because the constraint becomes much looser. As expected, each curve converges to its maximum, which is the same as the channel capacity without total power constraint.

V. Conclusions

In this paper, we contributed to the challenging problem of finding a reduced, optimal set of signaling points. We approached this problem by assuming a uniform distribution on the selected signaling points. By doing this, we obtain a lower bound for the unconstrained-channel capacity. We considered both an unconstraine-channel capacity and cutoff rate maximizing subset selection. We investigated these problems by forming a semidefinite programing problem and solving it with two different relaxation techniques. The obtained lower bound is close to the channel capacity of the full set of signaling points. This shows that using only a small size subset can indeed achieve a very high capacity even compared with the large full input set. This approach helps to highly simplify the receiver design while maintaining a high transmission rate over the channel.

Future research will be devoted to finding the corresponding optimal probability distribution which is the next step in tackling the full problem. Further, the analysis will be extended to non-regular constellation points, thereby allowing the I and Q components to be demodulated independently.

Acknowledgment

This work was supported by the UMIC Research Center in the framework of the German government excellence initiative.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[2] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, no. 6, pp. 203–219, 1971.

[3] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060 – 1071, Jul. 1995.

[4] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[5] T. H. Chan and F. R. Kschischang, "On the discreteness of the capacity achieving probability measure of conditional gaussian channels," in *Proceedings ISIT*, Chicago, June 2004, p. 347.

[6] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[7] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memory-less rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[8] M. C. Gursoy, H. V. Poor, and S. Verdu, "The noncoherent rician fading channel–part I: structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2193–2206, Sep. 2005.

[9] S. Sharma and S. Shamai, "Characterizing the discrete capacity achieving distribution with peak power constraint at the transition points," in *IEEE International Symposium on Information Theory and Its Applications (ISITA)*, Dec. 2008, pp. 1–6.

[10] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Transactions on Communications*, vol. 22, no. 1, pp. 28–38, January 1974.

[11] R. Mathar, A. Schmeink, and M. Zivkovic, "Optimum discrete signaling over channels with arbitrary noise distribution," in *2nd International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, Australia, Dec. 2008.

[12] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, August 2001.

[13] A. Mezghani, M. Ivrlac, and J. Nossek, "Achieving near-capacity on large discrete memoryless channels with uniform distributed selected input," in *IEEE International Symposium on Information Theory and Its Applications (ISITA)*, Dec. 2008, pp. 1–6.

[14] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," 1999.

[15] A. Wiesel, Y. Eldar, and S. Shitz, "Semidefinite relaxation for detection of 16-qam signaling in mimo channels," *Signal Processing Letters, IEEE*, vol. 12, no. 9, pp. 653–656, Sept. 2005.

[16] W.-K. Ma, P.-C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for m-ary psk multiuser systems," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 2862–2872, Oct. 2004.