Capacity-achieving probability measure for a reduced number of signaling points

Anke Schmeink · Haoqing Zhang

Published online: 9 March 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Putting bounding constraints on the input of a channel leads in many cases to a discrete capacity-achieving distribution with a finite support. Given a finite number of signaling points, we determine reduced subsets and the corresponding optimal probability measures to simplify the receiver design. The objective for the subset selection is to keep the channel quality high by maximizing mutual information and cutoff rate. Two approaches are introduced to obtain a capacity-achieving probability measure for the reduced subset. The first one is based on a preceded signaling point selection while the second one chooses the signaling points and corresponding probabilities simultaneously. Numerical results for both approaches show that using only a small number of signaling points achieves a very high mutual information compared to channels utilizing the full set of signaling points.

Keywords Bounded-input constraint · Capacity-achieving measure · Blahuta-arimoto algorithm · Reduced signaling set · Discrete signaling

1 Introduction

By Shannon, the scalar additive Gaussian noise channel subject to average power constraints achieves capacity if

the input distribution is Gaussian as well. This result transfers to complex circularly symmetric Gaussian vector channels, as Telatar showed in [1]. His general model particularly applies to multiple-input multiple-output (MIMO) transmission systems. However, due to the unlimited support of the normal distribution, this input is not realizable in practice. Thus, peak power constraints of different types have been imposed to avoid unbounded power requirements for the transmitter. It is a very interesting implication, that the capacity achieving input distribution then becomes discrete with finite support. Many challenging questions arise in this context. A good overview of previous research on this topic is given in [2]. The discreteness of the capacity achieving distribution was shown in [3] for the real and in [4] for the complex Gaussian channel. By considering conditionally Gaussian vector channels subject to bounded-input constraints by some bounded set $S \in \mathbb{R}^n$ reference [5] generalizes a number of previous papers on the subject. Under certain conditions on 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 minimized.

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

A. Schmeink (✉) · H. Zhang
UMIC Research Centre, RWTH Aachen University,
52056 Aachen, Germany
e-mail: anke.schmeink@rwth-aachen.de

H. Zhang
e-mail: haoqing.zhang@rwth-aachen.de

general problem of outermost interest arises. Given 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 $x_1, \dots, x_M \in \mathcal{S}$ and probabilities $P(X = x_i) = p_i$, $1 \leq i \leq M$, which maximizes the 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 $i \in \{1, \dots, M\}$ such that the number of effectively used points may be less than M . For the special case of conditional Gaussian channels and a non restricted number of signaling points, a partial answer is given in [2]. 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 both a small subset of prescribed cardinality and the associated probability measure such that the mutual information is highest, continuing some of the topics in [11]. As analytical results are extremely difficult to achieve, it is of interest to find algorithms that provide good input distributions and probabilities. The main purpose of using only a small number of signaling points is to simplify the receiver design and corresponding decoding algorithms.

The classical Blahut-Arimoto algorithm computes in an elegant way the capacity of a discrete memoryless channel [12, 13]. In the context of our distribution selection, we extend the algorithm to the considered case of discrete input and continuous output. We prove convergence of the extended algorithm by means of some optimality criterion. As the standard Blahut-Arimoto algorithm is computationally quite complex, several enhancements have been proposed. The most important ones, the natural-gradient-based algorithm and the accelerated Blahut-Arimoto algorithm are mentioned in [14]. They converge significantly faster to the capacity-achieving distribution and can be extended to our system model. References [15–17] suggest another interesting enhancement, the iterated Blahut-Arimoto algorithm. It was proposed for discrete memoryless channels and we extend it to a discrete input continuous output channel.

The material in this correspondence is organized as follows. First, we introduce the precise system model and the problem description in Sect. 2. In Sect. 3 we solve the subset selection by utilizing semidefinite programming and two different relaxation techniques. The probability measure optimization is considered in Sect. 4. Two different approaches are analyzed. The first one is a successive subset probability measure selection, while the second one updates both chosen signaling points and probability measure simultaneously. Numerical results are presented in Sect. 5. The paper concludes with a short summary and outlook in Sect. 6.

2 System model and prerequisites

We consider a channel with discrete input and continuous output. Random variable X denotes the channel input with finite input alphabet $\mathcal{X}$ of possible signaling points $x_1, \dots, x_M \in \mathbb{R}^n$. These signaling points are used by the transmitter according to a certain input distribution $\mathbf{p} = (p_1, \dots, p_M)$. The channel output Y is randomly distorted by noise. The distribution of Y given input $X = x_i$ is assumed to have Lebesgue density

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

An example for such a channel is the additive channel $Y = X + \mathbf{n}$ with $f_i(y) = g(y - x_i)$ where g denotes the noise density. Further the transition probability of the input x_i given output y is given by $P(x_i|y)$. The set of all $P(x_i|y)$ is called $\mathcal{P}$.

As outlined in the introduction, a challenging task is to choose a fixed size subset from the whole finite support and the corresponding distribution such that the channel is utilized optimally in terms of maximizing the mutual information between the channel input and the channel output. Being more precise, the task is to find a subset $\mathcal{X}' \subset \mathcal{X}$ with cardinality K for some fixed number K and the associated probability measure $\mathbf{p} = (p_1, \dots, p_M)$ where $p_i = 0$ for $i \in \mathcal{X} \setminus \mathcal{X}'$.

As was shown in [18], a necessary and sufficient condition for a distribution to be capacity achieving is the following proposition. This proposition will be used to show the optimality of our converging algorithm in Sect. 4.

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

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

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 $H(f_i) = - \int f_i(y) \log f_i(y) dy$ is independent of i , then $\mathbf{p}^*$ is capacity achieving if and only if

$$\int f_i(y) \log \left(\sum_{j=1}^M p_j^* f_j(y) \right) dy = \text{const}$$

for all i such that $p_i > 0$.

3 Selecting the subset

The problem of choosing both a fixed size subset and the corresponding distribution is rather complex. Therefore, we first consider a simplified approach by adding the assumption of a uniform distribution for the subset. Hence,

the task is reduced to only decide which signaling points should be included in the subset and which not, i.e., the task is to find a subset $\mathcal{X}' \subset \mathcal{X}$ with cardinality K such that $\mathbf{X}$ with distribution

$$p_i = \begin{cases} 1/K & \text{if } \mathbf{x}_i \in \mathcal{X}' \\ 0 & \text{else} \end{cases}$$

is capacity-achieving in the set of uniform distributions on $\mathcal{X}$ with at most K support points.

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

The capacity achieving subset: The most intuitive approach is to find the subset that maximizes the mutual information between the channel input and the channel output,

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

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

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

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 [19]. Figure 1 depicts the gap between the cutoff rate and the channel capacity using the setup given in Sect. 5, thereby showing the good performance of the lower bound. The cutoff rate with respect to our channel model is

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

Using again that $p_i = \frac{1}{K}$ and that $\log$ is a monotone function, the above equation transforms to

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

In what follows, we will show that this problem can be approached by exploiting semidefinite programming. We will use problem (2) to generate a selection of subsets and then decide on the one maximizing the mutual information. Thereby, we will see that we obtain a very good approximation to the solution of problem (1).

3.1 Subset selection using semidefinite programming

We assume, the power consumption of signaling point i to be w_i and introduce a binary switching algorithm to solve (2). As highest cutoff rate does not always lead to highest mutual information, we use the cutoff rate approach to find distributions with relatively high mutual information first,

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

and then choose the best one among them. The methodological approach is as follows. Problem (2) can be transformed into a constrained binary quadratic minimization problem, see also [20] where a similar approach was applied to increase the cutoff rate in discrete memoryless channels. We obtain problem

$$\begin{aligned} \min_{\mathbf{b} \in \{0,1\}^M} & \mathbf{b}^T \mathbf{A} \mathbf{b} \\ \text{s.t.} & \mathbf{1}_M^T \mathbf{b} = K \\ & \mathbf{w}^T \mathbf{b} + b_{M+1} = W. \end{aligned}$$

where the ij -th entry of $\mathbf{A}$ is $\int \sqrt{f_i(\mathbf{y})f_j(\mathbf{y})} d\mathbf{y}$. Vector $\mathbf{b}$ is a binary vector where a “1” indicates that the corresponding symbol is included in the subset, a “0” that the corresponding element is excluded. Further, slack variable $b_{M+1} \geq 0$, transforms the power constraint $\mathbf{w}^T \mathbf{b} \leq W$ with $\mathbf{w} = (w_1, \dots, w_M)^T$ to $\mathbf{w}^T \mathbf{b} + b_{M+1} = W$. For the derivation of the new objective, assume that $\mathcal{X}' = \{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_K}\}$. Then, it holds that

$$\begin{aligned} \int \left[\sum_{\mathbf{x}_i \in \mathcal{X}'} \sqrt{f(\mathbf{y}|\mathbf{x}_i)} \right]^2 d\mathbf{y} &= \int \left[\sum_{\mathbf{x}_i \in \mathcal{X}'} \sqrt{f_{i_1}(\mathbf{y})} \right]^2 d\mathbf{y} \\ &= \int \left[\sqrt{f_{i_1}(\mathbf{y})} + \dots + \sqrt{f_{i_K}(\mathbf{y})} \right]^2 d\mathbf{y} \\ &= \int \sum_{i,j \in \{1, \dots, K\}} \sqrt{f_{i_1}(\mathbf{y})} \sqrt{f_{i_j}(\mathbf{y})} d\mathbf{y} \\ &= \sum_{i,j \in \{1, \dots, K\}} \int \sqrt{f_{i_1}(\mathbf{y})} \sqrt{f_{i_j}(\mathbf{y})} d\mathbf{y} \\ &= \sum_{i,j} b_i b_j \int \sqrt{f_{i_1}(\mathbf{y})} \sqrt{f_{i_j}(\mathbf{y})} d\mathbf{y} \\ &= \mathbf{b}^T \mathbf{A} \mathbf{b}, \end{aligned}$$

thus, we obtain the new objective.

To make the optimization problem symmetrical and thus suitable for a convex formulation, we introduce another vector $\mathbf{s} = (s_1, \dots, s_n)^T$ with $n = M + 2$ and

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

where $s_n \in \{-1, 1\}$. The values of s_i result from s_n . If $s_n = 1$, then $s_i \in \{0, 1\}$; if $s_n = -1$, then $s_i \in \{-1, 0\}$. It follows that the above problem is equivalent to

$$\begin{aligned} \min_{\mathbf{s}} \quad & \mathbf{s}^T \mathbf{B} \mathbf{s} \\ \text{s.t.} \quad & s_i s_n = s_i^2 \quad \forall i, \\ & s_n^2 = 1, \\ & s_n \cdot \mathbf{1}_M^T \cdot [s_1, \dots, s_M]^T = K, \\ & s_n \cdot \mathbf{w}^T \cdot [s_1, \dots, s_M]^T + s_n s_{M+1} = W, \end{aligned} \quad (3)$$

with

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

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

$$\begin{aligned} \hat{\mathbf{S}} = \operatorname{argmin}_{\mathbf{S} \succeq \mathbf{0}} \quad & \operatorname{trace}(\mathbf{B} \mathbf{S}) \\ \text{s.t.} \quad & S_{ii} = S_{in} \forall i, S_{nn} = 1, \sum_{i=1}^M S_{ni} = K, \\ & \sum_{i=1}^M w_i S_{ni} + S_{n,M+1}, \operatorname{rank}(\mathbf{S}) = 1. \end{aligned} \quad (4)$$

The semidefinite programming (SDP) relaxation of (4) can efficiently be solved in polynomial time [21]. If the resulting matrix $\hat{\mathbf{S}}$ has rank one, then the relaxation is tight. Otherwise, special techniques are required to convert the SDP relaxation solution back to the solution of an binary quadratic problem, see, e.g., [22, 23]. Using any of these relaxation techniques, we then obtain the set of estimators $\{\tilde{\mathbf{s}}\}$. Before remapping it back to vector $\mathbf{b}$, it has to be checked whether the estimator fulfills the average power constraint and find the one which maximizes the channel capacity among all those. Originally, $\tilde{\mathbf{s}}^{(i)}$ 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 is already added before. The resulting vector $\tilde{\mathbf{s}}$ leads to the approximate solution of the original binary vector $\mathbf{b}$. This successive subset selection algorithm is summarized in Algorithm 1. Note that the metric to choose $\tilde{\mathbf{s}}$ in Algorithm 1 differs from [20]. It is a traditional way to use

$$\tilde{\mathbf{s}} = \operatorname{argmin}_{\tilde{\mathbf{s}}^{(i)}} \tilde{\mathbf{s}}^{(i)T} \mathbf{B} \tilde{\mathbf{s}}^{(i)}$$

which directly maximizes the cutoff rate as expected and where $\tilde{\mathbf{s}}^{(i)}$ denotes a scale for a certain input probability. However, there exists a gap between cutoff rate and channel capacity. To guarantee the best result among the randomizations, we directly calculate the capacity for each candidate of $\tilde{\mathbf{s}}$ in Step 15. Further, Step 15 helps to reduce the computation load since sometimes different randomization loops result the same or equivalent selection due to a symmetric noise distribution. We do not need to include any selection which already occurs.

3.2 Heuristic improvement based on weights of signaling points

The solution can further be improved by applying the following heuristic method. For readability, we mention this sub-algorithm separately from Algorithm 1. Consider the vector $\tilde{\mathbf{s}}^{(i)}$ in Step 10 of Algorithm 1 which is not yet quantized and sort the first M entries in descending order. In Algorithm 1, the highest K entries are directly chosen. It is easy to see that this selection is not optimal in general. 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 with respect to our objective of maximizing the cutoff rate. As exhaustive search is too complex to apply for our scenarios, we use a simple heuristic method to reduce the computational complexity of an optimal exhaustive search. In the method, the unquantized first M entries of $\tilde{\mathbf{s}}^{(i)}$ are kept in the buffer, adding $\tilde{\mathbf{s}}_{un}^{(i)} = [\tilde{s}_1^{(i)}, \dots, \tilde{s}_M^{(i)}]$ between Step 10 and 11. In Step 18, we obtain $\tilde{\mathbf{s}}$ and the paired $\tilde{\mathbf{s}}_{un} = \tilde{\mathbf{s}}_{un}^{(i)}$, 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 $K_1, K_2 \in \mathbb{R}$ 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, mutual information maximizing K_1 entries to obtain the new subset. Figure 2 illustrates how this method works. Simulations show that the solution of Algorithm 1 improves by applying this heuristic method even for rather small K_1 and K_2 such that the computational complexity of this heuristic method is low. Note that the actual choice of K_1 and K_2 is a trade-off between complexity and performance. We denote the subset selection algorithm (Algorithm 1) endowed with the above heuristic

Algorithm 1 Subset selection algorithm**1: Initialization:**

$$n = M + 2,$$

$$A = \int \begin{bmatrix} \sqrt{f_1(\mathbf{y})} \\ \vdots \\ \sqrt{f_M(\mathbf{y})} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{f_1(\mathbf{y})} \\ \vdots \\ \sqrt{f_M(\mathbf{y})} \end{bmatrix}^T d\mathbf{y},$$

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

define an empty vector $\mathbf{r}$;

2: Solve the semi-definite problem:

$$\hat{S} = \operatorname{argmin}_{S \succeq \mathbf{0}} \operatorname{trace}(BS) \quad \text{s.t.}$$

$$S_{ii} = S_{in}, \quad \forall i, \quad S_{nn} = 1, \quad \sum_{i=1}^M S_{ni} = K, \quad \sum_{i=1}^M w_i S_{ni} + S_{n,M+1} = W.$$

3: Adjustment:

$$n = M + 1,$$

delete the $(M + 1)$ th row and $(M + 1)$ th column of $\hat{S}$,

$$B = \begin{bmatrix} A & \mathbf{0}_{1 \times M} \\ \mathbf{0}_{M \times 1} & 0 \end{bmatrix}.$$

4: Cholesky factorization:

$$\hat{S} = \hat{V}^T \hat{V}.$$

5: Randomization:**6: for** $i = 1, \dots, N_{rand}$ **do**

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

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

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

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

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

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

13: **else continue**

14: **end if**

15: **if** t is not yet in vector $\mathbf{r}$ **then**

$$r_i = t,$$

calculate channel capacity C_i based on $\tilde{\mathbf{s}}^{(i)}$.

16: **end if**

17: **end for**

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

19: **Take** $\mathbf{b} = [\tilde{s}_1, \dots, \tilde{s}_M]^T$ as approximate solution.

improvement by SSA. For the sake of clarity, Table 1 briefly summarizes the algorithms introduced in this work.

4 Selecting both signaling constellation and input distribution

In the previous section, we investigated the task of finding an optimum signaling constellation in a bounded set by considering a uniform distribution on the selected subset. Our numerical evaluations in Sect. 5 show that the corresponding

mutual information is quite close to the capacity of the channel using the full signaling constellation. However, the uniform distribution we used for the obtained set is, of course, suboptimal and we would like to approach the question of optimizing both the set of signaling points and the corresponding input probability. We investigate this task by two different procedures, which are compared in Sect. 5. First, the subset is chosen according to Sect. 3 and the input distribution is improved using an idea which is based on the Blahut Arimoto algorithm. Second, we improve the selected subsets and probabilities simultaneously.

Fig. 2 Heuristic method to improve the subset selection

4.1 Successive subset and distribution selection

The approach described in Sect. 3 gives a subset of signaling points which is capacity-achieving in the set of uniform distributions with at most K support points. An open question is, if the chosen subset can be utilized in a better way by using a non-uniform distribution. This question is answered in the following.

The idea is based on the classical Blahut-Arimoto algorithm which computes in an elegant way the capacity of a discrete memoryless channel [12, 13]. We extend the algorithm to the here considered case of discrete input and continuous output. Extending Theorem 1 in [12] to our system model, we obtain the following proposition.

Proposition 2 *With*

$$I_0(\mathbf{p}, \mathcal{P}, \mathbf{f}) = \sum_{i=1}^M \int p_i f_i \log \frac{P(\mathbf{x}_i|\mathbf{y})}{p_i} d\mathbf{y}$$

the following holds.

1. *The channel capacity is given by*

$$C = \max_{\mathbf{p}} I(\mathbf{p}, \mathbf{f}) = \max_{\mathbf{p}} \max_{\mathcal{P}} I_0(\mathbf{p}, \mathcal{P}, \mathbf{f}).$$

2. *For fixed $\mathbf{p}$, $I_0(\mathbf{p}, \mathcal{P}, \mathbf{f})$ is maximized by*

$$P(\mathbf{x}_i|\mathbf{y}) = \frac{p_i f_i}{\sum_{j=1}^M p_j f_j}. \quad (5)$$

3. *For fixed $\mathcal{P}$, $I_0(\mathbf{p}, \mathcal{P}, \mathbf{f})$ is maximized by*

$$p_i = \frac{\exp(\int f_i \log P(\mathbf{x}_i|\mathbf{y}) d\mathbf{y})}{\sum_{j=1}^M \exp(\int f_j \log P(\mathbf{x}_j|\mathbf{y}) d\mathbf{y})}. \quad (6)$$

Inserting (5) into (6) gives

$$\begin{aligned} p_i &= \frac{p_i \exp\left(\int f_i \log \frac{f_i}{\sum_{j=1}^M p_j f_j} d\mathbf{y}\right)}{\sum_{j=1}^M p_j \exp\left(\int f_j \log \frac{f_j}{\sum_{k=1}^M p_k f_k} d\mathbf{y}\right)} \\ &= \frac{p_i \exp(D(f_i||f_0))}{\sum_{j=1}^M p_j \exp(D(f_j||f_0))}, \end{aligned} \quad (7)$$

where $f_0 = \sum_{i=1}^M p_i f_i$. This yields the idea for the updating algorithm. The probabilities $\mathbf{p}^{(k)}$ of iteration $k \geq 1$ are obtained by

$$p_i^{(k)} = \frac{1}{N^{(k)}} p_i^{(k-1)} \exp(D(f_i||f_0^{(k-1)})), \quad i = 1, \dots, M,$$

where $N^{(k)}$ is the normalizing factor

$$N^{(k)} = \sum_{j=1}^M p_j^{(k-1)} \exp(D(f_j||f_0^{(k-1)})),$$

and

$$f_0^{(k)} = \sum_{j=1}^M p_j^{(k)} f_j.$$

Note that $f_0^{(k)}$ is updated not because of f_i but $p_i^{(k)}$. Starting from a distribution $\mathbf{p}^{(1)}$ with all components strictly positive, this algorithm converges, see [12]. Let $\hat{\mathbf{p}}$ denote the limit. From (7) it follows that

$$\begin{aligned} \exp(D(f_i||f_0)) &= \sum_{j=1}^M \hat{p}_j \exp(D(f_j||f_0)), \\ \text{for all } i &= 1, \dots, M, \end{aligned}$$

which implies

$$D\left(f_i \parallel \sum_{j=1}^M \hat{p}_j f_j\right) = \text{const.}$$

Thus, by Proposition 1, which is a necessary and sufficient optimality criterion for optimality, $\hat{\mathbf{p}}$ is capacity-achieving.

As the standard Blahut-Arimoto algorithm is computationally quite complex, several enhancements have been proposed. The most important ones, the natural-gradient-based algorithm and the accelerated Blahut-Arimoto algorithm are mentioned in [14]. They converge significantly faster to the capacity-achieving distribution and can be extended to our system model.

Table 1 Overview of algorithm names and abbreviations

Name	Abbreviation	Explanation
Subset selection algorithm with heuristic improvement	SSA	Chooses a signaling set of size K with respect to a uniform distribution
SSA plus iterated Blahut-Arimoto algorithm	SSA_IBA	Chooses a subset first and then improves the distribution
Truncation of iterated Blahut-Arimoto algorithm	TIBA	Simultaneous subset and distribution selection

The accelerated Blahut-Arimoto algorithm updates $\mathbf{p}^{(k)}, k \geq 1$, according to

$$p_i^{(k)} = \frac{1}{N^{(k)}} p_i^{(k-1)} \exp\left(\mu^{(k)} D(f_i \| f_0^{(k-1)})\right), \quad i = 1, \dots, M,$$

where $\mu^{(k)}$ denotes an accelerating step size. As suggested in [14], the factor $\mu^{(k)}$ is adjusted by

$$\mu^{(k+1)} = \frac{D(\mathbf{p}^{(k)} \| \mathbf{p}^{(k-1)})}{D(f_0^{(k)} \| f_0^{(k-1)})} = \frac{\sum_{j=1}^M p_j^{(k)} \log(p_j^{(k)} / p_j^{(k-1)})}{\sum_{j=1}^M p_j^{(k)} f_j \log \frac{\sum_{j=1}^M p_j^{(k)} f_j}{\sum_{j=1}^M p_j^{(k-1)} f_j}}$$

for $k > 1$ with initial value $\mu^{(1)} = 1$.

The natural-gradient-based algorithm uses the update rule

$$p_i^{(k)} = p_i^{(k-1)} \left[1 + \mu^{(k)} \left(D(f_0^{(k)} \| f_0^{(k-1)}) - I_L^{(k-1)} \right) \right],$$

for all $k \geq 1$,

where $\mu^{(k)}$ is the same as in the accelerated Blahut-Arimoto algorithm.

The accelerated Blahut-Arimoto algorithm converges at least linear for fixed step size μ and convergence speed increases by increasing μ . This holds as

$$\lim_{k \rightarrow \infty} \frac{\|\mathbf{p}^{k+1} - \mathbf{p}^*\|}{\|\mathbf{p}^k - \mathbf{p}^*\|} \leq \frac{c}{\mu}$$

with constant c and $\mathbf{p}^*$ denotes the optimum distribution, as was shown in [14]. Furthermore, it can be shown that the accelerated Blahut-Arimoto algorithm converges superlinearly for properly chosen step sizes. However, obtaining the appropriate step sizes is computationally very complex [14]. Explicit results are difficult to obtain for the natural-gradient-based algorithm. Reference [14] suggests however, that convergence properties are similar to the latter ones. The convergence speed of the ordinary Blahut-Arimoto algorithms corresponds to $\mu = 1$ and thus, the proposed heuristics clearly converge faster.

The performance of the above algorithms can be improved by applying a heuristic method related to the one introduced in Sect. 3.2 and further performing Blahut-Arimoto for any combined new subset. This improves the performance, since a better subset for a uniform distribution does not always imply to be a better one for the non-uniform case. In addition, the parameters K_1 and K_2 should be chosen rather small to allow for additional Blahut-Arimoto applications in each step.

References [15–17] suggest another interesting enhancement, the iterated Blahut-Arimoto algorithm. Though it was proposed for discrete memoryless channels, it can be extended to a discrete input continuous output channel. We abbreviate the successive subset selection obtained by our subset selection algorithm followed by the

iterated Blahut-Arimoto algorithm by SSA_IBA, see also Table 1. This algorithm is motivated by the fact that capacity-achieving distributions usually include only a small number of inputs that are assigned non-zero probabilities, especially when the input alphabet size is very large. Thus, great efforts can be made in eliminating inputs which will end up with probability zero. Thereby the algorithm operates on a subset of the whole input alphabet. The algorithm starts from an input alphabet that consists of only two symbols. The alphabet grows by one symbol at every iteration until it includes all symbols with non-zero probabilities. At every iteration, a Blahut-Arimoto algorithm is used to compute a capacity relative to a partial input alphabet. This approach is discussed in more detail in the following subsection.

The ordinary Blahut-Arimoto algorithm is slow when the input alphabet is large. The IBA utilizes the Blahut-Arimoto algorithm only for small sets of signaling points and thus converges faster.

4.2 Simultaneous subset and distribution selection

As shown in Fig. 3 for the scenario given in Section 5, the channel capacity achieved by the iterated Blahut-Arimoto algorithm increases fast at first, when the partial alphabet size is small. Hence, by truncating the algorithm, i.e., by terminating it once the alphabet size reaches the subset size requirement, we obtain an approximate solution of the non-uniform distribution over the subset. Let $C_{\mathcal{X}'}$ denote the capacity of the discrete input continuous output channel induced when all but the symbols in the subset $\mathcal{X}'$ of $\mathcal{X}$ are assigned a probability of zero. Our truncation of the iterated Blahut-Arimoto algorithm (TIBA) with respect to our

Fig. 3 Increase in mutual information with the iterated Blahut-Arimoto algorithm

system model is based on the following steps which are also summarized in Algorithm 2:

1. Determine $\{i, j\} \in \mathcal{X}^2$ such that $C_{\{i, j\}} = D(f_i || f_0) = D(f_j || f_0)$ is maximized over all choices of i and j where $f_0 = p_i f_i + p_j f_j$ in this case. Define $\mathcal{X}' = \{i, j\}$ and $C' = C_{\mathcal{X}'}$.
2. If $\mathcal{X}' = \mathcal{X}$, then $C = C'$ and the algorithm terminates. Otherwise, for all $m \in \mathcal{X} \setminus \mathcal{X}'$, compute $D(f_m || f_0)$. If the values computed are all smaller or equal to C' , then $C = C'$ and the algorithm can be terminated at this point.
3. Add the symbol m that maximized $D(f_m || f_0)$ in Step 2 to the set $\mathcal{X}'$. Recompute $C' = C_{\mathcal{X}'}$ using the Blahut-Arimoto algorithm and update f_0 with the new $\mathcal{X}'$. Return to step 2.

The function BA mentioned in the algorithms is a placeholder for the iterated or non-iterated Blahut-Arimoto algorithm with either the accelerated Blahut-Arimoto algorithm or natural-gradient-based algorithm. The actual choice depends on the size of K .

To further enhance the performance of the truncation, we exploit the following ideas. Zero probability inputs characterized above should be eliminated to allow other signaling points to be included. To exclude zero probability inputs more efficiently, we set a certain threshold and exclude the input if its probability $p_i < \max\{10^{-3}, \frac{1}{k^2}\}$, where k is the current subset size, i.e., the number of already included points. The choice for a threshold for p_i is a balance between resolution and speed. We found in examples $\max\{10^{-3}, \frac{1}{k^2}\}$ to be quite good. The cardinality of $\mathcal{X}'$ increases by one during each iteration. Nevertheless, excluding low probability points as mentioned above, the cardinality of $\mathcal{X}'$ may shrink while improving the mutual information. Therefore, we postpone the truncation of the iterated algorithm until the cardinality of $\mathcal{X}'$ is equal to $K + K_1$ for some $K_1 > 0$. Instead of directly choosing $\mathcal{X}'$ with size K , we obtain it by reducing it from a certain larger set of size $K + K_2$ by excluding the least probable K_2 points. The explained steps are summarized in Algorithm 3.

The initialization of two input symbols can be done as outlined in Algorithm 4. It describes an efficient way to apply exhaustive search to all two point combinations.

Algorithm 2 Truncation of Iterated Blahut-Arimoto Algorithm (TIBA)

```

1: Initialization:
    $\mathcal{X}' = \{i, j\} = \operatorname{argmax}_{\{i, j\} \in \mathcal{X}^2} C_{\{i, j\}}, \quad C' = C_{\{i, j\}}, \quad f_0 = p_i f_i + p_j f_j.$ 
2: while true do
3:   if  $\operatorname{size}(\mathcal{X}') == M$  then
4:     break
5:   else if  $\operatorname{size}(\mathcal{X}') < K + K_1$  then
6:     for  $m = 1, \dots, M$  do
7:       if  $m \notin \mathcal{X}'$  then
8:          $D(f_m || f_0) = \int f_m \log \frac{f_m}{f_0} dy$ 
9:       end if
10:    end for
11:     $[D_{\max}, d_{\max}] = \max_{m=1, \dots, M} D(f_m || f_0)$ 
12:    if  $D_{\max} < C'$  then
13:      break
14:    else
15:       $\mathcal{X}' = \{\mathcal{X}', d_{\max}\}$ 
16:       $[p', C'] = BA(\mathcal{X}')$ 
17:      Run Algorithm 3.
18:       $f_0 = \sum_{m \in \mathcal{X}'} p_m f_m.$ 
19:    end if
20:  else
21:    break
22:  end if
23: end while
24: Take  $[\mathcal{X}', p']$  as approximate solution and  $C = C'$ .
```

Algorithm 3 Sub Algorithm for TIBA

```

1:  $k = \text{size}(\mathcal{X}')$ 
2: if  $p_{d_{\max}} > \max\{10^{-3}, \frac{1}{k^2}\}$  then
3:    $\mathcal{X}' = \{i | i \in \mathcal{X}', p_i > \max\{10^{-3}, \frac{1}{k^2}\}\}$ 
4:    $[\mathbf{p}', C'] = \text{BA}(\mathcal{X}')$ 
5: end if
6: if  $\text{size}(\mathcal{X}') == K + K_2$  then
7:    $\mathcal{X}' = \{i | i \in \mathcal{X}', p_i \text{ is among the top } K \text{ of } \mathbf{p}'\}$ 
8:    $[\mathbf{p}', C'] = \text{BA}(\mathcal{X}')$ 
9: end if

```

5 Simulation results

For our simulations, we use the following scenario. We aim to choose $K = 16$ signaling points from an M -QAM scenario with $M = 64$. The 64-QAM points in the square $[-3, 3]^2$ are chosen as initial situation. We consider 2-dimensional Gaussian noise with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with varying parameter ρ .

5.1 Reducing the number of signaling points

As outlined above, using a uniform distribution over the selected subset provides a lower bound for the capacity-achieving non-uniform distribution.

5.1.1 The subset selection algorithm with heuristic improvement: SSA

SSA depends on the chosen SDP relaxation technique for solving (4). Figure 4 compares different SDP relaxation techniques. The blue curve depicts the channel capacity obtained by applying the dominant eigenvector approximation, see [23], to acquire the SDP relaxation. The green and red curve are obtained by using randomization with 30 loops, i.e., an estimator of s is obtained after each randomization loop. In examples, we observed that 30 loops are close to maximize the mutual information. The red one aims at maximizing the cutoff rate, while the green one directly maximizes the mutual information. As can be seen in the figure and as expected, maximizing the mutual information with randomization always performs best, though the difference is interestingly not very large especially for small and large SNR values.

The resulting signaling schemes found by SSA depend on the total power constraint. Figure 5 shows the mutual information 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, be observed from the plot that each curve converges to its maximum, which is the same as the channel capacity without total power constraint.

We find in examples that the performance of a channel that only exploits a small set of signaling points and a uniform distribution is remarkably good. For low SNR values, the achieved mutual information is quite close to the capacity of the full set of signaling points. This is depicted in Fig. 6. Only for large SNR values, the capacity gap is rather large. The capacity for the optimal case (light blue) was obtained by using the Blahut-Arimoto algorithm. The blue, green and red curve show the mutual information result for a subset of size 16, 32 and 64, respectively, obtained by SSA. Interestingly, for small SNR values, a smaller subset size achieves a higher capacity than a larger subset size. Note that we only consider uniform distributions so far. For the full problem it is trivial, that a larger subset size achieves higher capacity. However, as can be seen here, this does not hold for the uniform case.

5.2 Selecting both signaling points and probability measure

So far, we only considered a uniform distribution over the signaling points. The channel can be exploited even better by choosing a more appropriate probability distribution. In Sect. 4, we discussed two algorithms for finding a better probability distribution.

5.2.1 The successive subset and distribution selection: SSA_IBA

In Fig. 7 we investigated the constellation of the capacity achieving distribution for the full set of signaling points using SSA_IBA for $K = M = 64$. The red points depict the optimal signaling points, the numbers refer to the optimal probability multiplied by 1,000. The tiny blue dots are the signaling points that are assigned a probability of zero by the algorithm, i.e., they are not chosen as signaling point. Interestingly, the obtained distribution only uses 28 signaling points. This shows that by restricting ourselves to a smaller set of signaling points, we can still achieve a remarkably good mutual information. Figure 9 shows the improvement of the heuristic method introduced in Sect. 4.1. The parameters were set to $K_1 = 8$ and $K_2 = 2$.

5.2.2 Simultaneous subset and distribution selection: TIBA

To show the performance of TIBA, we consider the following two plots. Figure 10 shows the performance of TIBA compared to the channel capacity. As can be seen, TIBA performs quite good. The arising question is, if

Algorithm 4 Initialization of Iterated Blahut-Arimoto Algorithm

```

1:  $C' = 0, \epsilon = 10^{-4}$ 
2: for  $i = 1, \dots, M - 1$  do
3:   for  $j = i + 1, \dots, M$  do
4:      $l = 0, u = 1$ 
5:      $\lambda = \frac{1}{2}(l + u), f_0 = \lambda f_i + (1 - \lambda)f_j$ 
6:     if  $D(f_i||f_0) > D(f_j||f_0)$  then
7:        $l = \lambda$ 
8:     else
9:        $u = \lambda$ 
10:    end if
11:    if  $|D(f_i||f_0) - D(f_j||f_0)| < \epsilon$  then
12:      if  $\min(D(f_i||f_0), D(f_j||f_0)) > C'$  then
13:         $C' = \min(D(f_i||f_0), D(f_j||f_0))$ 
14:         $\mathcal{X}' = \{i, j\}$ 
15:      end if
16:      Continue
17:    else
18:      Go to Step 5
19:    end if
20:  end for
21: end for
22:  $\mathcal{X}'$  is the best subset of size 2 and  $C'$  the corresponding capacity.

```

Fig. 4 SDP relaxation (Step 1 in Algorithm 1) comparison by changing SNR. The dominant eigenvalue approximation (*blue*) vs. randomization with 30 loops aiming at maximizing the cutoff rate (*green*) and the capacity (*red*)

TIBA finds the best possible subset and probability distribution or if exhaustive search can yield a result which is even closer to the full set capacity. As exhaustive search is computationally too complex, we include three simple aspects in the following pseudo exhaustive search, thereby reducing run-time enormously. As only 28 signaling points are assigned a non-zero probability in the optimum

Fig. 5 Mutual information vs. the total power constraint for different correlation parameters ρ

probability measure, see Fig. 7, we consider only those and reduce ourselves to choosing 16 out of 28 rather than 16 out of 64. Moreover, we include the four vertices $(3, 3)'$, $(-3, 3)'$, $(-3, -3)'$, $(3, -3)'$ to our choice, as they are assigned the highest probabilities in the optimum distribution. Thus, the search further reduces to 12 out of 24 with the four vertices fixed. We also included symmetry aspects in the pseudo exhaustive search which excludes many repeated combinations. The extended Blahut-

Fig. 6 Comparison of mutual information for different numbers of signaling points

Fig. 7 The optimum input distribution

Arimoto algorithms converges really fast in the first several iterations. That implies, that we can discard a combination if the capacity achieved after several iterations is still too low compared to other combinations. Thereby we do not waste any time in finding the exact capacities of not interesting combination. Taking all these considerations into account, we obtain our pseudo exhaustive search, which, for the scenario given, yields the results shown on in Fig. 11. Running TIBA gives exactly the same points and probabilities, i.e., our algorithm yields the best possible result.

Fig. 8 Mutual information comparison for different subset selection algorithms

Fig. 9 Comparison of mutual information for different numbers of signaling points

5.2.3 Comparison of SSA, SSA_IBA and TIBA

As can be seen in Fig. 8, TIBA performs best while the uniform distribution over the subset always gives the smallest mutual information. The intuitive and sequential approach of first finding a subset and then improving the according probability still yields a quite reasonable high mutual information, it is quite close to the sophisticated TIBA.

6 Conclusions

In this paper, we investigated the challenging problem of finding a reduced, optimal set of signaling points and a

Fig. 10 Comparison of TIBA with the channel capacity

Fig. 11 Comparison of TIBA with pseudo exhaustive search

corresponding capacity achieving probability measure. By assuming first a uniform distribution on the selected signaling points, we obtained a lower bound for the original problem. We considered both a mutual information and cutoff rate maximizing subset selection and solved these problems by forming a semidefinite programming problem and applying two different relaxation techniques. A heuristic improvement based on weights of signaling points improves the performance of this lower bound, which is close to the channel capacity of the full set of signaling points. Two different ways to tackle the full problem of selecting a subset and a distribution were introduced. Building on a Blahut Arimoto algorithm and a subsequent

heuristic improvement, the uniform distribution obtained in the first step is replaced by a more sophisticated one. Using a simultaneous update on both the chosen signaling points and the probability measure, we obtained the so called TIBA algorithm, which performs best under the proposed methods and gives remarkable results compared to exhaustive search.

Our numerical results show that using only a subset of small size can indeed achieve a very high mutual information even compared to the large full input set. This approach helps to highly simplify the receiver design while maintaining a high transmission rate over the channel.

An interesting open problem for future research is the analysis of the choice of K compared to M , e.g., the ratio K/M and analytical results proving our numerical conclusions.

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

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Telatar, I. E. (1999). Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6), 585–595.
2. Chan, T. H., Hranilovic, S., & Kschischang, F. R. (2005). Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE Transactions on Information Theory*, 51(6), 2073–2088.
3. Smith, J. G. (1971). The information capacity of amplitude and variance-constrained scalar Gaussian channels. *Information and Control*, 18(6), 203–219.
4. Shamai, S., & Bar-David, I. (1995). The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Transactions on Information Theory*, 41(4), 1060–1071.
5. Chan, T. H., & Kschischang, F. R. (2004). On the discreteness of the capacity achieving probability measure of conditional gaussian channels. In *Proceedings ISIT*, Chicago, June, p. 347.
6. Katz, M., & Shamai, S. (2004). On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels. *IEEE Transactions on Information Theory*, 50(10), 2257–2270.
7. Abou-Faycal, I. C., Trott, M. D., & Shamai, S. (2001). The capacity of discrete-time memory-less rayleigh fading channels. *IEEE Transactions on Information Theory*, 47(4), 1290–1301.
8. Gursay, M. C., Poor, H. V., & Verdu, S. (2005). The noncoherent rician fading channel—part I: structure of the capacity-achieving input. *IEEE Transactions of Wireless Communications*, 4(5), 2193–2206.
9. Sharma, S., & Shamai, S. (2008). 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. Foschini, G. J., Gitlin, R. D., & Weinstein, S. B. (1974). Optimization of two-dimensional signal constellations in the presence of Gaussian noise. *IEEE Transactions on Communications*, 22(1), 28–38.
11. Schmeink, A., Mathar, R., & Zhang, H. (2010). Reducing the number of signaling points keeping capacity and cutoff rate high. In *The seventh international symposium on wireless communication systems (ISWCS)*, 2010, submitted.
12. Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4), 460–473.
13. Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1), 14–20.
14. Matz, G., & Duhamel, P. (2004). Information geometric formulation and interpretation of accelerated blahut-arimoto-type algorithms. In *Information theory workshop, 2004. IEEE*, Oct. 2004, pp. 66–70.
15. Gallager, R. G. (1968). *Information theory and reliable communications*. New York: Wiley.
16. Sayir, J. (2000). Iterating the arimoto-blahut algorithm for faster convergence. In *Information Theory, 2000. Proceedings of IEEE International Symposium on*, 2000, p. 235.
17. Sayir, J. (1999). *On coding by probability transformation*. ETH Zürich, PhD Dissert. Nr. 13099, Hartung-Gorre Verlag Konstanz, Germany.
18. Mathar, R., Schmeink, A., & Zivkovic, M. (2008). Optimum discrete signaling over channels with arbitrary noise distribution. In *2nd International conference on signal processing and communication systems (ICSPCS)*, Gold Coast, Australia, Dec.
19. Proakis, J. G. (2001). *Digital communications*, 4th ed. McGraw-Hill, August.
20. Mezghani, A., Ivrlac, M., & Nossek, J. (2008). 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.
21. Sturm, J. F. (1999). Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones.
22. Wiesel, A., Eldar, Y., & Shitz, S. (2005). Semidefinite relaxation for detection of 16-qam signaling in mimo channels. *IEEE Signal Processing Letters*, 12(9), 653–656.
23. Ma, W.-K., Ching, P.-C., & Ding, Z. (2004). Semidefinite relaxation based multiuser detection for m-ary psk multiuser systems. *IEEE Transactions on Signal Processing*, 52(10), 2862–2872.

Author Biographies

Anke Schmeink received her Diploma degree in mathematics and the PhD degree in electrical engineering and information technology from RWTH Aachen University, Germany, in 2002 and 2006, respectively. From September to December 2005, she was a Visiting Researcher with the Department of Electrical and Electronic Engineering, The University of Melbourne. From November 2006 till May 2008 she was a research scientist with the medical

signal processing group at Philips Research Laboratories. Anke Schmeink joined the faculty of electrical engineering and information technology at RWTH Aachen University as Assistant Professor in May 2008. She is a member of the Young Academy at the North Rhine-Westphalia Academy of Science. Anke Schmeink received the E-plus best dissertation award 2006, the Vodafone young scientist award 2007 and the Helene Lange young scientist award 2009. Her research interests are in information theory and systematic design of communication systems, particularly game theory, optimization and medical applications.

Haoqing Zhang received his Bachelor's degree in Information Engineering in 2007 from Southeast University, Nanjing, China, with the President Scholarship award in 2005. Within this period, he took a one-year exchange study in 2006 at the University of Ulm, Germany. Then he started his Master study in Communications Engineering at RWTH Aachen University, Germany, as a recipient of Siemens Master Scholarship. He finished his

Master thesis in 2009 at UMIC Research Centre in the framework of German government excellence initiative. Currently he is doing an internship at Infineon Technology in Duisburg, Germany.