Constellation Shaping for Communication Channels with Quantized Outputs

Chandana Nannapaneni, Matthew C. Valenti, and Xingyu Xiang
Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, West Virginia 26506-6109
Email: nannapaneni.chandana@gmail.com, valenti@ieee.org, xiang.xxy2008@gmail.com

***Abstract*—This paper considers a technique to maximize the mutual information between the channel input and channel output of an AWGN channel when the input is chosen from a fixed signal constellation and the output is quantized by a uniform scalar quantizer. The goal is to jointly optimize the input distribution and the quantizer spacing. These parameters have previously been optimized individually, by Le Goff et. al, who optimize the input distribution, and by Singh et. al, who optimize the quantizer spacing. The contribution of our paper is the joint optimization. We consider 16-PAM modulation and investigate the performance of the system with a quantizer resolution of between 4 and 8 bits. Shaping gains in excess of 0.6 dB are observed, even when the output is quantized, when the data is transmitted at rates between 2.5 and 3.2 bits per channel use.**

I. Introduction

The seminal 1948 Shannon paper [1] defines the *capacity* of a channel as the highest rate at which data can be transmitted with arbitrarily low error probability. The capacity is found by maximizing the mutual information between channel input and output, a quantity that is also called the *information rate*, over all the input distributions under the constraints of the system. In this paper, we consider a system that is constrained to have inputs drawn from a fixed constellation and outputs that must be quantized prior to decoding (e.g. by passing the received signal through an analog-to-digital converter). We do not require that the signals be drawn from the constellation with equal probability, and therefore a goal of this paper is to determine the optimal distribution for the input when the output is quantized. The key to the optimization is to compute the mutual information between the input and quantized output, and to maximize this quantity with respect to the input distribution and quantizer spacing. The optimal parameters will depend on the given signal-to-noise ratio (SNR), and since each SNR point requires its own optimization, the optimization task is computationally demanding.

The issues of input-distribution optimization and quantizer-spacing optimization have been considered independently in the recent literature. With respect to quantization, Krone and Fettweis [2] consider the effect of uniform quantizer design on information rate. In addition, the impact of a phase offset between the transmitter and receiver is considered. However, in [2] the input is assumed to be drawn from the constellation according to a uniform distribution. Our present paper continues this line of research by allowing the input to have a nonuniform distribution.

This work was supported in part by the National Science Foundation under Award No. CNS-0750821.

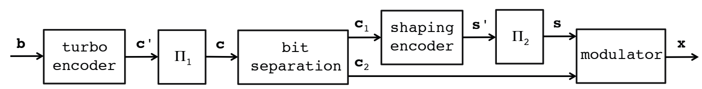

Fig. 1. Transmitter under consideration.

In [3] and [4], Le Goff et al. propose a shaping technique suitable for bit-interleaved systems typical of modern communication standards. The technique involves the use of the short nonlinear shaping codes described in [5] to select from among a plurality of subconstellations. The constellation is partitioned into subconstellations such that subconstellations with lower average energy are selected more frequently than constellations with higher energy. However, the results in [3] are limited to convolutionally-coded 16-QAM, while the results in [4] are limited to turbo-coded PAM. These papers assume that the output is continuous, and do not consider the effect of quantization. Our present paper builds upon this previous research by considering the impact of quantization upon shaping.

There have been some related papers that consider quantization and shaping, though from a very different approach. Wu et. al [6] optimize the input distribution in conjunction with a particular quantization scheme called "modulo-quantization", but they do not optimize the quantizer. Singh et. al [7] discusses a cutting-plane algorithm that gives bounds for the optimal input distribution with quantizer optimization. Our approach is different in the sense that it gives achievable rates and suggests approaches for achieving the shaped constellations.

II. System Model

The system considered in this paper uses the transmitter shown in Fig. 1. We assume that pulse amplitude modulation (PAM) is used, but the extension to other constellations such as QAM or APSK is straightforward. Let $\mathcal{X} = \{x_0, x_1, ..., x_{M-1}\}$ represent the PAM signal set, and $M = |\mathcal{X}|$ be the number of signals in the constellation. Let $m = \log_2 M$ represent the number of bits labeling each symbol. In our numerical examples, we consider $M = 16$ and $m = 4$.

The signal constellation $\mathcal{X}$ is normalized so that its average energy, $\mathcal{E}_s$, is unity. In particular,

$$\mathcal{E}_s = \sum_{j=0}^{M-1} p(x_j)|x_j|^2 \quad = \quad 1 \tag{1}$$

where $p(x_j)$ is the probability that the modulator selects symbol x_j for transmission.

The input to the transmitter is a sequence of equally-likely data bits, denoted as $\mathbf{b}$ in the diagram. The information bits are first encoded by a rate r_c binary turbo encoder to produce the codeword $\mathbf{c}'$. This codeword is permuted by interleaver Π_1 to produce the vector $\mathbf{c}$. The bits in $\mathbf{c}$ are separated into two groups, $\mathbf{c}_1$ and $\mathbf{c}_2$. The vector $\mathbf{c}_1$ is passed through a rate r_s shaping encoder, whose implementation is discussed below, to produce the vector of shaping bits $\mathbf{s}'$. The shaping encoder is designed to ensure that its output is more likely to be a zero than a one. The shaping bits $\mathbf{s}'$ are permuted by a second interleaver Π_2 to produce the vector $\mathbf{s}$.

The bits in $\mathbf{s}$ and $\mathbf{c}_2$ are input to a M-ary modulator. Each M-ary signal is selected by using one "shaping" bit from $\mathbf{s}$ and the remaining $m-1$ bits from $\mathbf{c}_2$. The constellation labeling map is designed such that the shaping bit is used to select the most significant bit, while the unshaped bits correspond to the remaining bits. It follows that the sequence $\mathbf{c}_2$ is $(m-1)$ times longer than the sequence $\mathbf{s}$.

A. Constellation Shaping

With *constellation shaping*, the modulator will pick certain symbols more frequently than the others. Shaping the constellation such that low energy signals are highly probable will conserve the overall energy needed for transmission. For a fixed average signal power, constellation shaping has the potential to increase the information rate of the system. In this paper, we follow the constellation shaping technique proposed in [4], which relies on the use of a *shaping encoder*, which is a nonlinear code designed to produce 0's more frequently that 1's. The overall signal constellation can be partitioned into two subconstellations, one with high energy and one with low energy. The shaping bit can be used to pick from among the two subconstellations. The idea can be generalized to more than two subconstellations, as long as the number of subconstellations is a power of two. In such a case, there will necessarily be more than just one shaped bit per chosen symbol.

Let $\mathcal{C}$ denote a shaping code with k_s input bits and n_s output bits. The code may be constructed according to the methodology of [5], which we summarize as follows. The goal of the construction process is for $\mathcal{C}$ to contain 2^{k_s} distinct codewords of lowest possible Hamming weight. Construction is a recursive process, with $\mathcal{C}$ initialized to contain the all-zeros codeword of length n_s. Codewords of higher and higher weight are recursively added to $\mathcal{C}$ until $|\mathcal{C}| = 2^{k_s}$. Suppose that $\mathcal{C}$ contains all codewords of weight $w-1$ or lower. During the next recursion, a codeword of weight w is drawn without replacement from the set of all weight w codewords and added to $\mathcal{C}$. Weight w codewords are repetitively drawn and added to $\mathcal{C}$ until either the number of distinct codeword in $\mathcal{C}$ is 2^{k_s} or all weight-w codewords have been used. In the former case, the code construction is complete, while in the latter case it moves on to begin adding codewords of weight $w+1$.

TABLE I
(5,3) SHAPING CODE.

3 input bits			5 output bits				
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1
0	1	0	0	0	0	1	0
0	1	1	0	0	1	0	0
1	0	0	0	1	0	0	0
1	0	1	1	0	0	0	0
1	1	0	0	0	0	1	1
1	1	1	1	0	1	0	0

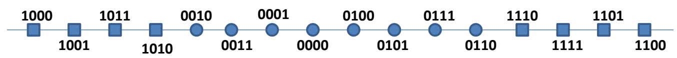

Fig. 2. 16-PAM Constellation and its symbol-labeling map.

As an example, consider the $(n_s, k_s) = (5, 3)$ code. The number of codewords in $\mathcal{C}$ is $2^3 = 8$. There are

$$\binom{5}{0} + \binom{5}{1} + \binom{5}{2} \quad = \quad 16 \tag{2}$$

binary 5-tuples of weight two or less. It follows that $\mathcal{C}$ is a subset of these 5-tuples. In particular, $\mathcal{C}$ will contain all 5-tuples of weight one or less, and will contain two of the weight-two 5-tuples. A possible construction of this code is shown in Table I. Assume that the bits at the input to the shaping encoder are equally likely to be one or zero. Let p_0 be the probability of a zero at the output of the shaping encoder and p_1 be the probability of a one. The value of p_0 may be found by dividing the total number of 1's in all codewords by the total number of bits in all codewords. For the example shown in Table I, we can see that there are a total of nine 1's in the codewords, while the total number of bits is $8 \times 5 = 40$. It follows that $p_0 = 9/40$ and that $p_1 = 31/40$.

To shape a PAM constellation, the constellation should be partitioned into two subconstellations: The $M/2$ signals of lowest energy and the $M/2$ signals of highest energy. The shaping bit is used to select from among these two subconstellations. An example symbol labeling map is shown in Fig. 2, where the most-significant bit (MSB) is used to distinguish between the high-energy and low-energy constellations. Note that a Gray labeling has been used. If the shaping bit is used for the MSB, then each of the $M/2$ low-energy signals will be selected with probability $2p_0/M$, and likewise each of the high-energy signals will be selected with probability $2p_1/M$.

B. Quantization

The selected signal is transmitted through an additive-white Gaussian noise channel (AWGN) with two-sided power-spectral density $N_0/2$. The received signal is passed through a matched-filter and sampled at the baud rate. Assuming perfect timing synchronization, the discrete-time received signal is

$$y \quad = \quad x_j + n \tag{3}$$

where n is a zero-mean Gaussian variable with variance $N_0/2$. Since PAM is one-dimensional, all values in (3) are scalars, but higher-dimension constellations can be considered through the use of a vector model.

To complete the processing in the digital domain, the sampled signal must be *quantized*. The quantizer will map y to the closest quantization level in the set $\mathcal{Y} = \{y_0, y_1, ..., y_{L-1}\}$ of quantization levels, where $L = |\mathcal{Y}|$ is the number of levels and $\ell = \log_2 L$ is the number of bits of resolution for the quantizer. Let D_i be the *quantization region* associated with level y_i, defined as subset of real space whose elements are closer to y_i than any other quantization level. The quantized value of y will be equal to y_i if $y \in D_i$.

Scalar quantizers may be classified as *uniform* if the spacing between adjacent levels is constant, or *nonuniform* if it is not. We assume uniform quantization because most commercially available analog-to-digital converters implement uniform quantization. Let δ represent the quantizer spacing, which is the difference between consecutive quantization levels. When the signal constellation is symmetric about the origin, as is the case for PAM modulation, the quantization levels should also be symmetric about the origin. If L is even, the the quantizer levels will be at half-integer multiples of δ. In particular, the i^{th} level is

$$y_i = \frac{\delta}{2}(2i - L + 1). \tag{4}$$

The boundary between two consecutive levels of a uniform quantizer level lies halfway between the two levels. Let $\beta_{i,i+1}$ represent the boundary between D_i and D_{i+1}. It follows that

$$\beta_{i,i+1} = \frac{y_i + y_{i+1}}{2} = \frac{\delta}{2}(2i - L + 2). \tag{5}$$

Note that the two quantization regions on the ends have only one finite boundary, i.e. $D_0 = (-\infty, \beta_{0,1})$ and $D_{L-1} = (\beta_{L-2,L-1}, \infty)$.

III. Information Rate

The two main parameters to be optimized are p_0, which is the probability of zero at the output of the shaping encoder, and δ, which is the quantization spacing. The optimization is with respect to the mutual information between the channel input and output. In general, the *mutual information* between two random variables quantifies the mutual dependence of one random variable on the other. For a digital communication system, mutual information can be interpreted as the certainty of predicting the input symbol given the output symbol and the input symbol distribution.

The mutual information between channel input X and channel output Y is given as

$$I(X;Y) = E\left[\log_2 \frac{p(Y|X)}{p(Y)}\right] \tag{6}$$

where the expectation $E[\cdot]$ is with respect to the joint distribution of X and Y. The mutual information between the input and output of a channel is commonly called *information rate*, which is a term we use throughout this paper. The numerical value of (6) depends on the type of channel, the type of receiver (e.g. quantized or continuous), and the distribution of the input symbols.

The capacity of the channel is found by maximizing (6) over all possible input distributions,

$$C = \max_{p(x)} I(X;Y). \tag{7}$$

In general, the input can take on any distribution $p(x)$. However, when limited to use the shaping scheme under consideration, $p(x)$ is constrained such that the probability of low energy signals is $2p_0/M$ and the probability of the high energy signals is $2p_1/M$.

When the input is discrete and output is continuous, (6) may be evaluated using

$$I(X;Y) = \sum_{j=0}^{M-1} p(x_j) \int p(y|x_j) \log_2 \frac{p(y|x_j)}{p(y)} dy. \tag{8}$$

The integral in (8) may be numerically evaluated using the Gauss-Hermite quadrature [8].

When the input is discrete and output is quantized, the channel may be modeled as a discrete-memoryless channel (DMC). For the DMC, (6) can be expressed as [9]

$$I(X;Y) = \sum_{j=0}^{M-1} \sum_{i=0}^{L-1} p(x_j) p(y_i|x_j) \log_2 \left(\frac{p(y_i|x_j)}{p(y_i)}\right) \tag{9}$$

where the *transition probability* $p(y_i|x_j)$ is the probability that the output of the quantizer is level y_i given that the input to the channel is x_j. The probability $p(y_i)$ that the quantizer output is level y_i may be found by

$$p(y_i) = \sum_{j=0}^{M-1} p(y_i|x_j) p(x_j). \tag{10}$$

The transition probability may be found by integrating the conditional pdf of y given x_j over level y_i's quantization region,

$$p(y_i|x_j) = \int_{D_i} p(y|x_j) dy = \int_{\beta_{i-1,i}}^{\beta_{i,i+1}} p(y|x_j) dy. \tag{11}$$

The quantization regions D_0 and D_{L-1}, which correspond to the endpoints, can be handled by setting $\beta_{-1,0} = -\infty$ and $\beta_{L-1,L} = \infty$. Because y conditioned on x_i is Gaussian, (11) may be expressed in terms of the Gaussian Q-function as follows:

$$p(y_i|x_j) = Q\left(\frac{x_j - \beta_{i,i+1}}{\sqrt{N_0/2}}\right) - Q\left(\frac{x_j - \beta_{i-1,i}}{\sqrt{N_0/2}}\right). \tag{12}$$

IV. Optimization Strategy

The constrained capacity of the system is found by maximizing the information rate. The constraints of the system are: (1) PAM modulation is used with $M = 16$ signal points, (2) The constellation is shaped using the method described in Subection II-A, (3) The channel is AWGN and the ratio of

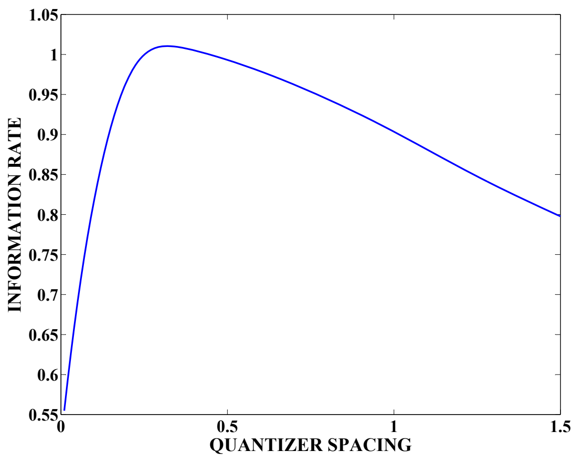

Fig. 3. Information rate as a function of δ (the quantizer spacing) of shaped 16-PAM at $\mathcal{E}_s/N_0 = 2$ dB with $p_0 = 0.81$ and using an $\ell = 4$ bit quantizer.

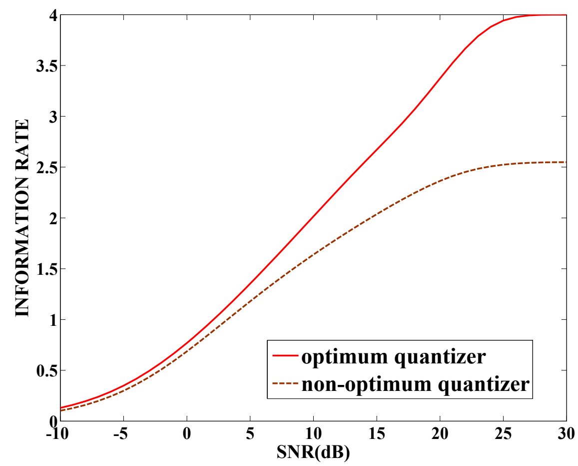

Fig. 4. Information rate as a function of $\mathcal{E}_s/N_0$ for uniform 16-PAM with a 4-bit quantizer using optimal and suboptimal quantizer spacings.

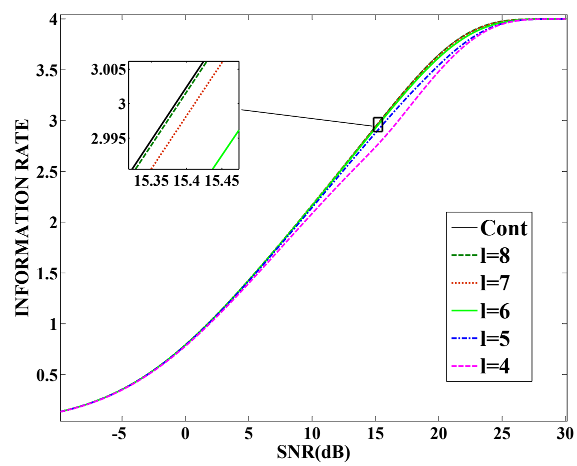

Fig. 5. Information rate as a function of $\mathcal{E}_s/N_0$ of a shaped 16-PAM constellation for a variety of quantizer resolutions.

signal energy to noise-spectral density is $\mathcal{E}_s/N_0$, and (4) The output is quantized by uniform quantizer with a resolution of ℓ bits. As mentioned at the start of Section III, the parameters to be optimized are p_0 and δ. The search space is infinite, since δ may take on any continuous value and p_0 is nearly continuous (though in fact, the set of possible p_0 is finite when the length of the shaping code is limited).

The optimization was run for each value of $\mathcal{E}_s/N_0$ ranging from -10 dB to 30 dB (in 0.1 dB increments) and ℓ ranging over the integers from 4 to 8. The optimization involved computing the information rate as a function of both δ and p_0 using (9) and finding the maximum. The optimization was also performed for the continuous-output case by using Gauss-Hermite quadratures to compute (8) as a function of p_0. Furthermore, the capacity with uniformly-distributed inputs was computed for both the continuous-output and quantized channels by setting $p_0 = 1/2$.

The values of p_0 were varied from 0.5 to 0.99 in increments of 0.005, which provides a range of values that are similar to those achievable with shaping codes of reasonable length. For each value of p_0, the mutual information was computed as a function of δ. However, since it is impossible to test every possible value of δ, a directed search was performed. Fig. 3 shows the dependency of the information rate on the quantizer spacing for the case that $\ell = 4$ bits of resolution, $p_0 = 0.81$, and $\mathcal{E}_s/N_0 = 2$ dB.

Notice that the curve shown in Fig. 3 is well behaved: it is monotonically increasing until it reaches its maximum value, then it is monotonically decreasing. Curves for other values of p_0 and $\mathcal{E}_s/N_0$ exhibit the same type of behavior. Because of this behavior, the search for the maximum can proceed by recursively reducing the search space. In the first recursion, the mutual information is computed over a large range of δ, such as $(0, 10)$ using very coarse increments, such as an increment of $\iota = 1$ so that only integers are considered. The value of δ that maximizes the function is selected, and a second recursion is conducted about that value of δ. Suppose that δ_1 was the maximum delta found in the first recursion, which used increments of ι_1. In the second recursion, the range $(\delta_1 - \iota_1, \delta_1 + \iota_1)$ is searched using a smaller increment $\iota_2 = \iota_1/10$ and a new optimal δ is selected, which we denote δ_2. The process continues recursively so that in the $(j+1)^{th}$ recursion, the search is conducted over the range $(\delta_j - \iota_j, \delta_j + \iota_j)$ in increments of $\iota_{j+1} = \iota_j/10$. The process continues until there is no change in the fourth digit of precision, indicating convergence of δ.

The importance of choosing the optimum δ is illustrated in Fig. 4 for a quantizer resolution of $\ell = 4$. The figure compares the performance of uniform 16-PAM with an arbitrary, suboptimal spacing, against that with an optimal spacing. For the suboptimal system, $\delta = 0.1$ for every value of $\mathcal{E}_s/N_0$. For the optimally-spaced quantizer, the optimal δ is found at each $\mathcal{E}_s/N_0$ and used for information rate calculations. Significant gains can be observed when optimum δ is used.

V. Optimization Results

For each value of ℓ and $\mathcal{E}_s/N_0$, the optimal values of δ and p_0 were found using the optimization strategy described in the previous section. The information rate achieved using these values is shown in Fig. 5 for the five values of ℓ that we considered. Also shown is the information rate when the

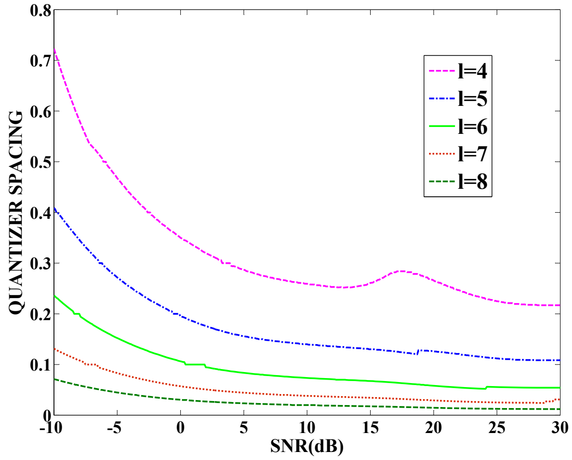

Fig. 6. Optimal quantizer spacing δ.

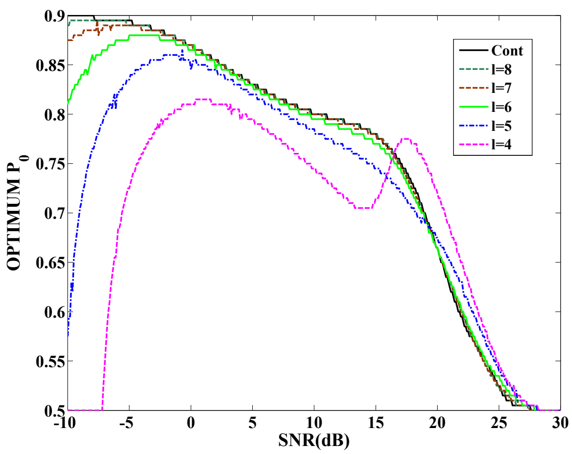

Fig. 7. Optimal shaping probability p_0.

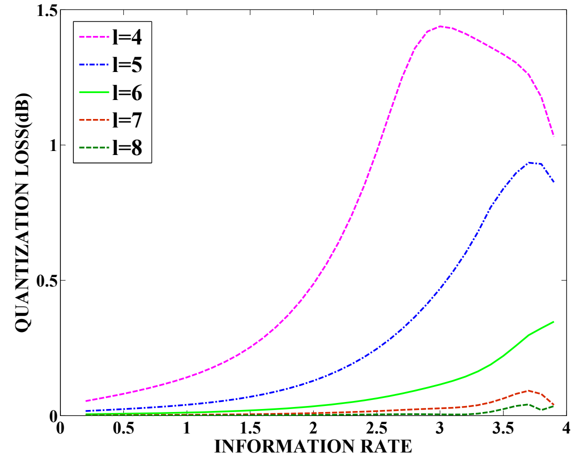

Fig. 8. Quantization loss of a shaped 16-PAM constellation.

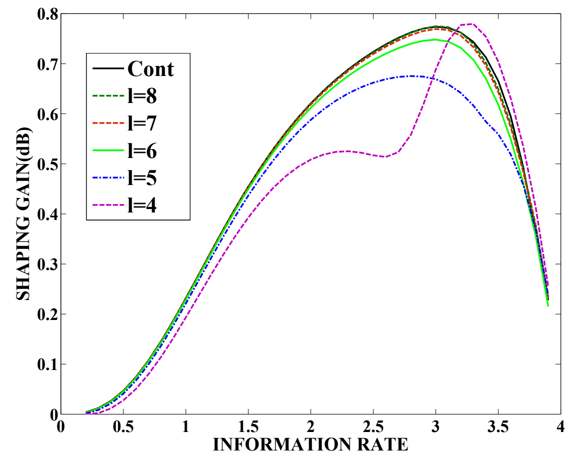

Fig. 9. Shaping gain of a 16-PAM constellation.

output is continuous and p_0 is optimized. The values of δ and p_0 that maximize the information rate are shown in Fig. 6 and Fig. 7, respectively.

The curves shown in Fig. 5 for each of the cases with quantization all indicate a worse performance than the continuous-output case, which can be expected from the information-processing theorem. Let the *quantization loss* be the dB difference between the information rate curve for the continuous-output case and the information rate curve for the case that ℓ bits of quantization are used. From Fig. 5, it can be seen that the quantization loss can be significant for $\ell = 4$, which corresponds to hard-decision decoding, but for $\ell = 7$ and $\ell = 8$ the loss is barely perceptible. The quantization loss is shown for each ℓ in Fig. 8. From this figure, it is clear that an 8-bit uniform quantizer provides nearly the same performance as the continuous output channel.

A key measure of the effectiveness of constellation shaping is the *shaping gain*, defined to be the dB difference between information rate curves for a uniform and shaped constellation. The shaping gain achieved by the 16-PAM shaped constellation is shown in Fig. 9 for the considered values of ℓ and also for the continuous-output channel. While in general, the shaping gain with quantization is less than it is for the continuous-output channel, it is interesting to note that the 4-bit quantizer can actually achieve a higher shaping gain than the continuous-output channel at high information rates.

VI. Implementation

In this section, we compare the actual bit error performance of a turbo-coded 16-PAM both with and without shaping. The turbo code is the one specified in the UMTS standard [10] and the iterative receiver of [4] is used. Let R be the overall rate of the system, taking into account both types of coding (turbo and shaping) and the modulation. For the uniform 16-PAM constellation $R = 4r_c$, i.e. the rate of the turbo code times the number of code bits transmitted per 16-PAM symbol. For the shaped 16-PAM constellation, $R = (3 + r_s)r_c$. For both the shaped and uniform constellations, we pick the code rates such that the overall rate $R \approx 3$, since the shaping gain is high at this rate. For the shaped system, we pick $r_s = 0.7$ which is close to the optimal value for information rate $r = 3$. The actual parameters used for the simulated turbo-coded system are given in Table II.

Bit error rate (BER) results are shown for the shaped and uniform systems in Fig. 10 and Fig. 11. In both figures, the BER is shown as a function of $\mathcal{E}_b/N_0$, where $\mathcal{E}_b = \mathcal{E}_s/R$ is the energy per information bit. The results in Fig. 10 show performance with 7 or 8 bits of resolution. The BER

TABLE II
PARAMETERS USED BY THE SIMULATED 16PAM SYSTEM.

	R	r_c	r_s	Quantizer Spacing				
				$\ell=8$	$\ell=7$	$\ell=6$	$\ell=5$	$\ell=4$
Uniform	2.9940	2000/2672	1	0.0139	0.0272	0.0534	0.1058	0.2165
Shaping	2.9836	2000/2479	7/10	0.0172	0.0339	0.0670	0.1270	0.2781

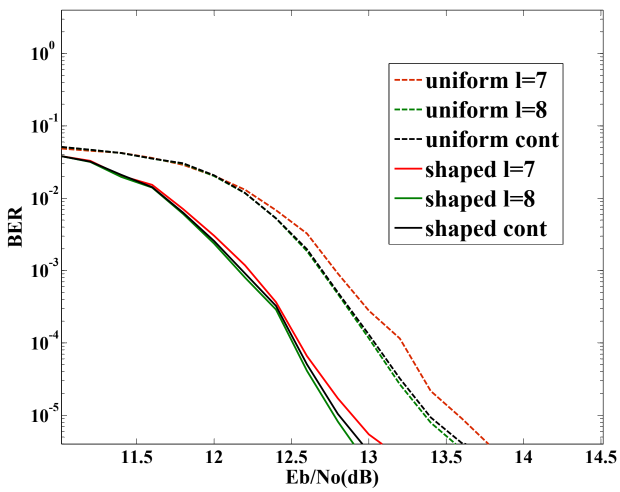

Fig. 10. BER of the turbo-coded system with fine quantizer resolution.

of the continuous-output channel is shown for comparison purposes. The results in Fig. 10 show performance with the coarser resolutions $\ell = \{4, 5, 6\}$. While shaping gains are still achieved, the performance is significantly worse with coarse resolution. Comparing the two curves, we see that a system with shaping and only a 5-bit quantizer performs worse than a uniform system with an 8-bit quantizer.

Table III shows the theoretical values of $\mathcal{E}_b/N_0$ required to achieve an arbitrary low error rate for the rates used in the simulations. These values are found by determining the value of $\mathcal{E}_s/N_0$ required to achieve the information rates listed in Table II, then converting $\mathcal{E}_b/N_0$. Also shown in the table are the actual values of $\mathcal{E}_b/N_0$ required to achieve a bit error rate of 10^{-5} for the turbo-coded system. These are found from the simulation curves. By inspecting the table, it is clear that the actual turbo-coded system performs about 2 dB away from the theoretical bound, and that the shaping gains realized by the actual turbo-coded system (about 0.6 dB) are close to those predicted by the theory (about 0.7 dB).

TABLE III
SNR REQUIRED BY 16PAM AT RATE $R = 3$.

		$\mathcal{E}_b/N_0$ in dB					
		Cont	$\ell=8$	$\ell=7$	$\ell=6$	$\ell=5$	$\ell=4$
Theoretical	Uniform	11.387	11.393	11.410	11.476	11.751	12.74
	Shaping	10.613	10.618	10.640	10.728	11.082	12.051
Actual	Uniform	13.378	13.394	13.585	13.703	14.184	15.524
	Shaping	12.789	12.811	12.922	13.154	13.742	15.112

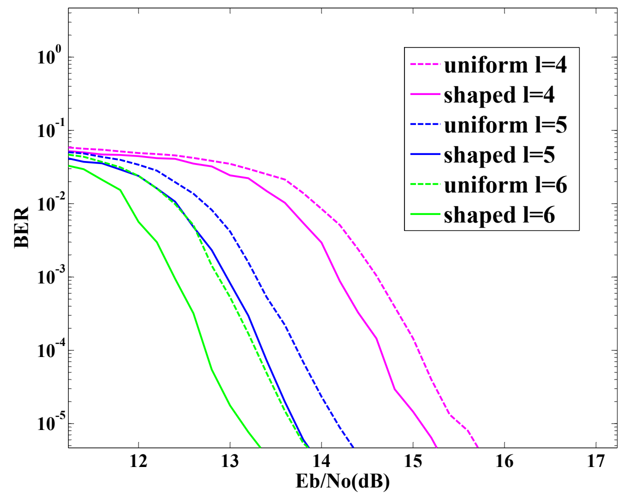

Fig. 11. BER of the turbo-coded system with coarse quantizer resolution.

VII. CONCLUSION

The simple constellation-shaping strategy considered in this paper can achieve shaping gains of over 0.7 dB when used with 16-PAM modulation. However, care must be taken when quantizing the received signal to ensure that the benefits of shaping are not lost. When a finite-resolution quantizer is used, there will necessarily be a quantization loss. The loss can be minimized by using an optimal quantizer spacing. The quantization loss can be further minimized by jointly optimizing the quantizer spacing and the shaping encoder (as parameterized by p_0). When properly optimized, a resolution of 8 bits is sufficient to provide performance that is very close to that of an unquantized system. This work can be further extended to a more-complex two-dimensional modulations like 16-APSK and to higher-dimensional modulations. Using a non-uniform vector quantizer can also be investigated for such systems.

REFERENCES

[1] C. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379–423,623–656, 1948.
[2] S. Krone and G. Fettweis, "Achievable rate of single-carrier systems with optimal uniform quantization at the receiver," in *Proc. IEEE Information Theory Workshop (ITW)*, (Cairo, Egypt), Jan. 2010.
[3] B. K. Khoo, S. Y. L. Goff, and B. S. Sharif, "Bit-interleaved coded modulation with iterative decoding using constellation shaping," *IEEE Trans. Commun.*, vol. 54, pp. 1517–1520, Sept. 2006.
[4] S. Y. L. Goff, B. K. Khoo, and C. C. Tsimenidis, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 2223–2233, Jun. 2007.
[5] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 726–740, Jul. 1990.
[6] Y. Wu, L. M. Davis, and R. Calderbank, "On the capacity of the discrete-time channel with uniform output quantization," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, (Seoul, South Korea), p. 2194, June 2009.
[7] J. Singh, O. Dabeer, and U. Madhow, "Capacity of the discrete-time awgn channel under output quantization," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, (Toronto, Canada), p. 1218, July 2008.
[8] P. E. McIllree, "Channel capacity calculations for M-ary N-dimensional signal sets," Master's thesis, The University of South Australia, Feb 1995.
[9] J. G. Proakis and M. Salehi, *Digital Communications*. New York, NY: McGraw-Hill, Inc., fifth ed., 2008.
[10] European Telecommunications Standards Institute, "Universal mobile telecommunications system (UMTS): Multiplexing and channel coding (FDD)," *3GPP TS 25.212 version 7.4.0*, Jun. 2006.