

## Article

# Combining the Burrows-Wheeler Transform and RCM-LDGM Codes for the Transmission of Sources with Memory at High Spectral Efficiencies

Imanol Granada <sup>1,\*</sup> , Pedro M. Crespo <sup>1</sup> and Javier Garcia-Frías <sup>2</sup><sup>1</sup> Department of Basic Science, Tecnun—University of Navarra, 20018 San Sebastian, Spain; pcrespo@tecnun.es<sup>2</sup> Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA; jgf@udel.edu

\* Correspondence: igranada@tecnun.es

Received: 11 March 2019; Accepted: 4 April 2019; Published: 8 April 2019



**Abstract:** In this paper, we look at the problem of implementing high-throughput Joint Source-Channel (JSC) coding schemes for the transmission of binary sources with memory over AWGN channels. The sources are modeled either by a Markov chain (MC) or a hidden Markov model (HMM). We propose a coding scheme based on the Burrows-Wheeler Transform (BWT) and the parallel concatenation of Rate-Compatible Modulation and Low-Density Generator Matrix (RCM-LDGM) codes. The proposed scheme uses the BWT to convert the original source with memory into a set of independent non-uniform Discrete Memoryless (DMS) binary sources, which are then separately encoded, with optimal rates, using RCM-LDGM codes.

**Keywords:** Burrows-Wheeler Transform; hidden Markov models; Markov processes; rate adaptation; RCM-LDGM codes

## 1. Introduction

When considering sources with memory, Shannon's JSC coding theorem states that reliable transmission is only possible if

$$\mathcal{H}(S)R \leq C, \quad (1)$$

where  $\mathcal{H}$  is the entropy rate of the source in bits per source symbol,  $C$  is the capacity of the channel in information bits per channel use and  $R$  is the JSC code's rate (source symbols per channel use). A fundamental result of information theory is the Separation Theorem, which states that provided unbounded delay, no optimality is lost by designing the joint decoder as the concatenation of an optimal source code with compression rate  $\mathcal{H}$ , and a capacity-achieving channel code of rate  $C$ . This independent design of source and channel codes allows for diverse sources to share the same digital media. Therefore, source coding and channel coding have traditionally been addressed independently of each other.

Nevertheless, when complexity is an issue and the length of the input block is constrained, the overall performance can be improved by using a JSC coding scheme. In this case, the joint decoder employs the inherent redundancy of the uncompressed source [1]. The main approaches to JSC can be categorized as follows:

- Ad hoc approaches where the channel encoder is applied to a given source compression format [2–4]. High-level information from the source code is used in the decoding process, which makes this approach highly dependent on the source encoder.

- Schemes where, assuming a Markovian model for the source [5,6], the factor graph of the source is combined with the one representing the channel code in order to perform joint decoding, such as in [7–9] (using Turbo codes) or in [10,11] (using LDPC codes).
- Strategies where context estimation techniques such as the Discrete Universal Denoiser (DUDE) [12] or the Burrow-Wheeler Transform [13–15] are used to help the decoding process.

To the best of our knowledge, when considering discrete sources with memory, existing JSC schemes in the literature [7–16] are restricted to encoders that produce binary output symbols, such as Turbo, LDPC, or Polar codes. This leads to throughputs (or spectral efficiencies) bounded by 2 binary source symbols per complex channel use. To achieve higher rates, larger alphabets must be used for the encoded symbols. To that end, we will consider a high-throughput, hybrid analog-digital JSC coding scheme, recently proposed for non-uniform discrete memoryless sources (DMSs) in [17,18]. This hybrid scheme is constructed by generating most of the output symbols from weighted linear combinations of the input bits, as in Rate-Compatible Modulation (RCM), and a few of them using a Low-Density Generator Matrix (LDGM) code. The basic idea is that the RCM scheme corrects most of errors, leaving residual errors to be corrected by the LDGM code. Source non-uniformity is exploited at the decoder and the encoder is optimized depending on the source non-uniformity to further improve the system performance. In what follows we will denote these codes by RCM-LDGM. These codes provide smooth rate adaptation in a broad dynamic range in a very simple way, adding or removing rows to the encoding matrix.

The proposed high-throughput JSC scheme belongs to the third group of techniques. It uses the BWT to convert the original source with memory into a set of independent non-uniform discrete (binary) memoryless sources (DMSs). The resulting DMSs are each RCM-LDGM encoded with rates adapted to the entropy of the corresponding DMS. It should be pointed out that the authors in [13,14] use the BWT as a context estimation tool to help in the iterative decoding process. Specifically, they apply the BWT at each iterative decoding step and then pass the first order probability distribution of its output to the constituent decoders of the LDPC or Turbo codes. Differently, and different from [15], which applies the BWT in the transmitter before coding and then uses the first order probability distribution of the BWT output sequences to optimize the output energy of the binary modulator, the proposed scheme uses this first order probability distribution to optimize the rates at which different segments of the BWT output sequences are transmitted. Thus, the main contribution of this paper is the proposal of a novel high-throughput JSC scheme for sources with memory based on the application of the BWT and optimal rate allocation. To the best of our knowledge, for sources with memory no high-throughput JSC system has appeared in the literature.

The remainder of this paper is organized as follows. Section 2 briefly reviews some preliminary concepts required for the explanation of the proposed BWT-JSC scheme. Section 3 presents our proposed JSC scheme, leaving for Section 4 the corresponding performance evaluation. Finally, Section 5 provides the concluding remarks.

## 2. Preliminaries

In this section, we briefly review the statistical characterization of a binary source with memory by Markov Models, the Burrows-Wheeler Transform, and the design of parallel RCM-LDGM codes, which are the building blocks of our proposed BWT-JSC scheme.

### 2.1. Markov Sources

We consider binary sources with memory such that the stationary output sequence follows a time-invariant, HMM with  $\lambda$  states  $\{S_1, \dots, S_\lambda\}$ . We denote the states of the source at time  $k$  as  $q_k$ . Complete specification of a binary HMM requires the specification of three probability measures,  $A$ ,  $B$ , and  $\pi$ , defined as:

- $A = [a_{ij}]$  is the state transition probability matrix of dimension  $\lambda \times \lambda$ , with  $a_{ij}$  the probability of transition from state  $S_i$  to state  $S_j$ , i.e.,  $a_{ij} = P(q_{k+1} = S_j | q_k = S_i)$  for all  $k$ .
- $B = [b_j(v)]$  is the observation symbol probability matrix, with  $b_j(v)$  the probability of getting in the binary symbol  $v$  in state  $S_j$ , i.e.,  $b_j(v) = P(v | S_j)$ ,  $1 \leq j \leq \lambda$ ,  $v \in \{0, 1\}$ .
- $\pi$  is the initial state distribution vector, with  $\pi_j$  the probability for the initial state to be  $S_j$ , i.e.,  $\pi_j = P(q_1 = S_j)$ ,  $1 \leq j \leq \lambda$ .

**Remark 1.** For stationary sources,  $\pi$  should be taken as the stationary distribution of the chain, i.e.,  $\pi = A\pi$ .

**Remark 2.** When matrix  $B$  has entries 0 and 1, the HMM reduces to a MC.

## 2.2. Burrows-Wheeler Transform (BWT)

The BWT [19] is a lexicographical permutation of the characters of a string such that the transformed sequence is easier to be compressed. It is obtained from the last column of an array whose rows are all cyclic shifts from the input in dictionary order, which tend to have long runs of identical characters. From this last string we can recover the entire array, making the BWT reversible. The BWT has been widely analyzed in [20–22] and employed for the general problem of data compression [23,24]. More recent contributions have focused on the applicability of the BWT to coded transmission of Markov sources through AWGN channels via LDPC [13] and non-systematic Turbo codes [14].

Let  $T = \{T_k\}_{k=1}^K$ ,  $T_k \in \{0, 1\}$  denote the output block of the reversible block-sorting BWT when its input is the block of binary source symbols  $\{U_k\}_{k=1}^K$ . For sources modeled by MCs with  $\lambda$  states, it was shown in [20] that the joint probability mass function,  $P_T(t)$ , of the random block  $T$  is approximately memoryless and piecewise stationary, in the sense that there exist  $\lambda$  index sets,  $\mathcal{L}_i = \{w_{i-1} \dots w_i\}$ ,  $i = 1, \dots, \lambda$  with  $w_0 = 1$  and  $w_\lambda = K + 1$ , and a probability distribution

$$Q_T(t) = \prod_{i=1}^{\lambda} \prod_{k=w_{i-1}}^{w_i-1} Q_i(t_k) \quad (2)$$

such that the normalized divergence between both distributions can be made arbitrarily small for sufficiently large  $K$ , i.e.,

$$\frac{1}{K} D(Q_T(t) \| P_T(t)) \rightarrow 0 \quad (3)$$

as  $K \rightarrow \infty$ .

As the block length  $K$  goes to infinity, the normalized length of the index set in expression (2) converges to  $c_i \in \mathbb{R}$ , i.e.,  $\lim_{K \rightarrow \infty} \frac{|\mathcal{L}_i|}{K} = c_i$ .

**Definition 1.** Let  $T_i$  denote the binary random sequence of length  $K_i = c_i K$  at the output of the BWT corresponding to the index set  $\mathcal{L}_i$ ,  $i = 1, \dots, \lambda$ . That is,  $T_i = \{T_k\}_{k \in \mathcal{L}_i}$ .

Observe from (2) that for large blocks of length  $K$ , the binary random symbols  $T_k \in T_i$ , with  $k \in \mathcal{L}_i$ , can be considered independent and identical distributed (i.i.d.), with probability distribution

$$Q_i(t_k) \triangleq \begin{cases} p_i^0 & \text{if } t_k = 0 \\ p_i^1 = 1 - p_i^0 & \text{if } t_k = 1 \end{cases} \quad (4)$$

for some  $p_i^0 \in (0, 1)$ . These approximations should be understood under the convergence criterium (3).

Therefore, we will model the non-stationary BWT output sequence  $T$  as the concatenation of  $\lambda$  blocks of length  $K_i = c_i K$ ,  $i = 1, \dots, \lambda$  generated by  $\lambda$  independent DMS binary sources  $S_1, S_2, \dots, S_\lambda$ , with entropies

$$H_i = -p_i^0 \log p_i^0 - (1 - p_i^0) \log(1 - p_i^0), \quad i = 1, 2, \dots, \lambda.$$

By the independence of the sources and their symbols, the entropy rate of the original source can be expressed as

$$\mathcal{H}(S) = \sum_{i=1}^{\lambda} \frac{K_i}{K} H_i = \sum_{i=1}^{\lambda} c_i H_i. \quad (5)$$

### 2.3. Parallel RCM-LDGM Codes

The  $N$ -length codeword of a parallel concatenation of RCM and LDGM  $\mathbf{x}$ , is composed of  $M$  RCM coded symbols and  $I = N - M$  LDGM coded bits. Next, we provide a succinct overview of the constituent RCM and LDGM codes.

#### 2.3.1. Rate-Compatible Modulation (RCM) Codes

RCM codes [25] are based on random projections which generate multilevel symbols from weighted linear combinations of the source binary symbols. More precisely, an RCM code of rate  $K/M$  is generated by an  $M \times K$  sparse mapping matrix  $G$ . The non-zero entries of each row of  $G$  belong to a multiset  $\pm\mathcal{D}$ , with  $\mathcal{D} \subset \mathbb{N}$ , the set of natural numbers (positive integers). Given the binary source sequence  $\mathbf{u} = \{u_1, u_2, \dots, u_K\}$ , the RCM coded sequence  $\mathbf{c}$  of length  $M$  is obtained as

$$\mathbf{c} = [c_1, c_2, \dots, c_M] = G\mathbf{u}$$

where these operations are in the real field. Finally, rate adaptation is achieved by adjusting the number of rows in  $G$ .

#### 2.3.2. Low-Density Generator Matrix (LDGM) Codes

LDGM codes are a subclass of the well-known LDPC codes with the particularity that the generator matrix  $G_L$  is also sparse. This allows the decoding algorithm to use the graph generated by  $G_L$ . In this paper, we consider systematic LDGM codes, whose generator matrix is of the form  $G_L = [I_K | P]$ , where  $I_K$  is the identity matrix of size  $K$  and  $P$  is a regular  $K \times I$  sparse matrix with  $d_{LDGM}^{(v)}$  non-zero elements in each column. The LDGM coded sequence  $\mathbf{c}$  of length  $N = K + I$  is obtained as

$$\mathbf{c} = [c_1, c_2, \dots, c_N] = \mathbf{u}^T G_L = \mathbf{u}^T [I_K | P] = [u_1, u_2, \dots, u_K, x_1, x_2, \dots, x_I],$$

where  $\mathbf{u} = \{u_1, u_2, \dots, u_K\}$  is the binary source sequence to be transmitted and the operations are in the binary field. Unlike general LDPC codes, LDGM codes suffer from high error floor [26]. However, it has been shown that they can help to lower the error floor of other codes as explained next.

#### 2.3.3. Parallel RCM-LDGM Code

Consider an RCM code of rate  $K/M$  generated by a matrix  $G$ , and the non-systematic part of a high rate binary regular LDGM of rate  $K/I$ , generated by  $P$ . Then, the parallel RCM-LDGM coded sequence  $\mathbf{x}$  of length  $M + I$  is given by

$$\mathbf{x}^T = \left[ (G\mathbf{u})^T \mid 2 \cdot \left( (\mathbf{u}^T P \bmod 2) - \frac{1}{2} \right) \right],$$

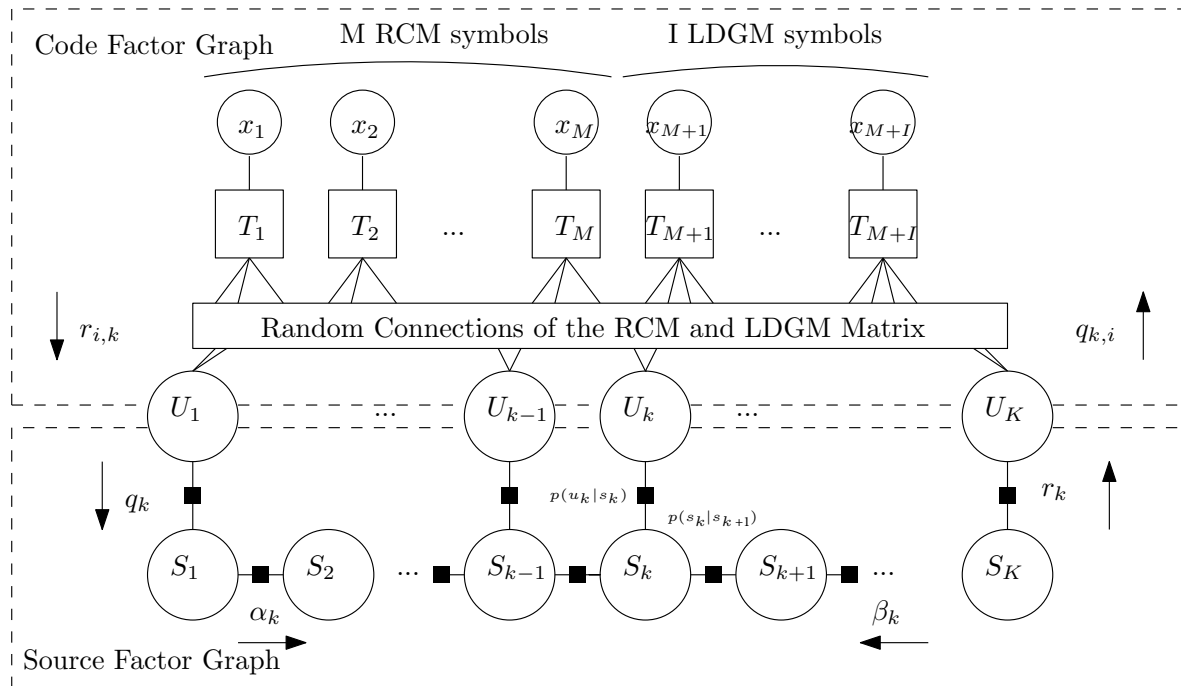
where the last  $I$  symbols are encoded using a BPSK modulator. Recall that the objective of the LDGM code is to correct the residual error of the RCM code, lowering the error floor but without degrading the RCM waterfall region.

Finally, the coded symbols of  $\mathbf{x}$  are grouped two by two and transmitted using a QAM modulator, so that the spectral efficiency,  $\rho$ , is

$$\rho = \frac{2 \cdot K}{M + I}$$

binary source symbols per complex channel use.

The performance of RCM-LDGM codes when encoding uniform and non-uniform DMSs can be found in [17,18]. An efficient way to design these codes was shown in [27]. However, no results have been found in the literature regarding the use of parallel RCM-LDGM codes to encode discrete binary sources with memory. The conventional approach in this situation would be to encode the correlated source symbols at the transmitter by the RCM-LDGM encoder, and to modify the decoder at the receiver to exploit the correlation of the source. This may be done by incorporating the factor graph that models the source into the factor graph of the RCM-LDGM code, and running the sum-product algorithm [28] over the whole factor graph represented in Figure 1. We will denote this approach as NON-BWT-JSC, and we will compare it with our proposed coding scheme defined in the next section.



**Figure 1.** Factor graph of the parallel RCM-LDGM code incorporating the factor graph modeling the source.

### 3. Proposed BTW-JSC Scheme

The main idea behind the proposed BWT-JSC scheme is to transform the original source with memory  $S$ , into a set of  $\lambda$  independent non-uniform memoryless binary sources. This is accomplished by partitioning the source sequence into blocks of length  $K$ ,  $\mathbf{U}^{(l)} = \{U_{l \cdot K + k}\}_{k=1}^K$ ,  $l \in \mathbb{N}$ , and then applying the BWT to each of these blocks. The corresponding output segment  $i$ , inside output block  $l$ , is given by

$$\mathbf{T}_i^{(l)} = \{T_{l \cdot K + k}\}_{k=w_{i-1}}^{w_i}.$$

Observe that the sequence blocks  $\mathbf{T}_i^{(l)}$ ,  $i = 0, 1, \dots, \lambda$  can be considered to have been generated by a non-uniform DMS with entropy  $H_i$ ,  $i = 1, 2, \dots, \lambda$ . Therefore, we have reduced the encoding problem of sources with memory to a simpler one, namely the problem of JSC coding of non-uniform *memoryless* binary sources, with entropies  $H_i$ . Notice that the previously mentioned RCM-LDGM high-throughput, JSC codes for non-uniform DMS sources [17], can now be applied to each of the  $\lambda$  independent sources as shown in Figure 2.

More concretely, let us consider a source with memory,  $S$ , and with entropy rate  $\mathcal{H}(S)$ , which generates blocks of  $K$  binary symbols to be transmitted at rate  $R = K/N$  by the parallel JSC coding system of Figure 2. Let  $\mathbf{T}_i$  (refer to Definition 1) be the input sequence to the corresponding  $i$ -JSC code

of rate  $R_i = K_i / N_i$ , under the constraint  $N = \sum_{i=1}^{\lambda} N_i$ . Denote by  $\{\text{SNR}_i\}_{i=1}^{\lambda}$  the set of signal-to-noise ratios allocated to each parallel channel. Define by

$$\overline{\text{SNR}} = \sum_{i=1}^{\lambda} \frac{N_i}{N} \text{SNR}_i$$

the average SNR over all parallel channels. The following Theorem proves that the proposed scheme achieves the Shannon limit.

**Theorem 1.** *Given a target rate  $R$ , the minimum overall  $\overline{\text{SNR}}$  in the coding scheme of Figure 2 is achieved when all the  $\text{SNR}_i$ 's take the same value, given by the SNR Shannon limit from expression (1), i.e.,  $\text{SNR}_i^* = 2^{R\mathcal{H}(\mathcal{S})} - 1$ . The individual rates  $R_i$  are given by  $R_i = \frac{R\mathcal{H}(\mathcal{S})}{H_i}$ ,  $i = 1, \dots, \lambda$ .*

**Proof.** Given a set of signal-to-noise ratios  $\{\text{SNR}_i\}_{i=1}^{\lambda}$ , the rates of the JSC encoders in Figure 2 are given by the Shannon's separation theorem as

$$R_i = \frac{K_i}{N_i} = \frac{C(\text{SNR}_i)}{H_i}, \quad i = 1, \dots, \lambda,$$

where by the BWT hypothesis,  $K = \sum_{i=1}^{\lambda} K_i$ .

We seek to minimize the average signal-to-noise ratio  $\overline{\text{SNR}}$  over all the  $\lambda$  parallel AWGN channels, i.e.,

$$\overline{\text{SNR}} = \sum_{i=1}^{\lambda} \frac{N_i}{\sum_{j=1}^{\lambda} N_j} \text{SNR}_i = \sum_{i=1}^{\lambda} \frac{K_i}{N} \frac{H_i}{C(\text{SNR}_i)} \text{SNR}_i,$$

under the constraint of achieving a rate

$$R = \frac{K}{N} = \sum_{i=1}^{\lambda} \frac{K_i}{\sum_{j=1}^{\lambda} N_j}.$$

Please note that since  $K = \sum_{i=1}^{\lambda} K_i$  is fixed, the constraint in  $R$  reduces to the constraint

$$N = \sum_{j=1}^{\lambda} N_j = \sum_{j=1}^{\lambda} \frac{H_j K_j}{C(\text{SNR}_j)}. \quad (6)$$

By applying the Lagrange multipliers method, we define  $F$  as

$$F = \sum_{i=1}^{\lambda} \frac{K_i H_i}{N} \frac{\text{SNR}_i}{C(\text{SNR}_i)} + \gamma \left( \sum_{i=1}^{\lambda} \frac{K_i H_i}{C(\text{SNR}_i)} - N \right),$$

and by searching for an extreme of  $F$ , we obtain that the optimal  $\text{SNR}_i^*$  are all equal to some value  $\Gamma$ . Therefore, from constraint (6)

$$N = \sum_{i=1}^{\lambda} N_i = \sum_i \frac{K_i H_i}{C(\Gamma)} = \frac{K\mathcal{H}(\mathcal{S})}{C(\Gamma)},$$

where the last equality follows from expression (5). Thus, the rate can be written as

$$R = \frac{K}{N} = \frac{C(\Gamma)}{\mathcal{H}(\mathcal{S})}.$$

Consequently, the value of  $\Gamma$  is given by the signal-to-noise ratio required to achieve the same rate  $R$  in the standard point-to-point communications system. That is,

$$\Gamma = 2^{RH(S)} - 1.$$

We conclude that

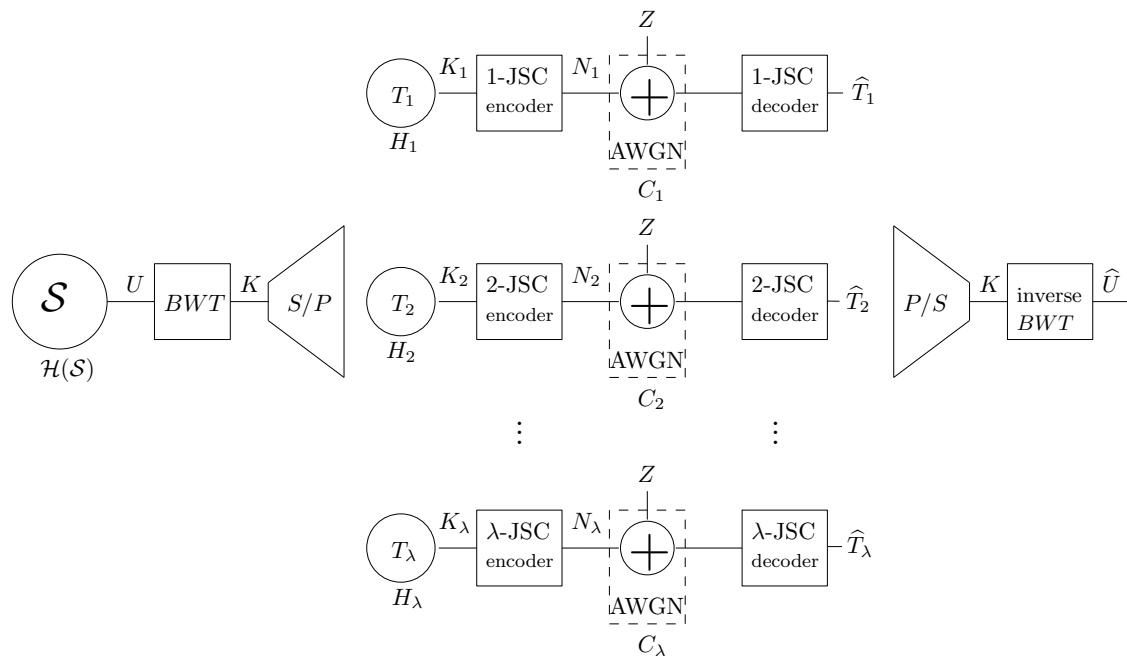
$$\overline{SNR}^* = SNR_i^* = 2^{RH(S)} - 1$$

and

$$R_i = \frac{K_i}{N_i} = \frac{C(\overline{SNR}^*)}{H_i} = \frac{RH(S)}{H_i}. \quad (7)$$

□

**Remark 3.** Observe that the BWT-JSC is asymptotically optimal in the sense that can achieve the SNR Shannon limit given by the Separation Theorem.



**Figure 2.** BWT-based proposed communication system. Please note that  $K = \sum_{i=1}^{\lambda} T_i$ .

## 4. Results

In this section, we evaluate the proposed scheme, comparing its performance with the conventional NON-BWT-JSC approach described in Section 2.3, which is based on a single code. Without any loss of generality, the spectral efficiency of the communication system has been set to 7.4 binary source symbols per complex channel use, and the source block length to  $K = 37,000$ . Thus, the total number of coded symbols at the output of the JSC encoder is  $N = 10,000$ . We begin by specifying the Markov sources used in the simulations.

### 4.1. Simulated Sources and Their Output Probability Profile

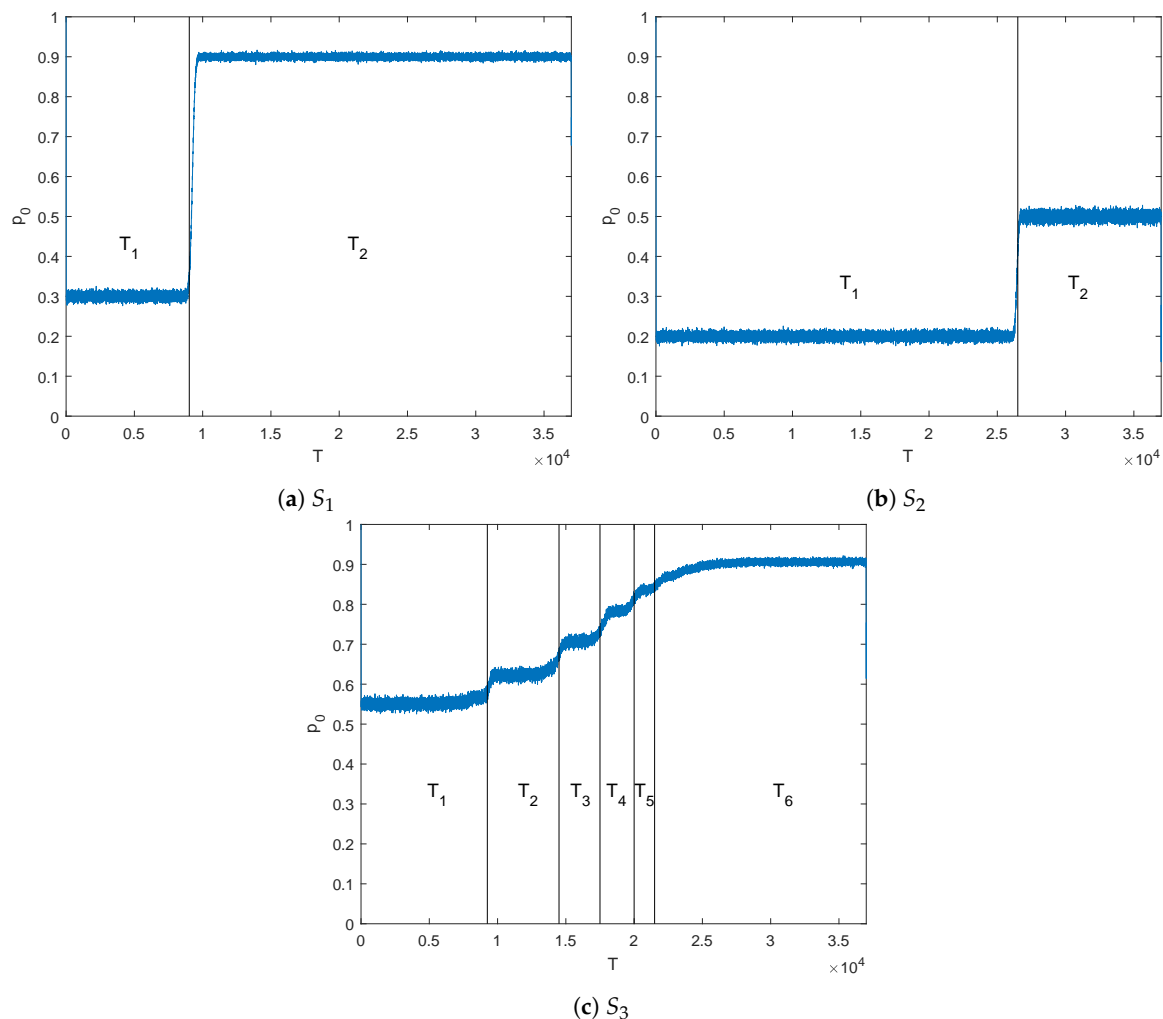
Three different 2-state ( $\lambda = 2$ ) Markov sources have been chosen. Two are modeled by MCs, with entropy rates 0.57 and 0.80 bits per source symbol, whereas the third is modeled by a HMM with entropy rate 0.73. For the sake of notation, they will be referred as  $S_1$ ,  $S_2$  and  $S_3$ . Table 1 summarizes their corresponding Markov parameters.



**Table 1.** Markov Source Parameters.

Source	Matrix A	Matrix B	Vector $\pi$	Entropy $\mathcal{H}$
$S_1$	$a_{11} = 0.90$ $a_{22} = 0.70$	$b_{11} = 1.0$ $b_{22} = 1.0$	$[0.75 \ 0.25]$	0.57
$S_2$	$a_{11} = 0.80$ $a_{22} = 0.50$	$b_{11} = 1.0$ $b_{22} = 1.0$	$[0.71 \ 0.29]$	0.8
$S_3$	$a_{11} = 0.90$ $a_{22} = 0.90$	$b_{11} = 0.5$ $b_{22} = 0.995$	$[0.5 \ 0.5]$	0.73

Figure 3 shows the probability mass function  $P_T(t)$  (refer to (4)) of the binary random block  $T$  of length  $K = 37,000$  at the output of the BWT for sources  $S_1$ ,  $S_2$ , and  $S_3$ . Observe that due to the fact that sources  $S_1$  and  $S_2$  follow a 2-state MC behavior, the BWT will produce approximately two i.i.d. segment  $T_1$  and  $T_2$ . This is clearly shown in Figure 3a,b, with segments of length ( $K_1 = 9020$ ,  $K_2 = 27,980$ ) with first order probabilities ( $p_0^{(1)} \approx 0.3$ ,  $p_0^{(2)} \approx 0.9$ ) for  $S_1$  and ( $K_1 = 9020$ ,  $K_2 = 27,980$ ) with probabilities ( $p_0^{(1)} \approx 0.2$ ,  $p_0^{(2)} \approx 0.5$ ) for  $S_2$ . On the contrary, the source  $S_3$  is characterized by a 2-state hidden Markov model, and the hidden property has the effect of increasing the number of states, should the HMM source be approximated by a pure MC. This is observed in Figure 3c, where a 6-state MC source will fairly approximate the statistics of source  $S_3$ . The partition into 6 segments has been decided by the authors based on significant change in the a priori probability of the bits forming the segments. In this case, the first order probabilities of segments  $T_1 - T_6$  of sizes ( $K_1 = 9250$ ,  $K_2 = 5250$ ,  $K_3 = 3000$ ,  $K_4 = 2500$ ,  $K_5 = 1500$ ,  $K_6 = 15,500$ ) are given by  $p_0^{(1)} \approx 0.55$ ,  $p_0^{(2)} \approx 0.63$ ,  $p_0^{(3)} \approx 0.71$ ,  $p_0^{(4)} \approx 0.78$ ,  $p_0^{(5)} \approx 0.84$  and  $p_0^{(6)} \approx 0.9$ .



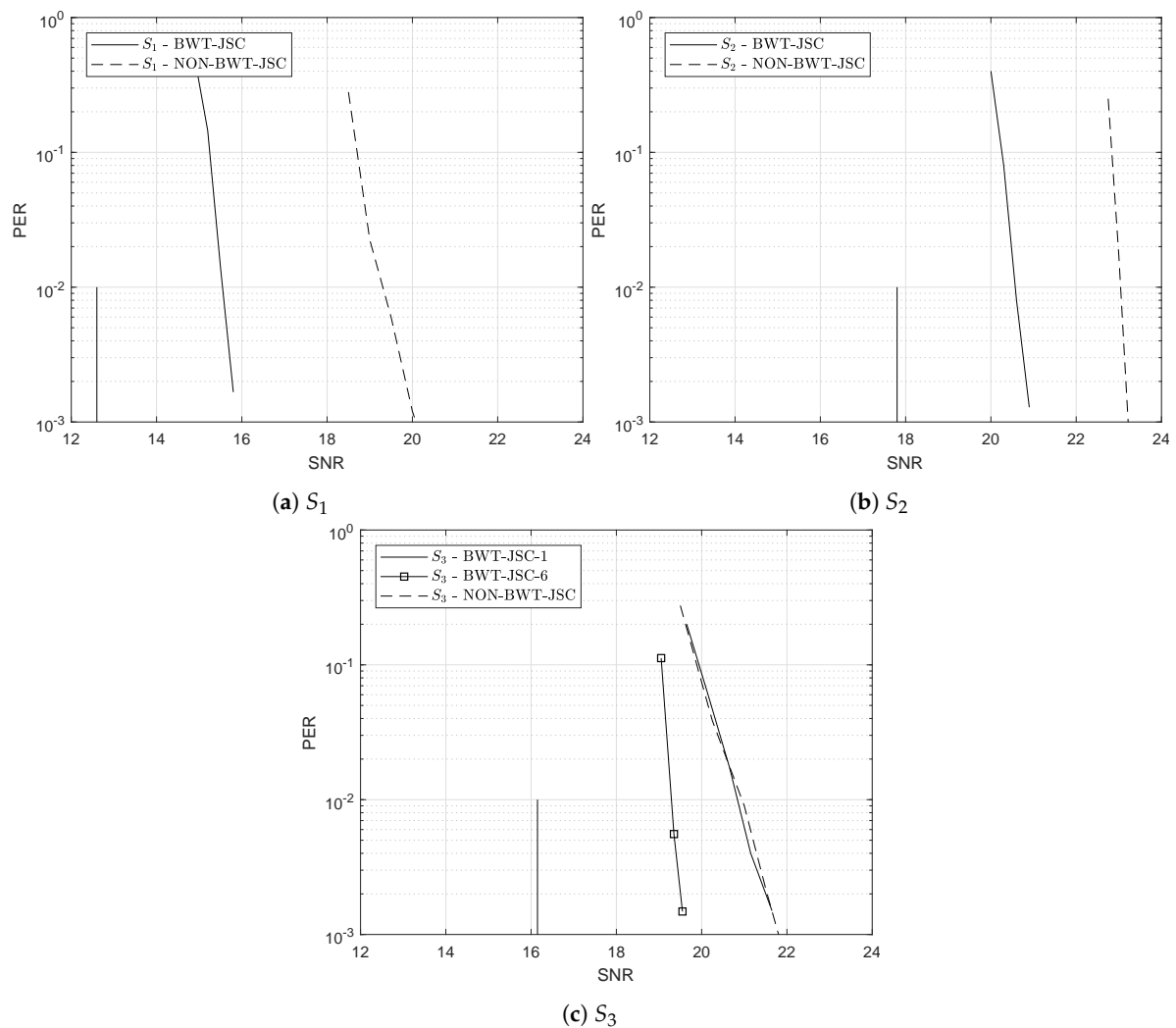
**Figure 3.** First order probability profiles of the output blocks of the BWT for example sources (a)  $S_1$ , (b)  $S_2$  and (c)  $S_3$ .



## 4.2. Numerical Results

In this section, we present the results obtained by Monte Carlo simulation for the proposed BWT-JSC and the conventional NON-BWT-JSC coding schemes. Observe that due to the BWT block, in our proposed scheme a single error at the output of the decoders will be propagated after applying the inverse-BWT. Therefore, to make a fair comparison, the results are presented in the form of Packet Error Rate (PER) versus SNR. It should be mentioned that for the correct recovery of the original transmitted source block, the inverse-BWT at the receiver side needs to know the exact position where the original End of File symbol has been moved by the BWT at the transmitted side. Therefore, this additional information should also be transmitted. Please note that for a 37,000 block length, this position can be addressed by adding 16 binary symbols. In this work, we have considered this rate loss as negligible, but in real scenarios it must be taken into account.

Figure 4 shows the PER vs SNR curves obtained by simulations for the example sources (a)  $S_1$ , (b)  $S_2$  and (c)  $S_3$  when using both the proposed system (BWT-JSC) and the conventional approach (NON-BWT-JSC) as a reference. In the proposed scheme, as stated in Section 3, after performing the BWT, each of the resulting  $\lambda$  independent non-uniform i.i.d. segments  $T_i(p_0^{(i)})$   $i = 1, \dots, \lambda$  (refer to Figure 3), are encoded by  $\lambda$  separated RCM-LDGM JSC codes of rates  $R_i$  as given by Theorem 1. The codes used for each DMS in the BWT-JSC approach, as well as the one used in the conventional NON-BWT-JSC scheme are summarized in Table 2.



**Figure 4.** Obtained PER vs SNR curves for the NON-JSC-BWT and JSC-BWT schemes when sources (a)  $S_1$ ; (b)  $S_2$  and (c)  $S_3$  are considered. The corresponding Shannon limits are plotted in vertical lines.

Observe from Figure 4a,b that for sources  $S_1$  and  $S_2$ , represented by a MC, our BWT-JSC scheme outperforms the NON-BWT-JSC approach by about 4.2 and 2.3 dB's, respectively. The reason behind this improvement lies in the fact that in the NON-BWT-JSC system, the Factor Graph (FG) of the decoder, results from a parallel concatenation of two sub-graphs: The RCM-LDGM code and MC source sub-graphs (refer to Figure 1). Consequently, in the overall FG decoder cycles between both sub-graphs appear, degrading in this way the performance of sum-product algorithm. However, in the proposed scheme, these cycles do not occur since in this case the sources are memoryless and non-uniform. The contribution of the sources sub-graphs is just to introduce the a priori probabilities of the non-uniform sources into the variable nodes of the corresponding RCM-LDGM factor sub-graphs.

Let us now consider the HMM source  $S_3$  with entropy rate  $\mathcal{H}(S_3) = 0.73$  and output probability profile as shown in Figure 3c. Note from this figure that the BW transformation of source  $S_3$  can be approximated by 6 memoryless non-uniform sources  $\{T_i\}_{i=1}^6$ , with blocks of lengths  $K_1 \approx 9250, K_2 \approx 5250, K_3 \approx 3000, K_4 \approx 2500, K_5 \approx 1500, K_6 \approx 15,500$ . Some of these blocks have short lengths, which is detrimental for the performance of the corresponding RCM-LDGM codes. To solve this problem, we build larger segments  $\tilde{T}_i$  that keep the same statistical properties as previous segments. In this approach, named BWT-JSC- $\kappa$ , we put together  $\kappa$  consecutive output blocks of the BWT to form the new segments as  $\tilde{T}_i^{(l)} = \{T_i^{(l \cdot \kappa)}, \dots, T_i^{(l \cdot (\kappa+1) - 1)}\}$  for  $i = 0, 1, \dots, \lambda$  and  $l \in \mathbb{N}$ . This is, in fact, similar to applying the BWT to source blocks of length  $\kappa \cdot K$ , but computationally it is more efficient. The RCM-LDGM codes used to transmit these segments have the same rate as before, but in this case their input and output block lengths are scaled by  $\kappa$ , i.e.,  $\tilde{K}_i = \kappa \cdot K_i, \tilde{M}_i = \kappa \cdot M_i$  and  $\tilde{I}_i = \kappa \cdot I_i, i = 0, 1, \dots, \lambda$ , respectively.

**Table 2.** Design parameters (refer to Section 2.3) used for sources (a)  $S_1$ , (b)  $S_2$  and (c)  $S_3$ .

		$\mathcal{D}$	$\mathbf{M}$	$\mathbf{I}$	$d_{LDGM}^{(v)}$
BWT-JSC	$K_1 = 9020$	$\{1, 1, 1, 2, 2\}$	3705	55	5
	$K_2 = 27980$	$\{1, 1, 1, 1, 2, 2, 2, 2\}$	6140	100	5
NON-BWT-JSC	$K = 37000$	$\{2, 2, 3, 3, 4, 7\}$	9860	140	5
$S_1$					
		$\mathcal{D}$	$\mathbf{M}$	$\mathbf{I}$	$d_{LDGM}^{(v)}$
BWT-JSC	$K_1 = 26500$	$\{2, 2, 3, 3, 4, 8\}$	6310	110	5
	$K_2 = 10500$	$\{2, 3, 4, 7\}$	3490	90	5
NON-BWT-JSC	$K = 37000$	$\{2, 3, 4, 4, 7\}$	9940	60	3
$S_2$					
		$\mathcal{D}$	$\mathbf{M}$	$\mathbf{I}$	$d_{LDGM}^{(v)}$
BWT-JSC- $\kappa$	$K_1 = 9250$	$\{2, 3, 4, 4, 7\}$	3376	29	7
	$K_2 = 5250$	$\{2, 3, 4, 4, 7\}$	1839	19	7
	$K_3 = 3000$	$\{2, 3, 4, 4, 7\}$	935	37	7
	$K_4 = 2500$	$\{2, 3, 4, 4, 7\}$	673	35	6
	$K_5 = 1500$	$\{2, 2, 3, 3, 4, 7\}$	351	18	6
	$K_6 = 15500$	$\{2, 2, 2, 3, 3, 4, 4, 7, 7\}$	2632	56	6
NON-BWT-JSC	$K = 37000$	$\{2, 2, 3, 3, 4, 8\}$	9880	120	3
$S_3$					

As before, Figure 4c plots the PER versus SNR curves for both strategies BWT-JSC (solid curves) and NON-BWT-JSC (dashed curves). When plotting the performance of the BWT-JSC- $\kappa$  approach, two different cases have been considered,  $\kappa = 1$  and  $\kappa = 6$ . Please note that when  $\kappa = 1$  the scheme

is the same as in previous MC examples. On the other hand, by concatenating 6 consecutive BWT output segments ( $\kappa = 6$ ), we force the length of smallest segment to be 9000. Notice that for  $\kappa = 6$  the proposed scheme outperforms the conventional approach by 2.3 dB. However, due to the bad performance of the short block-length RCM-LDGM codes, when  $\kappa = 1$  the performance is similar to that of the conventional approach. This clearly shows that by concatenating BWT segments the system performance improves thanks to the avoidance of blocks with short lengths.

As summarized in Table 3, the proposed scheme clearly outperforms the conventional approach, and the PER vs SNR curves are only about 3 dB away from the Shannon limits.

**Table 3.** Summary of numerical results. Labels BWT-JSC and NON-BWT-JSC represent the SNR required for a PER of  $10^{-3}$  with each scheme.

	Entropy Rate	Shannon Limit	BWT-JSC( $-\kappa$ )		NON-BWT-JSC
$S_1$	0.57	12.57 dB	15.8 dB		20 dB
$S_2$	0.80	17.78 dB	20.9 dB		23.25 dB
			1	6	
$S_3$	0.73	16.15 dB	21.8 dB	19.55 dB	21.8 dB

## 5. Conclusions

A new source-controlled coding scheme for high-throughput transmission of binary sources with memory over AWGN channels has been proposed. The proposed strategy is based on the concatenation of the BWT with rate-compatible RCM-LDGM codes. The BWT transforms the original source with memory into a set of independent non-uniform discrete memoryless binary sources, which are then separately encoded, with optimal rates, using RCM-LDGM codes. Simulations show that the proposed scheme outperforms the traditional strategy of using the FG of the source in the decoding process by up to 4.2 dB for a spectral efficiency of 7.4 binary source symbols per complex channel use and a source with entropy rate 0.57 bits per source symbol. The resulting performance lies within 3 dB of the Shannon limit.

**Author Contributions:** Funding acquisition, P.M.C.; Investigation, I.G.; Software, I.G.; Supervision, P.M.C. and J.G.-F.; Writing—original draft, I.G. and P.M.C.; Writing—review & editing, J.G.-F.

**Funding:** This work was supported in part by the Spanish Ministry of Economy and Competitiveness through the CARMEN project (TEC2016-75067-C4-3-R) and by NSF Award CCF-1618653.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

BWT	Burrows-Wheeler Transform
DMS	Discrete Memoryless Source
JSC	Joint Source-Channel
MC	Markov Chain
HMM	Hidden Markov Model
RCM	Rate-Compatible Modulation
LDGM	Low-Density Generator Matrix
LDPC	Low-Density Parity Check
AWGN	Additive White Gaussian Noise
QAM	Quadrature Amplitude Modulation
SNR	Signal-to-Noise Ratio
PER	Paquet Error Rate

## References

1. Sayood, K.; Borkenhagen, J.C. Use of residual redundancy in the design of joint source/channel coders. *IEEE Trans. Commun.* **1991**, *39*, 838–846. [\[CrossRef\]](#)
2. Ramzan, N.; Wan, S.; Izquierdo, E. Joint source-channel coding for wavelet-based scalable video transmission using an adaptive turbo code. *J. Image Video Process.* **2007**, *2007*, 047517. [\[CrossRef\]](#)
3. Zhonghui, M.; Lenan, W. Joint source-channel decoding of Huffman codes with LDPC codes. *J. Electron. China* **2006**, *23*, 806–809.
4. Pu, L.; Wu, Z.; Bilgin, A.; Marcellin, M.W.; Vasic, B. LDPC-based iterative joint source-channel decoding for JPEG2000. *IEEE Trans. Image Process.* **2007**, *16*, 577–581. [\[CrossRef\]](#)
5. Ordentlich, E.; Seroussi, G.; Verdu, S.; Viswanathan, K. Universal algorithms for channel decoding of uncompressed sources. *IEEE Trans. Inf. Theory* **2008**, *54*, 2243–2262. [\[CrossRef\]](#)
6. Hindelang, T.; Hagenauer, J.; Heinen, S. Source-controlled channel decoding: Estimation of correlated parameters. In Proceedings of the 3rd International ITG Conference on Source and Channel Coding, Munich, Germany, 17–19 January 2000.
7. Garcia-Frias, J.; Villasenor, J.D. Combining hidden Markov source models and parallel concatenated codes. *IEEE Commun. Lett.* **1997**, *1*, 111–113. [\[CrossRef\]](#)
8. Garcia-Frias, J.; Villasenor, J.D. Joint turbo decoding and estimation of hidden Markov sources. *IEEE J. Sel. Areas Commun.* **2001**, *19*, 1671–1679. [\[CrossRef\]](#)
9. Shamir, G.I.; Xie, K. Universal lossless source controlled channel decoding for i.i.d. sequences. *IEEE Commun. Lett.* **2005**, *9*, 450–452. [\[CrossRef\]](#)
10. Yin, L.; Lu, J.; Wu, Y. LDPC based joint source-channel coding scheme for multimedia communications. In Proceedings of the 8th International Conference Communication System (ICCS), Singapore, 28 November 2002; Volume 1, pp. 337–341.
11. Yin, L.; Lu, J.; Wu, Y. Combined hidden Markov source estimation and low-density parity-check coding: A novel joint source-channel coding scheme for multimedia communications. *Wirel. Commun. Mob. Comput.* **2002**, *2*, 643–650. [\[CrossRef\]](#)
12. Ordentlich, E.; Seroussi, G.; Verdu, S.; Viswanathan, K.; Weinberger, M.J.; Weissman, T. Channel decoding of systematically encoded unknown redundant sources. In Proceedings of the International Symposium on Information Theory, Chicago, IL, USA, 27 June–2 July 2004.
13. Shamir, G.I.; Wang, L. Context decoding of low density parity check codes. In Proceedings of the 2005 Conference on Information Sciences and Systems, New Orleans, LA, USA, 14–17 March 2005; Volume 62, pp. 597–609.
14. Xie, K.; Shamir, G.I. Context and denoising based decoding of non-systematic turbo codes for redundant data. In Proceedings of the International Symposium on Information Theory, Adelaide, Australia, 4–9 September 2005; pp. 1280–1284.
15. Del Ser, J.; Crespo, P.M.; Esnaola, I.; Garcia-Frias, J. Joint Source-Channel Coding of Sources with Memory using Turbo Codes and the Burrows-Wheeler Transform. *IEEE Trans. Commun.* **2010**, *58*, 1984–1992. [\[CrossRef\]](#)
16. Zhu, G.-C.; Alajaji, F. Joint source-channel turbo coding for binary Markov sources. *IEEE Trans. Wirel. Commun.* **2006**, *5*, 1065–1075.
17. Li, L.; Garcia-Frias, J. Hybrid Analog Digital Coding Scheme Based on Parallel Concatenation of Liner Random Projections and LDGM Codes. In Proceedings of the 2014 48th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 19–21 March 2014.
18. Li, L.; Garcia-Frias, J. Hybrid Analog-Digital Coding for Nonuniform Memoryless Sources. In Proceedings of the 2015 49th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 18–20 March 2015.
19. Burrows, M.; Wheeler, D. *A Block Sorting Lossless Data Compression Algorithm*; Research Report 124; Digital Systems Center: Gdańsk, Poland, 1994.
20. Visweswariah, K.; Kulkarni, S.; Verdú, S. Output Distribution of the Burrows Wheeler Transform. In Proceedings of the IEEE International Symposium of Information Theory, Sorrento, Italy, 25–30 June 2000.
21. Effros, M.; Visweswariah, K.; Kulkarni, S.; Verdú, S. Universal lossless source coding with the burrows wheeler transform. *IEEE Trans. Inf. Theory* **2002**, *48*, 1061–1081. [\[CrossRef\]](#)

22. Balkenhol, B.; Kurtz, S. Universal data compression based on the burrows wheeler transform: Theory and practice. *IEEE Trans. Comput.* **2000**, *49*, 1043–1053.
23. Caire, G.; Shamai, S.; Verdú, S. Universal data compression using LDPC codes. In Proceedings of the International Symposium on Turbo Codes and Related Topics, Brest, France, 1–5 September 2003.
24. Caire, G.; Shamai, S.; Shokrollahi, A.; Verdú, S. *Fountain Codes for Lossless Data Compression*; DIMACS Series in Discrete Mathematics and Theoretical Computer Science; American Mathematical Society: Providence, RI, USA, 2005; Volume 68.
25. Cui, H.; Luo, C.; Wu, J.; Chen, C.W.; Wu, F. Compressive Coded Modulation for Seamless Rate Adaption. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 4892–4904. [[CrossRef](#)]
26. Mackay, D.J. Good Error-Correcting Codes Based on Very Sparse Matrices. *IEEE Trans. Inf. Theory* **1999**, *45*, 399–431. [[CrossRef](#)]
27. Granada, I.; Crespo, P.M.; Garcia-Frias, J. Asymptotic BER EXIT chart analysis for high rate codes based on the parallel concatenation of analog RCM and digital LDGM codes. *J. Wirel. Commun. Netw.* **2019**, *11*. [[CrossRef](#)]
28. Kschischang, F.R.; Frey, B.J.; Loeliger, H.-A. Factor Graphs and the Sum-Product Algorithm. *IEEE Trans. Inf. Theory* **2001**, *47*, 498–519. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).