Signal Shaping Scheme based on the BWT for the Transmission of Non-Binary Sources with Memory over the AWGN Channel

(Invited Paper)

Pedro M. Crespo[†], Xabier Insausti[†], Javier Del Ser[‡] and Jesús Gutiérrez-Gutiérrez[†]

[†]: CEIT and TECNUN (University of Navarra)
20018, Donostia-San Sebastián, Spain
Email: {pcrespo,xinsausti,jgutierrez}@ceit.es

[‡]: TECNALIA-TELECOM
48170 Zamudio, Spain
Email: jdelser@robotiker.es

Abstract—This paper proposes a low-complexity *signal shaping* scheme for the transmission of non-binary symbols generated by sources with memory over the AWGN channel. It is based on using the Burrows-Wheeler Transform (BWT) [1] and standard two dimensional (QAM) constellations. The constellation points are driven based on the first order probabilities of the BWT output symbols so that shaping gain is maximized. The proposed system is an extension of the one introduced in [2] for binary sources. To evaluate its performance, the corresponding average energy per transmitted symbol is compared with that obtained when using an ideal source encoder followed by a standard uncoded QAM transmission scheme. In addition, it is shown that the power gain ratio between both systems can be assessed based on the Kullback-Leibler divergence between the first order probability distributions of the BWT output symbols and the Maxwell-Boltzmann probability distribution, obtained under the constraint of achieving an entropy equal to the entropy-rate of the original source.

Index Terms—Burrows-Wheeler Transform; Kullback-Leibler Divergence; Shaping Gain; Constellation Design.

I. INTRODUCTION

We consider non-binary sources with memory generating M-letter symbols ($M > 2$). It is assumed that the entropy rate $\mathcal{H}$ of the source is greater than 1 bit per source symbol and that the generated information is to be transmitted over an AWGN channel in the bandwidth limited regime, that is, using a transmission scheme capable of sending more than two bits per 2-*real dimensions*, i.e., with a spectral efficiency ρ larger than 2. The standard approach has been to separate the encoding process in two parts: first, a source encoder capable of compressing the source up to its theoretical limit, given by the source's entropy rate $\mathcal{H}$, followed by a channel coded modulation scheme able to support the required spectral efficiency, such as lattice codes or lattice-type trellis codes. The total channel coding gain of such a scheme is the sum of two coding gains, one due to the minimum separation between points in the high-dimensional lattice used by the constellation, and another coming from the choice of the particular set of $2^{\frac{N\rho}{2}}$ points from the N-dimensional lattice.

This latter contribution has come to be known as shaping gain [4]. Shaping gain is upper bounded by $\frac{\pi e}{6}$ (1.53 dB), and this maximum is achieved when implementing high-dimensional spherical constellations driven by an uniform distribution[1]. Practical implementations use trellis shape schemes in an infinite-dimensional sequence space. In this paper we focus on coding schemes that only implement shaping gain. In particular, it proposes a low complexity *signal shaping* scheme for two dimensional $M \times M$-QAM constellations. It uses the reversible Burrows-Wheeler Transform (BWT) [1] to convert the temporal redundancy of the memory source in a symbol-wise redundancy in such a way that the first order probabilities of the BWT output non-binary symbols are used to drive the QAM constellation so that the shaping gain is maximized. The proposed scheme is an extension of the uncoded scheme introduced in [2] suitable for binary sources and 2×2-QAM constellations. In [5] a channel coded transmission scheme is also proposed where the BWT is applied to a Turbo code ($\rho < 2$) in order to improve the coding efficiency when sending the information generated by binary sources.

Since the performance of the proposed scheme only depends on the statistical properties of the source, a fair comparison of our system would be with one that employs an ideal source encoder followed by a standard uncoded QAM modulation scheme. The conditions in which our proposed system outperforms this ideally encoded scheme will be analyzed and characterized in terms of shaping gain. To that end, the Kullback-Leibler divergence or relative entropy [3] will be computed between the first order probability distributions at the output of the BWT and the Maxwell-Boltzmann probability distribution obtained under the constraint of achieving an entropy equal to the entropy rate of the source. The reason behind using this measure of the difference between probability distributions lies on the fact that the Maxwell-Boltzmann distribution is

---

[1]By inducing in this way, near-Gaussian like distributions on its lower constituent constellations.

known to be optimal in achieving maximum shaping gain given a constellation and therefore, will give an upperbound on performance.

The remainder of the paper is outlined as follows. In Section II we describe the proposed system. To simplified the description of the scheme, we first focus on non-uniform memoryless sources. The natural extension to sources with memory is then made possible thanks to the properties of the Burrows-Wheeler Transform (BWT). The performance of the proposed scheme is compared with that of a standard uncoded QAM scheme with ideal source coding. A general expression to evaluate the power gain between both systems is provided in terms of the Kullback-Leibler divergence. Section III presents two numerical examples and the results are discussed. Finally, some concluding remarks are given in Section IV.

II. PROPOSED SCHEME

A. Particular case of Memoryless Sources

We begin by first considering a discrete memoryless source (DMS) with M output letters[2] $\{a_i\}_{i=1}^{M}$, and probability distribution $p \triangleq \{p(a_i)\}_{i=1}^{M}$. Section II-B will cover the more interesting case of sources with memory.

Let $\mathcal{A} \triangleq \{\mathbf{r}_{11}, \ldots, \mathbf{r}_{MM}\}$ denote the set of points of the considered $M \times M$-QAM constellation, and d_{min} the corresponding minimum distance between adjacent points. Observe that the projection of $\mathcal{A}$ into one dimension give rise to a M-PAM constellation with points located at $\mathbf{x} = (x_1, \ldots, x_M) = \frac{d_{min}}{2}(-(M-1), \ldots, -3, -1, 1, 3, \ldots, (M-1))$.

By mapping each pair (a_i, a_j) of source letters into a QAM constellation point $\mathbf{r}_{ij}$, the resulting energy per two dimensions (2-D) $E_s(p)$ is

$$E_s(p) = \sum_{i=1}^{M} \sum_{j=1}^{M} \|\mathbf{r}_{ij}\|^2 p(a_i)p(a_j), \tag{1}$$

and the normalized symbol rate is given by $R = \log_2 M$ binary symbols per 2-D. Define $E_s^*(p)$ to be the minimum achievable energy given a probability distribution p of the source, that is

$$E_s^*(p) \triangleq \min_{\substack{over\ all \\ permutations\ \pi}} \left[\sum_{i=1}^{M} \sum_{j=1}^{M} \|\mathbf{r}_{ij}\|^2 p(a_{\pi(i)})p(a_{\pi(j)}) \right], \tag{2}$$

where $\pi : M \to M$, denotes a permutation mapping. Notice that when transmitting the constellation points over an AWGN channel the resulting probability of error depends on the minimum distance d_{min} between the points of the constellation $\mathcal{A}$ and not in the particular mapping used. The probability of error per two dimensional symbol can be approximated by

$$P_{error} \approx 4 \left(\frac{M-1}{M} \right) Q \left(\sqrt{\frac{d_{min}^2}{2N_0}} \right), \tag{3}$$

where N_0 is the one sided noise power spectral density (W/Hz), and $Q(\cdot)$ denotes the standard Q-function.

[2]Without loss of generality we assume M to be even.

Consequently, given a source distribution $p \triangleq \{p(a_i)\}_{i=1}^{M}$ and a $M \times M$-QAM constellation, the minimum average energy for any probability of error, achieved by any uncoded scheme is given by $E_s^*(p)$.

1) Comparison between Optimal Point Assignment and Ideal Source Coding: We now would like to compare the above achieved average energy per source symbol with that obtained when using the concatenation of an ideal source encoder followed by a standard uncoded modulation using channel symbols belonging to the constellation $\mathcal{A}$. We assume that the encoder removes all the redundancy of the source and returns coded symbols with an equiprobable alphabet of V letters. Notice that the compression rate of this ideal source encoder is given by $H_V(p)$, where $H_V(\cdot)$ stands for the information entropy base V and p represents the probability distribution of the source letters. After compressing, each letter of the coded symbols are mapped to a point in the baseline constellation $\mathcal{A}$, but now having $V \times V$ signal points instead of $M \times M$. The energy per two dimensions of this concatenated system is given by

$$E_s^b = \frac{d_{min}^2(2^\rho - 1)}{6} \tag{4}$$

where $\rho = 2H_2(p)$ is the spectral efficiency of this concatenated system in bits per two dimensions.
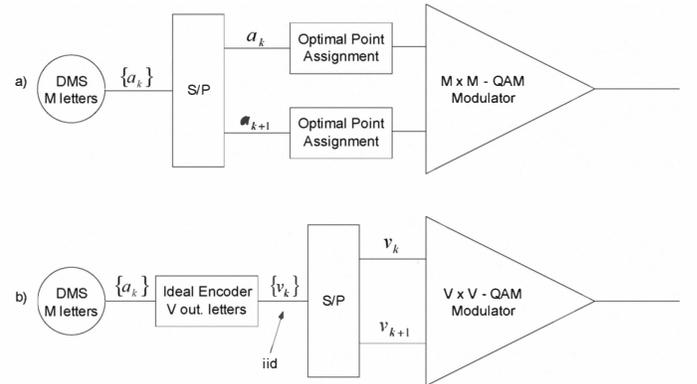

Fig. 1. The two communication systems considered for a memoryless source.

Notice that both systems will render practically the same probability of error since they use constellations with the same minimum distance d_{min}. However, to make a fair comparison we also need to maintain the same normalized symbol rate[3], that is, $R = \log_2 M$ binary symbols per 2-D. Since for each source symbol the encoder produces $H_V(p)$ symbols with an alphabet of V letters, the normalized rate of both schemes will be equal whenever V is such that makes $H_V(p) = 1$. Therefore, this restricts[4] the value of the encoded symbol alphabet to be $V = 2^{H_2(p)}$.

[3]This will render the same information rate (source symbols per second) and bandwidth for both schemes.
[4]Notice that $V = 2^{H_2(p)}$ may not be an integer and therefore, the concatenated system having same information rate and same bandwidth as the proposed scheme is not implementable. However, from the point of view of comparison is still useful.

Under these conditions, the ratio between the powers required by both systems to achieve the same BER is given by

$$G = \frac{E_s^*(p)}{E_s^b} = \frac{6E_s^*(p)}{d_{min}^2(2^\rho - 1)}. \quad (5)$$

Whenever this quantity is less than one (< 0 dB), our scheme outperforms the ideal source encoded concatenated system. Since our communication scheme is uncoded, the power ratio obtained above is merely shaping gain and therefore it should be lower bounded by $6/(\pi e)$ (-1.53 dB). This is shown in Appendix A.

2) Comparison with the Maxwell-Boltzmann distribution:
It is known that the Maxwell-Boltzmann (MB) distribution (p_{MB}) achieves the highest shaping gain for any given constellation and any given entropy $H_2(p_{MB})$ [4]. Consequently, the power gain G in expression (5) should be minimum whenever the distribution of the source symbols p equals p_{MB}. Based on this fact, one could compare the given distribution p with p_{MB} and obtain in this way a measure of how far we are from achieving this minimum gain. Notice that in this comparison the entropy of both distributions should be equal, that is, $\rho/2 = H_2(p) = H_2(p_{MB})$. The MB probability distribution $p_{MB} \triangleq \{p_{MB}(a_1), \ldots, p_{MB}(a_M)\}$ is defined as

$$p_{MB}(a_i) \triangleq \Theta(\lambda_M)e^{-\lambda_M x_i^2}, \quad i = 1, \ldots, M, \quad (6)$$

$$\Theta(\lambda_M) \triangleq \left(\sum_{i=1}^{M} e^{-\lambda_M x_i^2}\right)^{-1}, \quad (7)$$

where the parameter $\lambda_M > 0$ controls the entropy of the resulting p_{MB} distribution. Therefore, λ_M should be such that $\rho = 2H_2(p_{MB})$. It can be shown that the entropy $H_2(p_{MB})$ can be written as

$$H_2(p_{MB}) = \frac{\lambda_M E_s^*(p_{MB})}{2 \ln 2} - \log_2 \Theta(\lambda_M), \quad (8)$$

where $E_s^*(p_{MB})$ is the resulting energy per two dimensions of the constellation $\mathcal{A}$ when the probability distribution of the source follows the MB distribution, i.e.

$$E_s^*(p_{MB})$$
$$\triangleq \min_{\substack{over\ all \\ permutations\ \pi}} \left[\sum_{i=1}^{M}\sum_{j=1}^{M}\|\mathbf{r}_{ij}\|^2 p_{MB}(a_{\pi(i)})p_{MB}(a_{\pi(j)})\right]. \quad (9)$$

Therefore, for any other distribution p

$$E_s^*(p) \geq E_s^*(p_{MB}) = \frac{\rho + 2\log_2 \Theta(\lambda_M)}{\lambda_M} \ln 2.$$

On the other hand, $E_s^*(p)$ is upper bounded by $E_s^*(p_W)$ where p_W is the probability distribution that yields the maximum energy (or minimum shaping gain) after optimal mapping. This distribution is given by

$$p_W = \{\alpha, \alpha, \ldots, 1 - (M-1)\alpha\}, \quad (10)$$

where α is chosen so that p_W satisfies the fixed entropy condition. This can be checked by noticing that any distribution obtained by a perturbation of p_W (and keeping the

same entropy) will necessarily decrease the energy of the constellation after performing the optimal mapping.

Therefore, given M and a spectral efficiency ρ, we can calculate the minimum G_{MB} and maximum G_W achievable gains by using distributions p_{MB} and p_W, respectively. The power gain of any other distribution $\{p_i\}_{i=1}^{M}$ will be located between these two bounds. Figure 2 plots the values of these bounds versus ρ, for a source with $M = 8$ output letters. Shown in this figure is also the 0 dB line. Any scheme with a gain below this line will outperform to the standard ideal source encoded scheme.
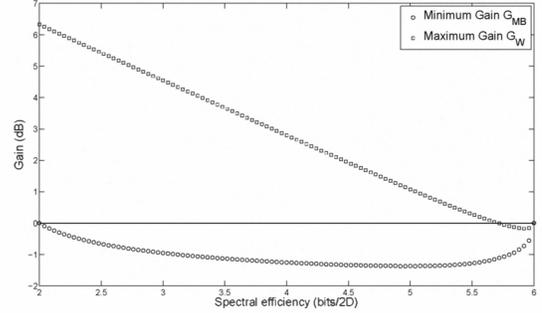

Fig. 2. Maximum and minimum shaping gains achievable for a source with $M = 8$ letters

Since the performance of our proposed scheme only depends on the disparity between the given source distribution p and the corresponding MB distribution p_{MB}, it is reasonable to try to express the power gain G in expression (5) as a function of the distance between these two distributions. A common measure of distance between two probability distributions is the Kullback-Leibler divergence or relative entropy defined as

$$D(p\|p_{MB}) \triangleq \sum_{i=1}^{M} p(a_i) \log_2 \frac{p(a_i)}{p_{MB}(a_i)}$$
$$= -H_2(p) - \log_2 \Theta(\lambda_M) + \frac{\lambda_M}{2 \ln 2} E_s^*(p). \quad (11)$$

The Kullback-Leibler divergence does not fulfill the conditions to be a real metric, but it turns out that there is a simple relationship between the Kullback-Leibler divergence $D(p\|p_{MB})$ and G. It is easy to show that

$$E_s^*(p) = 2 \ln 2 \frac{D(p\|p_{MB}) + \frac{\rho}{2} + \log_2 \Theta(\lambda_M)}{\lambda_M}, \quad (12)$$

and substituting this expression into (5) yields the sought relation

$$G = \frac{12 \ln 2}{d_{min}^2(2^\rho - 1)} \cdot \frac{D(p\|p_{MB}) + \frac{\rho}{2} + \log_2 \Theta(\lambda_M)}{\lambda_M}. \quad (13)$$

Notice that since $D(p\|p_{MB}) = 0$ iff $p = p_{MB}$, expression (12) is an alternative way of showing (by using an information-theoretic expression rather than by analysis) that for a given ρ, p_{MB} is indeed the distribution that minimizes the energy of the constellation. For large ρ Appendix A shows that (13) is lower-bounded by $6/(\pi e)$, and this bound is achieved when $M \to \infty$ and $p \to p_{MB}$.

B. Application to sources with memory

Let us now consider a source with memory, with an alphabet of M letters and entropy rate $\mathcal{H}_M(\mathbf{p}) < 1$. In this situation the standard scheme using an ideal source encoder will render an energy per 2D given by

$$E_s^b = \frac{d_{min}^2(2^{2\mathcal{H}_2(\mathbf{p})} - 1)}{6}. \tag{14}$$

We would like to compare this energy with the energy obtained by our previous scheme. However, it is clear that our previous scheme is no longer valid to exploit shaping gain, since the redundancy of the source is not seen symbol-wise but time-wise. In other words, although the entropy rate $\mathcal{H}_M(\mathbf{p}) < 1$, the entropy per letter $H_M(p)$ can be close to one. Therefore, one needs to transform the temporal redundancy of the source into spatial redundancy before applying the scheme. This is done by partitioning the output source sequence into blocks of K symbols and transforming these blocks using the Burrows-Wheeler Transform (BWT). The BWT algorithm will rearrange the symbols inside a block in such a way that they become approximately independent, with a probability distribution that is now time dependent (i.e., the BWT output process becomes cyclostationary with period K). Let p_k denote the probability distribution of the BWT transformed symbol at position $1 \leq k \leq K$, and $E_s^*(p_k)$ the corresponding energy with optimal point assignment. In this situation, everything stated before holds for every time instant k. However, the overall energy is now obtained by averaging $E_s^*(p_k)$ over a block, that is,

$$\overline{E_s^*}(\mathbf{p}) = \frac{1}{K} \sum_{k=1}^{K} E_s^*(p_k) \triangleq \langle E_s^*(p_k) \rangle. \tag{15}$$

Let p_{MB} be the MB distribution with the parameter λ_M such that $H_2(p_{MB}) = \mathcal{H}_2(\mathbf{p})$. Then from expression (12), the average energy can be written as

$$\overline{E_s^*}(\mathbf{p}) = \frac{1}{K} \sum_{k=1}^{K} 2 \ln 2 \frac{D(p_k \| p_{MB}) + \frac{\rho_k}{2} + \log_2 \Theta(\lambda_M)}{\lambda_M}$$
$$= 2 \ln 2 \frac{\langle D(p_k \| p_{MB}) \rangle + \langle \frac{\rho_k}{2} \rangle + \log_2 \Theta(\lambda_M)}{\lambda_M}, \tag{16}$$

where $\rho_k = 2H_2(p_k)$. The average power gain $\overline{G}$ can now be written as

$$\overline{G} = \frac{12 \ln 2 \left(\langle D(p_k \| p_{MB}) \rangle + \langle \frac{\rho_k}{2} \rangle + \log_2 \Theta(\lambda_M) \right)}{d_{min}^2 \left(2^{2\mathcal{H}_2(\mathbf{p})} - 1 \right) \lambda_M}. \tag{17}$$

Assuming that when the BWT block length K increases without bound the BWT transformed symbols become independent, the following limit will hold

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} H_2(p_k) = \mathcal{H}_2(\mathbf{p}). \tag{18}$$

Therefore, for K large it is reasonable to approximate $\langle \frac{\rho_k}{2} \rangle = \langle H_2(p_k) \rangle$ by $\mathcal{H}_2(\mathbf{p})$. Using this approximation and

the fact that $\mathcal{H}_2(\mathbf{p}) = H_2(p_{MB})$, expression (17) reduces to

$$\overline{G} \approx \frac{12 \ln 2 \left(\langle D(p_k \| p_{MB}) \rangle + H_2(p_{MB}) + \log_2 \Theta(\lambda_M) \right)}{d_{min}^2 \left(2^{2H_2(p_{MB})} - 1 \right) \lambda_M}, \tag{19}$$
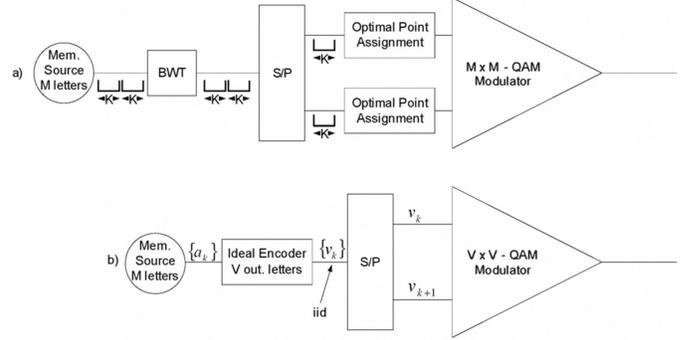
where the approximation is tight for large K.

Fig. 3. The two communication systems considered for a memory source.

Notice that if the sequence $\{p_k\}_{k=1}^{K}$ has all its terms equal to a fixed probability distribution $p_\mathcal{H}$ such that its entropy $H_2(p_\mathcal{H}) = \mathcal{H}_2$, the above expression for $\overline{G}$ reduces to expression (13) with $\rho = 2\mathcal{H}_2$. That is, the G that would be obtained by a memoryless source with an entropy H_2 equal to the entropy rate of the original memory source, and letter distribution $p_\mathcal{H}$.

Furthermore, $\overline{G}$ is lowerbounded by

$$\overline{G} \geq \frac{12 \ln 2}{d_{min}^2 (2^{2\mathcal{H}_2(\mathbf{p})} - 1)} \frac{\mathcal{H}_2(\mathbf{p}) + \log_2 \Theta(\lambda_M)}{\lambda_M}, \tag{20}$$

that is, the same lowerbound found for the case of a memoryless source but now using $\mathcal{H}_2(\mathbf{p})$ instead of $H_2(p)$.

III. EXAMPLES AND RESULTS

We give two examples of sources with memory and compute the corresponding power gain G. Both sources have an alphabet of $M = 6$ letters and are chosen to have a similar entropy rate $\mathcal{H}_2(\mathbf{p}) \approx 2.175$ bits and therefore the same effective spectral efficiency of $\rho = 4.35$ bits/2D. However, one is modeled by a Markov chain whereas the other by a Hidden Markov model (HMM).

A. Source modeled by a Markov Chain

This source is modeled by means of a Markov Chain with 6 states, each one associated with an output letter. The state transition probability matrix is given by,

$$\mathbf{A} = \begin{pmatrix} 0.495 & 0.101 & 0.101 & 0.101 & 0.101 & 0.101 \\ 0.101 & 0.495 & 0.101 & 0.101 & 0.101 & 0.101 \\ 0.101 & 0.101 & 0.495 & 0.101 & 0.101 & 0.101 \\ 0.101 & 0.101 & 0.101 & 0.495 & 0.101 & 0.101 \\ 0.101 & 0.101 & 0.101 & 0.101 & 0.495 & 0.101 \\ 0.101 & 0.101 & 0.101 & 0.101 & 0.101 & 0.495 \end{pmatrix} \tag{21}$$

To obtain the probability profile $\{p_k\}_{k=1}^{K}$ at the output of the BWT for such a source, we first set the source block

length to $K = 64000$ M-ary symbols drawn from the set $\{a_1, \ldots, a_M\}$. We approximate the sought probability profile $p_k \triangleq \{p_k^1, \ldots, p_k^M\}$ by using frequency of occurrence as

$$p_k^i \approx \frac{1}{Q} \sum_{q=1}^{Q} \mathbb{I}(t_k(q) = a_i), \quad i = 1, \ldots, M, \quad (22)$$

where $t_k(q)$ denotes the output symbol of the BWT at position $k \in \{1, \ldots, K\}$ and block $q \in \{1 \ldots, Q\}$, and $\mathbb{I}(t_k(n) = 0)$ is an indication function taking value 1 when its argument is true and 0 otherwise. The estimated probability profile averaged over $Q = 60000$ different MC source blocks is shown in Figure 4.
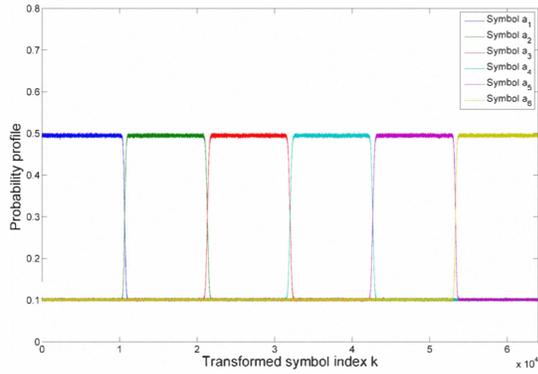
Fig. 4. Estimated probability profile for a source modeled by means of a Markov Chain.

Given the probability profile we can apply the optimal mapping method so that the minimum energy of the constellation can be calculated at every time instant k. The entropy of the probability profile at the output of the BWT varies between 2.15 and 2.38 bits, and the average entropy (which actually should meet the entropy rate) is close to 2.175 bits. Figure 5 shows the instantaneous gain for the considered source, which gives $\overline{G} \approx 0.60$ dB on average.

B. Source modeled by a Hidden Markov Model

In this case the source is modeled by a 6-state HMM with $M = 6$ output letters. The state transition and output probability matrices are given by

$$\mathbf{A} = \begin{pmatrix} 0.7 & 0.06 & 0.06 & 0.06 & 0.06 & 0.06 \\ 0.06 & 0.7 & 0.06 & 0.06 & 0.06 & 0.06 \\ 0.06 & 0.06 & 0.7 & 0.06 & 0.06 & 0.06 \\ 0.06 & 0.06 & 0.06 & 0.7 & 0.06 & 0.06 \\ 0.06 & 0.06 & 0.06 & 0.06 & 0.7 & 0.06 \\ 0.06 & 0.06 & 0.06 & 0.06 & 0.06 & 0.7 \end{pmatrix}, \quad (23)$$

$$\mathbf{B} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0 & 0.5 \end{pmatrix}. \quad (24)$$

We are using the same block length ($K = 64000$) and averaging over the same number of blocks ($Q = 60000$). The probability profile obtained after the BWT is shown in Figure 6 (next page).
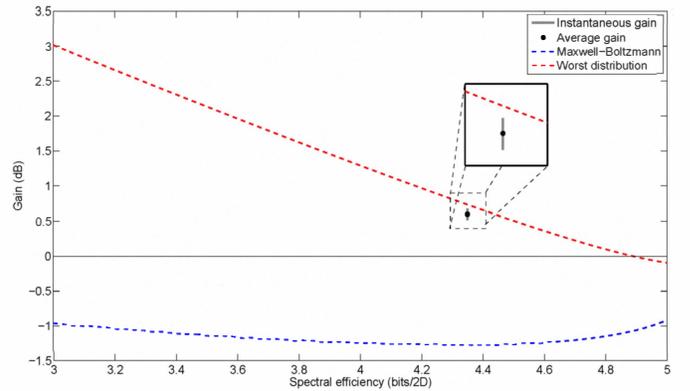

Fig. 5. Gains for the Maxwell-Boltzmann distribution and for the considered MC source.

For this source the average power gain is $\overline{G} \approx -0.69$ dB, which means that we achieve better performance by using the BWT and the optimal mapping rather than the ideal source compressor. Figure 7 (next page) depicts the gain of the considered source for each time instant and the average shaping gain.
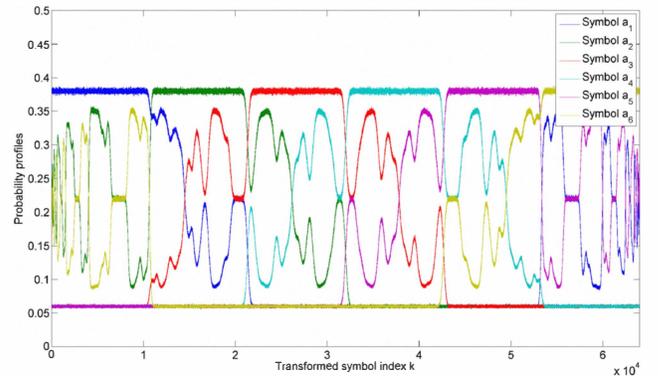

Fig. 6. Estimated probability profile for a source modeled by means of a Hidden Markov Model.

IV. Conclusions

We have proposed a communication system that integrates the block sorting Burrows-Wheeler Transform with an optimal energy allocation scheme for non-binary sources with memory. Depending on the joint probability distribution of the source, our proposed system outperforms a system consisting of an ideal source encoder followed by an uncoded QAM modulation scheme. Furthermore, we have shown that the power gain ratio between these two schemes can be determined solely by the entropy rate of the source and the average Kullback-Leibler divergence between the first order probability distributions of

the BWT output symbols and the Maxwell-Boltzmann probability distribution, obtained under the constraint of achieving an entropy equal to the entropy-rate of the source.
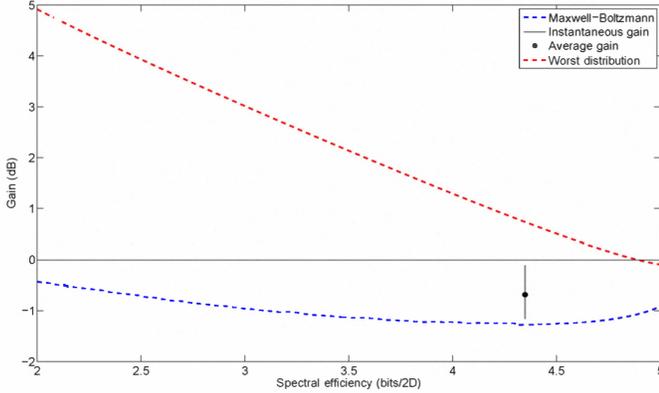

Fig. 7. Gains for the Maxwell-Boltzmann distribution and for the considered HMM source.

APPENDIX A

We want to prove that expression (5) is lower bounded by

$$\frac{2^\rho}{(2^\rho - 1)} \frac{6}{\pi e},$$

and this bound is achieved when $M \to \infty$ and $p \to p_{MB}$.

Proof: Recall from Section II-A that given a memoryless source with symbol alphabet of M letters, probability distribution p and entropy $H_2(p)$, the power gain ratio G is given by expression (13), that is,

$$G(p, M, \rho) = \frac{6}{d_{min}^2 (2^\rho - 1)} \cdot E_s^*(p, M, \rho), \quad (25)$$

and $E_s^*(p, M, \rho)$ can be written as

$$E_s^*(p, M, \rho) = \frac{D(p\|p_{MB}) + \frac{\rho}{2} + \log_2 \Theta(\lambda_M)}{\lambda_M} 2 \ln 2,$$

where p_{MB} denotes the Maxwell-Boltzmann distribution (6) obtained under the constraint of achieving the same entropy than the one of the source, and $\rho = 2H_2(p)$. Clearly, this expression is lower bounded by setting $p = p_{MB}$, i,e,

$$G(p, M, \rho) \geq \frac{12 \ln 2}{2^\rho - 1} \left(\frac{\frac{\rho}{2} + \log_2 \Theta(\lambda_M)}{\lambda_M} \right), \quad (26)$$

It is easily shown by an argument of reduction-to-the-absurd that $E_s^*(p_{MB}, M, \rho)$ is a decreasing sequence of M when all the other parameters are kept fixed. Therefore,

$$G(p, M, \rho) = \frac{6 E_s^*(p, M, \rho)}{d_{min}^2 (2^\rho - 1)} \geq$$

$$\geq \lim_{m \to \infty} \frac{12 \ln 2}{d_{min}^2 (2^\rho - 1)} \frac{\frac{\rho}{2} + \log_2 \Theta(\lambda_m)}{\lambda_m}, \quad (27)$$

where the MB parameter λ_m is such that makes the corresponding Maxwell-Boltzmann distribution on m letters to have an entropy $H_2(p_{MB}^{(m)}) = \rho/2$.

Next, we show that the last limit converges to

$$\frac{2^\rho}{(2^\rho - 1)} \frac{6}{\pi e}.$$

To that end, choose $d_{min} = 1/\sqrt{M}$ in constellation $\mathcal{A}$. For M large the Maxwell-Boltzmann distribution p_{MB} can be approximated by the sampled Gaussian distribution[5] $\mathcal{N}(0, \sigma_M^2)$ [4], i.e.,

$$p_{MB}(i) = \Theta(\lambda_M) e^{-\lambda_M x_i^2} \approx \frac{1}{\sqrt{2\pi \sigma_M^2}} e^{-\frac{x_i^2}{2\sigma_M^2}} \Delta_M, \quad (28)$$

where $\Delta_M = \frac{1}{\sqrt{M}}$ denotes the separation between samples, and the σ_M^2 variance depends on M. For M large, σ_M^2 can be approximated by using the condition that for all M, $H_2(p_{MB})$ should be remained constant with value $\rho/2$, and the fact that the following relation between the differential entropy h_2 of a continuous random variable and the corresponding discrete entropy H_2 of the quantized random variable holds ([6, Chapter 8]),

$$H_2(p_{MB}) \approx h_2(\mathcal{N}(0, \sigma_M^2)) - \log_2 \Delta_M, \quad (29)$$

Expression (29) yields

$$\frac{\rho}{2} \approx \frac{1}{2} \log_2(2\pi e \sigma_M^2) + \frac{1}{2} \log_2(M) = \frac{1}{2} \log_2(2\pi e \sigma_M^2 M).$$

Consequently, for large M

$$\sigma_M^2 \to \frac{2^\rho}{2\pi e M}, \quad \lambda_M \to \frac{1}{2\sigma_M^2}, \quad \Theta(\lambda_M) \to \frac{1}{\sqrt{2\pi \sigma_M^2 M}}.$$

Substituting these approximations in expression (27), we have

$$G(p, M, \rho)$$
$$\geq \lim_{m \to \infty} \frac{m \cdot 12 \ln 2}{2^\rho - 1} \left(\frac{\frac{\rho}{2} + \log_2 \Theta(\lambda_M)}{\lambda_m} \right) = \frac{2^\rho}{2^\rho - 1} \frac{6}{\pi e},$$

and we get the sought result. For large ρ this bound reduces to $6/(\pi e)$. ∎

[5]That is, p_{MB} can be approximated by the discrete probability distribution that results from quantizing a continuous Gaussian random variable $\mathcal{N}(0, \sigma_M^2)$, with quantization step Δ_M.

References

[1] M. Burrows, D. A. Wheeler, "A Block Sorting Lossless Data Compression Algorithm", *Digital Systems Center*, research report 124, 1994.

[2] P. M. Crespo, E. Loyo, J. Del Ser, "Uncoded Optimal Binary Communication for Sources with Memory using the Burrows Wheeler Transform" AEÜ *International Journal of Electronics and Communications*, Vol. 62, No. 8, pp. 597–609, August 2007.

[3] S. Kullback, R. A. Leibler, "On Information and Sufficiency" *Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79–86, 1951.

[4] R. F. H. Fischer, "Precoding and Signal Shaping for Digital Transmission", Ed. *John Wiley & Sons*, New York, 2002.

[5] J. Del Ser, P. M. Crespo, I. Esnaola, J. Garcia-Frias, "Turbo Joint Source-channel Coding of Sources with Memory using the Burrows Wheeler Transform." *IEEE Transactions on Communications*, to appear.

[6] T. M. Cover, J. A. Thomas, "Elements of Information Theory", *John Wiley & Sons*, New York, 1991.