STORM: Optimal Constellations for Noncoherent MIMO Communications at Low SNR under PAPR Constraints

Shivratna Giri Srinivasan and Mahesh K. Varanasi

***Abstract*— Reliable communication over the discrete input and continuous output noncoherent MIMO Rayleigh fading channel is considered when the SNR per degree of freedom is low. The input constellations are required to satisfy peak and average power constraints. When the peak-to-average power is constrained and in the low SNR regime, the mutual information upto second order in SNR is maximized jointly over the input signal matrices and their respective probabilities. Even though the problem considered is a finite dimensional *non-convex* optimization, it admits an elegant solution in *closed form*. The constellation obtained is referred to as Space Time Orthogonal Rank one Modulation (STORM), and it provides several new insights into noncoherent MIMO comunications in the low SNR regime. In contrast to most existing schemes for the noncoherent MIMO channel at low SNR, STORM is a practical constellation due to its discrete structure with a finite number of points and because it requires a moderate PAPR in general. Moreover, STORM is near-optimal with respect to the maximum mutual information with unconstrained cardinality, even with a moderate peak-to-average power ratio (PAPR). Numerical results show that STORM is also close to optimal from the important viewpoint of spectral efficiency.**

I. Introduction

In this paper, we consider the problem of communicating reliably over the noncoherent MIMO i.i.d. Rayleigh fading channel in the low SNR regime. This regime is commonly encountered in wideband (WB) and ultra-wideband (UWB) channels, where the signal power is spread over a large bandwidth, consequently making the SNR per degree of freedom low. A well known characteristic of the optimal signals in the low SNR regime is that they have a high peak-to-average power ratio (PAPR) in general. It is shown in [1] that under certain regularity conditions on the signal, which include making the fourth and sixth moments finite, the noncoherent MIMO capacity grows as $O(\mathrm{SNR}^2)$. This is the capacity that can be expected when the average power is low and the peak-to-average power ratio (PAPR) is not too large, which is the case with many practical modulation schemes. Similar expressions for the mutual information upto the second order is obtained in closed form in [2], [3] with different assumptions on the fading matrices and peak-power constraints. In this paper, we maximize this mutual information with respect to discrete input constellation matrices and their respective probabilities. To ensure that the optimal constellations are practically meaningful, we impose a peak-power constraint per antenna and per time slot in our optimizations, in addition to the average power constraint. In this paper, we consider the practically motivated case when the peak-power constraint is a constant times the average-power constraint, so that the PAPR of the optimal constellation is limited. This assumption ensures that the input constellation satisfies the regularity conditions in [1], and the mutual information grows as $O(\mathrm{SNR}^2)$.

The optimization of the mutual information at low SNR jointly over the signal matrices as well as their respective probabilities is a finite dimensional *non-convex* optimization problem. Inspite of this, we show that the problem admits an elegant solution in *closed* form due to its structure. We refer to this new constellation as **S**pace **T**ime **O**rthogonal **R**ank one **M**odulation (STORM). The key advantage of formulating the problem of maximizing mutual information as a finite dimensional optimization over constellation matrices and their probabilities is that the solution obtained offers insights simultaneously into information theoretic as well as coding-modulation aspects. A major benefit of STORM is that it has a mutual information very close to optimal while having a moderate value of PAPR. Unlike many existing noncoherent schemes for low SNR, STORM is practical because it is discrete, consists of a finite number of points, and has moderate PAPRs in general.

II. System Model

We consider a MIMO channel with N_t transmit and N_r receive antennas. The random channel matrix $\mathbf{H} \in \mathbb{C}^{N_t \times N_r}$ is assumed to be constant for a duration of T symbols after which it changes to an independent value. It has independent, identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries and the knowledge that the entries have this distribution is available to the transmitter and receiver. The realizations of $\mathbf{H}$ however, are unknown at both ends. Assuming that the transmitted symbol is $\mathbf{X} \in \mathbb{C}^{T \times N_t}$, the output of the channel can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{H} + \mathbf{N} \; . \tag{1}$$

The entries of $\mathbf{N}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables. The symbol $\mathbf{X}$ is drawn from a finite constellation $\boldsymbol{\mathcal{C}}$ which in turn is normalized so that $\frac{1}{T}\mathrm{E}[\mathrm{tr}(\mathbf{X}\mathbf{X}^*)] \leq P$ so that the average received SNR is constrained to be at most P. For convenience, we denote the average energy per block of T symbols by $E = PT$. We also impose an additional peak power constraint $\|\mathbf{X}\|_\infty = \max_{i,j} \; |[\mathbf{X}]_{ij}| \leq \sqrt{K}$ for each $\mathbf{X} \in \boldsymbol{\mathcal{C}}$. This peak-power constraint is most natural as it restricts the peak-power per antenna and per time slot to at most K.

This work was supported in part by NSF grants CCF-0434410 and CCF-0431170.

Srinivasan and Varanasi are with the Electrical & Computer Engineering Department at the University of Colorado, Boulder, CO 80309 {srinivsg,varanasi}@colorado.edu

We define the PAPR of a constellation $\mathcal{C}$ as

$$\eta = \max_{\substack{\mathbf{X}\in\mathcal{C}\\ m,n}} \frac{|[\mathbf{X}]_{mn}|^2}{\mathrm{E}\left\{|[\mathbf{X}]_{mn}|^2\right\}} . \quad (2)$$

The probability density function (p.d.f.) of $\mathbf{Y}$ conditioned on $\mathbf{X}$ being sent is given by

$$p(\mathbf{Y}|\mathbf{X}) = \frac{\exp\left\{-\mathrm{tr}\left(\mathbf{Y}^*\left(\mathbf{I}_T+\mathbf{X}\mathbf{X}^*\right)^{-1}\mathbf{Y}\right)\right\}}{\pi^{TN_r}\left|\mathbf{I}_T+\ \mathbf{X}\mathbf{X}^*\right|^{N_r}}$$

III. Maximizing the mutual information at low SNR

Throughout this paper, we will consider a discrete input (of cardinality L) and continuous output channel over which the constellation $\{\mathbf{X}_i\}_{i=1}^L$ with corresponding transmission probabilities $\{P_i\}_{i=1}^L$ is used.

The mutual information for the noncoherent MIMO Rayleigh fading channel under an average power constraint on the input signals is not known in closed form for a general SNR. At asymptotically low SNR and when the input signal satisfies certain regularity conditions to avoid inputs having prohibitively large PAPRs, the authors in [1] show that the mutual information is zero upto first order for the continuous input and continuous output channel. Moreover, the mutual information upto the second order in P is also obtained in closed form as follows.

$$I_{low} = \frac{N_r}{2T}\mathrm{tr}\left\{\mathrm{E}[(\mathbf{X}\mathbf{X}^*)^2] - (\mathrm{E}[\mathbf{X}\mathbf{X}^*])^2\right\} + o(P^2). \quad (3)$$

The above expression for mutual information upto second order was in essence derived earlier in [2] and [3], but with more stringent conditions on the input distribution.

A similar analysis as in [1], tailored to the discrete input and continuous output case yields the same expression as in (3) with the expectations involving a discrete input and probability mass function (p.m.f.), while those in [1] involve a continuous input and p.d.f.

In this section, we seek to maximize (3) jointly with respect to $\{\mathbf{X}_i\}_{i=1}^L$ and $\{P_i\}_{i=1}^L$ under an average power constraint $\sum_i P_i\mathrm{tr}(\mathbf{X}_i\mathbf{X}_i^*) \leq E$ and a peak power constraint $\|\mathbf{X}_i\|_\infty = \max_{m,n}|[\mathbf{X}_i]_{mn}| \leq \sqrt{K} \ \ \forall i$.

In [1], the authors use a continuous input and output version of the low SNR mutual information in (3) and maximize this expression with respect to the input distribution under a peak constraint. Their solution however relies on the assumption that the input signal has the form

$$\mathbf{S} = \mathbf{\Phi}\mathbf{V} , \quad (4)$$

where $\mathbf{\Phi}$ is an isotropically distributed unitary random matrix and $\mathbf{V}$ is a diagonal matrix with non-negative entries. While this assumption entails no loss of generality for the case when only the average power is constrained [4], it can be seen that this is not the case when there is also a peak-power constraint $|V_i|^2 \leq K$, which is utilized in [1]. Due to this reason, the maximizations in [1] lead to the conclusion that a single antenna alone should be used in the low SNR regime, which is misleading. The optimization problem that we consider here does not restrict the signals to be of the form (4). Moreover, we assume the most natural peak-power constraint from a practical standpoint, namely, $\|\mathbf{X}\|_\infty \leq \sqrt{K}$ for each $\mathbf{X}\in\mathcal{C}$, in addition to an average power constraint. In contrast to [1], our results indicate that at sufficiently low SNR, the maximum mutual information actually *grows* with the number of transmit antennas used.

Let $\mathbf{u}_i$ be the i^{th} column of the Discrete Fourier Transform (DFT) matrix of size $T\times T$, normalized such that $\|\mathbf{u}_i\| = \sqrt{T} \ \ \forall i$. Let $\mathbf{1}_{N_t}$ be the N_t-length column vector of ones. Let the set of all feasible constellations with a cardinality of L be $\boldsymbol{\mathcal{S}}_L$, which is defined as follows.

$$\Bigg\{(\mathbf{X}_i\ ,P_i)_{i=1}^L\ :\ P_i\geq 0\ ,\ \mathbf{X}_i\in\mathbb{C}^{T\times N_t}\ ,\ \sum_{i=1}^L P_i = 1,$$
$$\sum_{i=1}^L P_i\ \mathrm{tr}(\mathbf{X}_i\mathbf{X}_i^*)\leq E, \|\mathbf{X}_i\|_\infty\leq\sqrt{K},\ i=1\ldots,L\Bigg\} .$$

When $P_i = 0$, the symbol $\mathbf{X}_i$ is not used. Therefore, it can be seen that the set of feasible constellations in $\boldsymbol{\mathcal{S}}_L$ is also included in the set $\boldsymbol{\mathcal{S}}_{L'}$ for any $L' > L$. Let $I^*_{low,L}$ be the maximum possible mutual information achievable by any constellation in the set $\boldsymbol{\mathcal{S}}_L$. We define the maximum mutual information when there is no upper limit on the cardinality L as $I^*_{low} = \lim_{L\to\infty} I^*_{low,L}$. If we were to optimize jointly over $\{\mathbf{X}_i\}_{i=1}^L$, $\{P_i\}_{i=1}^L$ and L, it would amount to maximizing the mutual information at low SNR over the p.m.f. and hence result in I^*_{low}. It will be subsequently shown that the problem of finding $I^*_{low,L}\Big|_{L=T+1}$ provides a solution with respect to I^*_{low}. Another key insight is that even with moderate PAPR, there isn't much to be gained by considering a cardinality more than $T+1$. We thus succeed in bypassing the much harder problem of optimizing over the p.m.f.

A. Both peak and average power constraints are active ($KN_tT \geq E$)

In this subsection, we consider the case when both the peak and average power constraints are active, or $KN_tT \geq E$. The following theorem, which is one of the main results of this paper, characterizes $I^*_{low,L}\Big|_{L=T+1}$ and constellations that achieve it. The proof is omitted in this paper.

Theorem 1: Let $T\geq 2$. Consider the case $KN_tT\geq E$. Then the maximum MI with an unconstrained cardinality can be bounded as

$$I^*_{low,L}\big|_{L=T+1} \leq I^*_{low} < P^2\frac{N_r}{2}\frac{KN_tT}{P} = I^*_{low,UB} , \quad (5)$$

where

$$I^*_{low,L}\big|_{L=T+1} = P^2\frac{N_r}{2}\left(\frac{KN_tT}{P}-1\right) . \quad (6)$$

A constellation that achieves the equality in (6) has cardinality $L = T+1$, which is the smallest possible, and together with its p.m.f. is given by

$$(\mathbf{X}_i \ , P_i) \ = \ \left(\sqrt{K}\ \mathbf{v}_i \mathbf{1}_{N_t}^T \ , \ \frac{E}{(L-1)KN_tT}\right) \quad (7)$$

$$(\mathbf{X}_L \ , P_L) \ = \ \left(\mathbf{0}_{T\times N_t} \ , \ 1-\frac{E}{KN_tT}\right). \quad (8)$$

In (7), $1 \leq i \leq L-1$. In other words, each of the $L-1=T$ non-zero constellation points is formed from a column of a $T \times T$ DFT matrix and this column is repeated over the N_t antennas. ■

Theorem 1 specifies the maximum mutual information achievable by peak and average power constrained constellations which have a cardinality not more than $T+1$. Also, Theorem 1 provides an upper bound on the maximum mutual information achievable when there is no cardinality constraint (I^*_{low}). Since our optimal signal constellation is a set of $T+1$ matrices with their corresponding probabilities, it can be viewed as a space-time code employing unequal transmission probabilities in general, that achieves the maximum mutual information at low SNR. Based on its structure, we call the optimal constellation when $\frac{KN_tT}{E} \geq 1$ as **S**pace **T**ime **O**rthogonal **R**ank one **M**odulation (STORM). It is called so because each non-zero matrix consists of a distinct column of the $T\times T$ DFT matrix repeated along the columns, thereby making each signal matrix of unit rank. Furthermore, every constellation matrix is orthogonal to the other constellation matrices by construction.

It can be seen that for STORM described in (8), the PAPR as defined in (2), is $\eta = \frac{KN_tT}{E} = \frac{KN_t}{P} > 1$.

We may express (5) in terms of η as follows

$$I^*_{low,L}\big|_{L=T+1} \ \leq \ I^*_{low} \ < \ P^2\frac{N_r}{2}\eta T \ , \quad (9)$$

where

$$I^*_{low,L}\big|_{L=T+1} = P^2\frac{N_r}{2}\left(\eta T - 1\right) \ . \quad (10)$$

The next corollary describes when STORM is near-optimal for a fixed η.

Corollary 1: When $\frac{KN_tT}{E} \geq 1$ and $\eta T >> 1$, the ratio between the upper and lower bounds in (9) is nearly 1.

Since I^*_{low} is sandwiched between the strict upper bounds and lower bounds in (5) and (11), Corollary 1 indicates that when $\frac{KN_tT}{E} > 1$ and $\eta T >> 1$, the $T+1$ point constellation is nearly optimal, and there is no need to use more points. We emphasize here that this approximation is well justified even for moderate and practical values for η and T. To illustrate this point, Figure 2 plots the ratio between the lower and upper bounds on I^*_{low} in (5) against the PAPR η for $T=6$. As an example, for $\eta = 4$, the ratio is 0.96. It can hence be seen that in many practical cases of interest, the $T+1$ point STORM scheme is very close to optimal.

B. *Peak-power constraint alone active* $(KN_tT < E)$

The assumption $KN_tT < E$ corresponds to the case when the average power constraint is not active. The following theorem characterizes constellations that achieve $I^*_{low,L}\big|_{L=T}$ in this case. The proof is omitted in this paper.

Theorem 2: Let $T \geq 2$. Consider the case $KN_tT < E$. Then the maximum mutual information with an unconstrained cardinality can be bounded as

$$I^*_{low,L}\big|_{L=T} \leq I^*_{low} < \frac{N_r}{2}K^2N_t^2T \ , \quad (11)$$

where

$$I^*_{low,L}\big|_{L=T} = \frac{N_r}{2}K^2N_t^2(T-1) \ . \quad (12)$$

A constellation that achieves the leftmost equality in (11) has a cardinality $L=T$, which is the smallest possible, and is given by

$$(\mathbf{X}_i \ , P_i) = \left(\sqrt{K}\ \mathbf{v}_i\mathbf{1}_{N_t}^T, \frac{1}{L}\right), \quad 1 \leq i \leq L \ . \quad (13)$$

■

For the scheme described in (13), $\eta = 1$. Such a scheme may be viewed as a special case of STORM wherein there is no zero symbol. Even though this scheme is sub-optimal in general, it is near-optimal when $T >> 1$ since the ratio between the upper and lower bounds in (11) is nearly 1.

The advantage of the constellation indicated in (13) is that the constellation points are equiprobable and the PAPR is one, which facilitates practical implementation. Moreover, when $T >> 1$, it is also near-optimal.

Since STORM achieves a significant fraction of I^*_{low} even for moderate values of η and T, the following insights from its structure and mutual information are of interest.

1) It can be seen that the mutual information of STORM is an increasing function of the maximum peak power K. This is expected, since peaky signaling is known to achieve the noncoherent capacity in the low SNR regime when there is only an average power constraint.
2) When the i^{th} non-zero signal of STORM is transmitted, the received signal is

$$\mathbf{Y} = \sqrt{K}\ \mathbf{v}_i\mathbf{1}_{N_t}^T\mathbf{H} + \mathbf{N} = \sqrt{K}\ \mathbf{v}_i\mathbf{h}^T + \mathbf{N} \ , \quad (14)$$

where the $1 \times N_r$ vector $\mathbf{h}^T = \mathbf{1}_{N_t}^T\mathbf{H}$. Clearly, $\mathbf{h}$ would be distributed as $\mathcal{CN}(\mathbf{0}, N_t\mathbf{I})$. Since the effective channel (14) involves just N_r unknown channel coefficients, the implicit estimation of the channel for the noncoherent scheme involves only N_r coefficents and not the original N_tN_r coefficients. At low SNR, the implicit estimation of a large number of channel coefficients is difficult due to the low received power per path. The unit rank structure of the signals in STORM helps to circumvent this difficulty by focussing the power on just N_r effective unknown path gains, thereby making best use of the multiple transmit antennas.
3) In noncoherent constellation design literature, an often used design criterion is to maximize the worst case chordal distance in a constellation which is given by $\min_{j\neq i}\ \mathrm{tr}\left\{\mathbf{I} - \mathbf{X}_i^*\mathbf{X}_j\mathbf{X}_j^*\mathbf{X}_i\right\}$. For STORM, the worst case chordal distance is the maximum possible as for every $i \neq j$, $\mathbf{X}_i^*\mathbf{X}_j = \mathbf{0}_{N_t\times N_t}$.

4) It can be seen that the mutual information of STORM increases with T and N_t when K, P and N_r are fixed. An intuitive explanation for this is that when the maximum peak power per antenna and time slot is limited, it helps to spread the power across more spatio-temporal dimensions in the manner described by STORM. Further, we can infer from the proof of Theorem 1 that it helps to spread power across more constellation matrices in the low SNR regime in the manner indicated by STORM.
5) The difference between any two different matrices in STORM has unit rank, and hence the scheme has a theoretical diversity order of one at high SNR, based on the rank criterion for coherent space-time codes [5]. Constellation design at high SNR for the coherent MIMO channel is typically geared towards achieving the maximum (full) diversity. Theorem 1 shows that optimal noncoherent constellations at low SNR could have quite the opposite properties as good coherent constellations at high SNR.
6) The entries of all signal matrices in STORM have equal magnitudes, with the exception of the zero matrix. This shows that with the exception of the zero signal matrix, the power is spread equally, thereby minimizing the maximum peak power among all transmit antennas and time slots.

C. *Optimal spectral efficiency*

In the previous subsection, we obtained signals that nearly achieve the maximum mutual information at low SNR even for moderate values of block length T and PAPR η. In this subsection, we will justify that even when the criterion for optimality is to minimize the energy per nat [6] for reliable communications, STORM is close to optimal.

When the regularity conditions described in [1] are satisfied, the mutual information at low SNR therefore grows only as $O(\mathrm{SNR}^2)$. For the optimal scheme, $\lim_{\mathrm{SNR}\to 0} \frac{E_n}{N_0} = \infty$, and hence the minimum energy per nat occurs at an intermediate SNR and not at $\mathrm{SNR} = 0$. It is clear that a signal which achieves the minimum $\frac{E_n}{N_0}$ at a certain SNR, also achieves the capacity $C(\mathrm{SNR})$. Due to the above mentioned reasons, since STORM is nearly optimal with respect to the mutual information at low SNR even for moderate values of T and η, it is a good candidate in terms of minimum energy per nat as well in the low SNR regime. Figure 1 is a representative plot that compares the actual mutual information of STORM at different SNRs with the low SNR approximation given in (3). The actual mutual information is obtained through a Monte-Carlo simulation. Numerical results indicate that ${\frac{E_n}{N_0}}_{min}$ always occurs in the non-asymptotic low SNR regime. As an example, for the parameters indicated in Figure 1, ${\frac{E_n}{N_0}}_{min}$ occurs at a received SNR of $\approx \;\; -11.5$ dB. Indeed, it can also be seen that the difference between the low SNR approximation and the actual mutual information in the vicinity of ${\frac{E_n}{N_0}}_{min}$ is quite small. This appears to be the only way to assess performance, in the absence of the capacity at general SNR. Also, since the mutual information at a general SNR is not available in closed form, these numerical results support the tractable approach of maximizing the low SNR approximation with respect to input constellations to ascertain the most spectrally efficient constellations. Such an approach resulted in STORM as detailed in the earlier part of this section. With STORM, one should however not operate at vanishing SNRs, but instead operate in the vicinity of the SNR at which ${\frac{E_n}{N_0}}_{min}$ occurs.

IV. Conclusion

Under peak and average-power constraints, a new optimal constellation referred to as STORM is obtained via a finite dimensional non-convex optimization of the mutual information upto second order at low SNR jointly over the signal matrices and their respective probabilities, in closed form. The structure and mutual information of STORM provides many new insights into noncoherent MIMO communications in the low SNR regime. STORM is near-optimal with respect to the maximum mutual information under unconstrained cardinality even for a moderate PAPR. Numerical results indicate that STORM is also close to optimal from the important viewpoint of spectral-efficiency.

References

[1] C. Rao and B. Hassibi, "Analysis of multiple-antenna wireless links at Low SNR," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2123 – 2130, Sept. 2004.
[2] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
[3] B. Hajek and V. Subramaniam, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 828–839, Apr. 2002.
[4] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[6] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002, Special Issue on Shannon Theory: Perspective, Trends, and Applications.

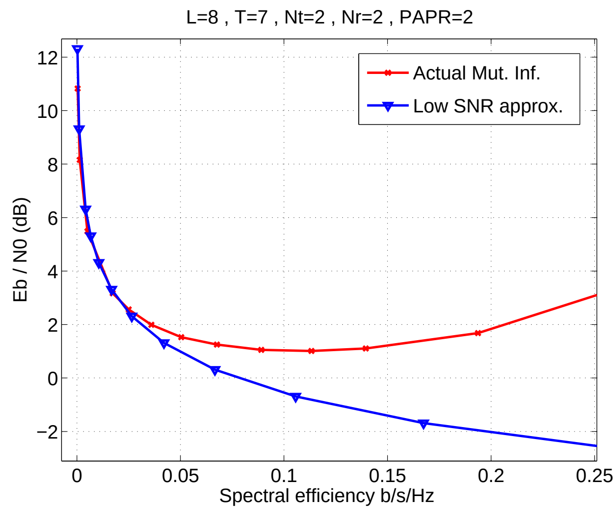

Fig. 1. Comparison of actual mutual information and low SNR approximation for STORM

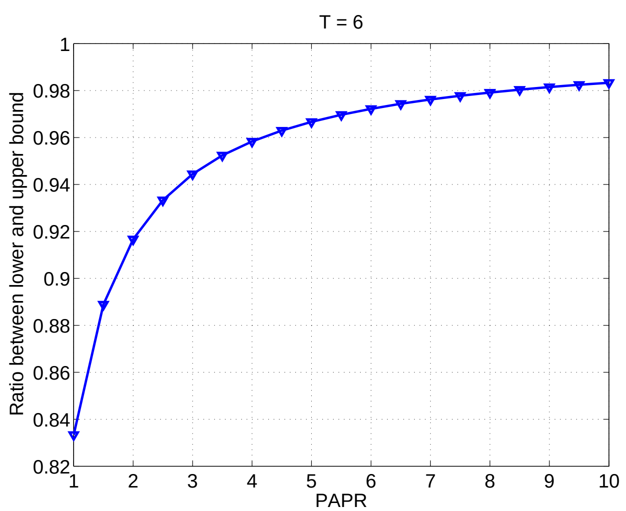

Fig. 2. Ratio of lower and upper bound in (5) vs. PAPR for $T = 6$