

Fountain Code Design for Data Multicast with Side Information

Dino Sejdinović, Robert J. Piechocki, Angela Doufexi, and Mohamed Ismail

Abstract—Fountain codes are a robust solution for data multicasting to a large number of receivers which experience variable channel conditions and different packet loss rates. However, the standard fountain code design becomes inefficient if all receivers have access to some side information correlated with the source information. We focus our attention on the cases where the correlation of the source and side information can be modelled by a binary erasure channel (BEC) or by a binary input additive white Gaussian noise channel (BIAWGNC). We analyse the performance of fountain codes in data multicasting with side information for these cases, derive bounds on their performance and provide a fast and robust linear programming optimization framework for code parameters. We demonstrate that systematic Raptor code design can be employed as a possible solution to the problem at the cost of higher encoding/decoding complexity, as it reduces the side information scenario to a channel coding problem. However, our results also indicate that a simpler solution, non-systematic LT and Raptor codes, can be designed to perform close to the information theoretic bounds.

Index Terms—Fountain codes, distributed compression, joint source-channel coding, side information.

I. INTRODUCTION

FOUNTAIN codes are a universal, capacity-approaching Forward Error Correction (FEC) solution for data transmission over lossy packet networks. The first practical fountain codes were Luby-Transform (LT) codes [2], whereas their extension, Raptor coding [3], represents a state-of-the-art digital fountain solution for lossy transmission with excellent performance and linear encoding/decoding complexity. The property of ratelessness, i.e., the ability to adapt the code rate on-the-fly, makes fountain codes an attractive solution for data broadcast/multicast and asynchronous data access based applications where users may experience varying channel conditions and packet loss rates, such as mobile environments. It is possible to adapt fountain codes for reliable transmission over noisy channels [4]–[7], as they are more amenable to soft-decision decoding than classical erasure correcting codes, such as Reed-Solomon codes. Even though it has been demonstrated that universal code parameters generally do not exist for memoryless symmetric channel models [7], fountain codes have shown promising performance in practice over channels such as the binary input additive white Gaussian

noise channel (BIAWGNC) [7] and block-fading channels [6]. Following the advent of interest in fountain codes, the notion of fountain capacity has been introduced [8], such that the definition of rate penalizes the reception of symbols at the receiver rather than the use of the channel at the transmitter. It has been shown that for stationary memoryless channels, fountain capacity suffers no rate loss compared to Shannon capacity. In addition to channel coding, fountain codes have been used in the context of lossless data compression [9], distributed source coding [10], [11], distributed joint source-channel coding [12]–[14] and transmission of video [13], [15], [16]. A systematic version of Raptor codes is being adopted as the application layer forward error correction scheme of choice for large scale multimedia content delivery in standards such as Multimedia Broadcast/Multicast Services (MBMS) within Third Generation Partnership Project (3GPP) [17] and the IP-Datcast services within Digital Video Broadcasting (DVB) [18].

In this contribution, we consider the design of fountain codes as applied to the problem of multicast transmission, where receivers have access to a priori side information Y about the source X . A common way to model the correlation between X and Y is to view them as the input and the output, respectively, of a certain communication channel, referred to as the correlation channel or virtual channel. We will focus on the fountain code design problem for two special cases of the correlation channel: (1) *coding with partial information* - Y is a partial information about X , i.e., the output of a binary erasure channel (BEC) when X is the input; and (2) *Gaussian correlation* - $Y = X + \mathcal{N}(0, \sigma^2)$, where X is the binary information source over the alphabet $\{-1, 1\}$. As coding with side information is an instance of the problem of distributed source coding, our scenario incorporates both distributed source compression and channel coding gains as goals of the fountain code design. A systematic Raptor design is employed as a possible solution to the problem, which reduces the side information scenario to a simple channel code design problem at the cost of higher encoding/decoding complexity. Similar model has been independently reported in [10]. In addition, we present novel code design methods which utilize the presence of decoder side information and aim to provide both the distributed source compression scheme and the channel coding scheme in a single (non-systematic) fountain code. For coding with partial information, we develop a code optimization procedure which yields superior performance compared to that of [19], where a similar adaptive rateless coding scenario has been independently studied.

In section II the relevant concepts about LT and Raptor codes are reviewed. In section III, the system model and our assumptions are outlined, while section IV considers

Manuscript received August 11, 2008; revised May 5, 2009 and July 13, 2009; accepted July 14, 2009. The associate editor coordinating the review of this paper and approving it for publication was M. Valenti.

D. Sejdinović, R. J. Piechocki, and A. Doufexi are with the Centre for Communications Research, Department of Electrical and Electronic Engineering, University of Bristol, Bristol, UK (e-mail: {d.sejdinovic, r.j.piechocki, a.doufexi}@bristol.ac.uk).

M. Ismail is with the Toshiba Research Europe Ltd. Telecommunications Research Laboratory, Bristol, UK (e-mail: mohamed@toshiba-trel.com).

The material in this paper was presented in part at the IEEE International Conference on Communications, Beijing, China, May 2008 [1].

Digital Object Identifier 10.1109/TWC.2009.081076

systematic Raptor codes for coding with side information. In section V results are presented for a non-systematic fountain code design for an erasure correlation channel, whereas the Gaussian correlation case is studied in section VI. Section VII concludes the paper.

II. FOUNTAIN CODES FOR CHANNEL CODING

A. The digital fountain paradigm

The digital fountain paradigm, introduced in [20], aims to enable the transmitter to produce a potentially infinite stream of encoding symbols which are all random and equally important descriptions of the source. Ideally, such a coding apparatus would be sufficient for a reliable transmission over any channel. The receiver could collect any set of encoding symbols as long as there is a sufficient number of them to guarantee a large probability of successful decoding. A practicable construction of fountain coding schemes allows broadcasting/multicasting systems where: (a) there is no need for feedback nor retransmissions of lost encoding symbols, and (b) users benefit from the asynchronous data access.

Similarly to low-density parity-check (LDPC) codes, practical fountain codes utilize their sparse graphical representation in order to perform the decoding procedure of low computational complexity. The decoding algorithm of choice is Belief Propagation (BP) [21].

BP decoding is especially simple in the case of erasure channels and simplifies to a greedy graph pruning procedure [22]. This means that BP decoding can be performed on data packets. The existence of an information sequence containing k binary data packets, i.e., binary vectors of some fixed length, is implicit in our discussion and it is assumed that the encoding packets are produced by bitwise XOR-ing original data packets. However, in the case of soft decision BP decoding, we assume that the information sequence consists of k symbols on binary alphabet \mathbb{F}_2 . This enables us to simplify the message-passing rules of the BP algorithm in terms of log-likelihood ratios (LLR). The performance of a fountain code is typically measured in terms of the average error rate per symbol, i.e., per bit or per packet, as the function of *code overhead* ε , where $n = \frac{k}{Cap(\mathcal{C})}(1 + \varepsilon)$ is the number of encoding symbols at the receiver, k is the block length and $Cap(\mathcal{C})$ is the capacity of the transmission channel \mathcal{C} . A sequence of fountain codes approaches capacity if it attains an arbitrarily small error rate at a code overhead arbitrarily close to zero.

B. LT codes and output degree distributions

LT (Luby Transform) codes [2] are the first class of fountain codes fully realizing the digital fountain paradigm. The only two parameters of an LT code ensemble $LT(k, \Omega(x))$ are the length of information sequence k and the probability distribution Ω on the set $N_k = \{1, 2, \dots, k\}$. The distribution $\Omega(x)$ is called the output degree distribution, since it determines the degrees of the output nodes in the decoding graph. The output degree distribution is usually written in the generating polynomial notation¹ as $\Omega(x) = \sum_{d=1}^k \Omega_d x^d$,

¹Throughout the paper, symbols $\Omega(x)$, $\Phi(x)$ and $\omega(x)$, $\phi(x)$ are respectively reserved for the node perspective and their corresponding edge perspective degree distributions, $\omega(x) = \Omega'(x)/\Omega'(1)$, $\phi(x) = \Phi'(x)/\Phi'(1)$

where Ω_d is the probability that degree d is chosen. The generation of a single LT encoding packet consists of two simple steps: (a) Sample an output degree d with probability Ω_d , (b) Sample d distinct input packets uniformly at random from the information sequence and XOR them. These steps can be performed as many times as necessary in order to produce enough encoding packets for successful decoding. The decoding of an LT code utilizes a BP algorithm on the factor graph of the linear encoder $\mathbb{F}_2^k \rightarrow \mathbb{F}_2^n$ obtained by restriction of the fountain code mapping to exactly those n coordinates in the fountain encoding stream observed at the receiver.

By applying the coupon collector's problem, it can be concluded that the expected number of edges M on the decoding graph needs to grow at least as $O(k \ln k)$ in order to ensure that every input node in the decoding graph is connected to at least one output node, which is a necessary condition for successful decoding with any decoding algorithm. Luby [2] showed the existence of an output degree distribution which meets this lower bound of $O(k \ln k)$ edges, while providing a high probability of successful BP decoding at rates just below the channel capacity on erasure channels. Furthermore, the probability of successful decoding can be made arbitrarily close to 1 in the asymptotic lengths of the data source, i.e., when $k \rightarrow \infty$. Luby named such distribution the *ideal soliton* distribution.

Definition 1: The ideal soliton distribution $\Psi(x)$ on N_k is given by:

$$\Psi_i = \begin{cases} 1/k, & i = 1, \\ \frac{1}{i(i-1)}, & 2 \leq i \leq k. \end{cases} \quad (1)$$

However, the ideal soliton distribution performs poorly in practice for finite and practicable values of k . This is due to its high sensitivity: the expected number of singly connected output nodes is one at each stage of graph pruning, and whenever it becomes zero prior to decoding completion, decoding fails. Thus, Luby introduced its modification, the *robust soliton* distribution, based on another distribution

$$T_i = \begin{cases} R/(ik), & 1 \leq i \leq \frac{k}{R} - 1, \\ (R/k) \ln(R/k), & i = \frac{k}{R}, \\ 0, & \frac{k}{R} + 1 \leq i \leq k, \end{cases} \quad (2)$$

where $R = c\sqrt{k} \ln \frac{k}{\delta}$, and c and δ are suitably chosen parameters [2]. Now, after adding together distributions $\Psi(x)$ and $T(x)$ and renormalizing, we obtain the robust soliton distribution with a characteristic spike at $i = \frac{k}{R}$. Luby showed that robust soliton LT codes have low encoding-decoding complexity of the order $O(k \ln \frac{k}{\delta})$ XOR operations for reconstruction probability of $1 - \delta$ and a vanishing code overhead, as $k + O(\sqrt{k} \ln^2 k)$ encoding packets are sufficient to decode an information sequence of length k .

C. Raptor codes

In [3], Shokrollahi introduced Raptor codes, a modification of LT codes obtained by precoding the input message block by a high rate sparse graph code, and by using a *light* output degree distribution (capped at some maximum degree d_{max} and essentially independent of k). Raptor codes were shown to have excellent performance and linear encoding and decoding

times, since the number of edges in the decoding graph scales as $O(k)$. The benefit of smaller encoding and decoding complexity lies within the relaxation of the condition that *all* the input packets from the source need to be decoded at the receiver end. Instead, only a certain *fraction* of the input packets is required to be decoded and precoding provides the additional redundancy within the set of input packets. The idea is that the redundancy provided by a high rate precode, e.g., an LDPC code, should be enough to finish off the decoding procedure after a sufficiently large fraction of the input packets. The precode can be optimized for a projected fraction of undecoded packets, viewed as the erasure probability of the channel.

D. Asymptotic analysis

The structure of the decoding graph, i.e., the choice of degree distribution, determines the performance of fountain codes. Unlike irregular LDPC codes [22] which have both check node and variable node degree distributions as their design parameters, the LT code ensemble depends on the choice of the output (check) node degree distribution, whereas the input (variable) node degree by construction follows a binomial distribution on αk trials with probability $1/k$, where α is the average input degree. The average input degree of a light degree distribution stays bounded as $k \rightarrow \infty$. This can be concluded from the equality $\alpha k = \Omega'(1)(1 + \varepsilon)k$, where both sides are the number of edges M on the decoding graph. For large k the input degree distribution can be approximated by the Poisson distribution $\Lambda(x) = \exp(\alpha(x - 1))$, which means that the asymptotic behaviour of an LT ensemble with constant average degree distribution is captured by the following version of the AND-OR lemma [23]:²

Lemma 1: The packet error rate of $LT(k, \Omega(x))$ with an average input degree α which stays bounded as $k \rightarrow \infty$, converges to $y = \lim_{l \rightarrow \infty} y_l$, where y_l is given by:

$$\begin{aligned} y_0 &= 1, \\ y_l &= \exp(-\alpha\omega(1 - y_{l-1})). \end{aligned} \quad (3)$$

The following related result is due to Sanghavi [24] and it is more general compared to Lemma 1 as it applies to all LT ensembles. It was adopted from the study of hypergraph collapse [25] and it determines a necessary and sufficient condition for a sequence of ensembles $LT(k, \Omega^{(k)}(x))$ to have a vanishing error rate at code overhead ε .

Lemma 2: Let δ_k be the error rate of $LT(k, \Omega^{(k)}(x))$, where degree distributions $\Omega^{(k)}(x)$ converge pointwise to $\Omega(x)$. Then, at code overhead ε , $\delta_k \rightarrow \delta$ as $k \rightarrow \infty$, where:

$$\delta = 1 - \inf\{x \in [0, 1) : (1 + \varepsilon)\Omega'(x) + \log(1 - x) < 0\}, \quad (4)$$

where such infimum exists, and $\delta = 0$ otherwise.

When optimizing degree distributions for LT codes, Lemma 1 can be transformed into a linear programming routine. Let us for the moment fix the average input node degree $\alpha = \Omega'(1)(1 + \varepsilon)$ and minimize the code overhead such that the desired error rate δ is achieved. The code overhead can be expressed in terms of the edge perspective distribution

$\omega(x)$ as $1 + \varepsilon = \alpha \sum_{d=1}^{d_{max}} \frac{\omega_d}{d}$. The set of linear programs $LP_1(\delta, d_{max}, N)$ is given by :

$$\begin{aligned} LP_1 : \quad & \min \alpha \sum_{d=1}^{d_{max}} \frac{\omega_d}{d} \\ & \alpha \sum_{d=1}^{d_{max}} \omega_d y_i^{d-1} \geq -\ln(1 - y_i), \quad i \in 1, 2, \dots, m, \\ & \omega_d \geq 0, \quad d \in 1, 2, \dots, d_{max}, \end{aligned} \quad (5)$$

where $0 = y_1 < y_2 < \dots < y_N = 1 - \delta$ are N equidistant points on $[0, 1 - \delta]$, δ is the desired error rate, and d_{max} is the maximum degree of the degree distribution which is being optimized. The node perspective distribution $\Omega(x)$ can be determined from $\omega(x)$ as $\Omega(x) = \frac{\int_0^x \omega(z) dz}{\int_0^1 \omega(z) dz}$. Also, note that the coefficient α is in fact not a design parameter as we can allow variables ω_d to sum to an undetermined α , instead of 1. In that case, $\omega(x)$ is the unnormalized edge perspective degree distribution, and in further, we will omit α from the linear programs.

III. SIDE INFORMATION SCENARIO: SYSTEM MODEL AND ASSUMPTIONS

Slepian and Wolf [26] produced a remarkable result which states that separate compression (Slepian-Wolf Coding - SWC) of two correlated sources suffers no rate loss compared to the case of joint compression. In this contribution, we study an instance of asymmetric SWC, i.e., when one of the sources is fully known at the decoder (decoder side information). In this case, the range of achievable compression rates is given by $R > H(X|Y)$ where $H(X|Y)$ is the entropy of source X conditional on the decoder side information Y . The authors of [10] address a similar problem. However, [10] disregards the non-systematic fountain code design, whereas our main aim in this paper is to advocate non-systematic fountain codes as a sub-optimal yet attractive solution to the problem of asymmetric SWC. A related work in [11] considers using fountain codes for symmetric SWC of two correlated sources, but concentrates on the decoding algorithm rather than on the code design. An alternative approach to constructing rateless SWC schemes, which uses layered LDPC codes, rather than fountain codes, is presented in [27].

The system model we are considering is presented in Fig. 1. The binary information source X is correlated with decoder side information Y^j available at the receiver j , via some “virtual” correlation channel $\mathcal{C}_V^j = \mathcal{C}_V$, $j \in \{1, 2, \dots, r\}$, which is identical for all the receivers. This means that different receivers merely see different realizations of the same random variable Y , which is the output of \mathcal{C}_V . The encoder processes an information sequence $\mathbf{x} = (x_1, \dots, x_k)$ of length k at a time, produces the potentially infinite binary stream $\mathbf{z} = (z_1, z_2, \dots)$, $\mathbf{z} = f_{enc}(\mathbf{x})$, of the encoding symbols and multicasts this stream. The receiver j receives the stream through an “actual” transmission channel \mathcal{C}_A^j , which can differ across the set of receivers. The channel outputs are depicted as the “noisy” stream $\mathbf{w}^j = (w_1^j, w_2^j, \dots)$. The receiver j picks up any t_j channel outputs $\mathbf{w}^{j*} = (w_{i_1}^j, w_{i_2}^j, \dots, w_{i_{t_j}}^j)$ from the incoming stream of symbols, aware of their coordinates

²packet error rate of the ensemble $LT(k, \Omega(x))$ is the probability that an input packet cannot be recovered by the decoder

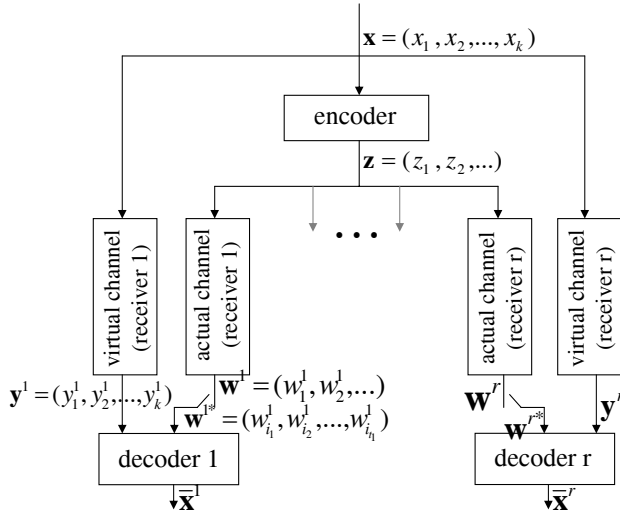


Fig. 1. Fountain coded data multicast with side information.

within a stream, where $t_j/k \geq H(X|Y)/\text{Cap}(\mathcal{C}_A^j)$, and tunes out from the multicast. By taking advantage of the side information sequence $\mathbf{y}^j = (y_1^j, \dots, y_k^j)$ the receiver j decodes: $\bar{\mathbf{x}}^j = f_{\text{dec}}(\mathbf{w}^{j*}, \mathbf{y}^j)$. Our objective is to devise the encoding strategy such that it is possible to have the rate t_j/k at the receiver j close to the optimal value, i.e., the Slepian-Wolf limit in the noisy channel case [28], $H(X|Y)/\text{Cap}(\mathcal{C}_A^j)$, and to still allow for a high probability of successful decoding, i.e., of $\bar{\mathbf{x}}^j = \mathbf{x}$, $j \in \{1, 2, \dots, r\}$.

In this contribution, we focus on two special cases of the above system model where \mathcal{C}_V is either a binary erasure channel (BEC) with erasure probability p or a binary input additive white Gaussian noise channel (BIAWGNC) with noise variance σ_V^2 . The following example motivates the study of rateless code solutions when the virtual channel is a BEC.

Example 1: A source node contains a large number k of data packets to be disseminated to a large number of receivers over lossy links. However, each receiver already knows a subset of the data packets, i.e., approximately $(1-p)k$ packets for $0 < p < 1$. Different receivers can possibly have knowledge of different packets. This could have arisen, e.g., as a result of transmission from other sources. Now, since the transmitter has no knowledge of which packets are available at which receivers, it must encode over all its packets for multicast transmission.

Ideally, a rateless code is the solution sought after for the setting outlined in the example, as it would be able to naturally adapt its rate to different or variable packet loss rates across the set of receivers. Alternatively, some kind of Hybrid-ARQ scheme could be employed, but due to a large number of receivers, feedback resources may be severely limited and this solution may lead to feedback implosion³. Clearly, each receiver must receive at least $p k$ encoding packets to successfully recover the unknown part of the message. But how close can we get to this lower bound in the multicast transmission? We will study this problem in detail in Sections IV and V.

³Feedback implosion is a common problem in broadcasting arising as different receivers may request retransmission of different data packets, which in turn may lead to the multiple transmission of all data packets.

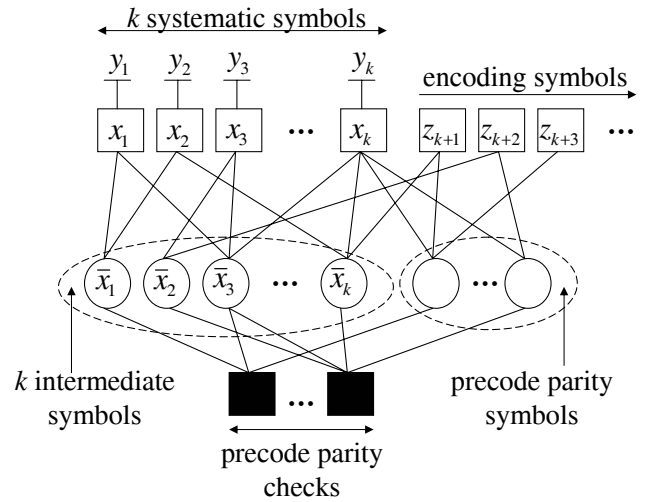


Fig. 2. Systematic Raptor for coding with partial information.

Note that the problem of coding with partial information is essentially equivalent to the problem of simultaneous dissemination of independent messages to r receivers with the same bitstream, provided that each receiver already has access to all messages intended to other receivers. Namely, if each receiver $j \in \{1, 2, \dots, r\}$ requests message a_j but already knows messages a_i , $i \neq j$, then it is decoding the source $\{a_i\}_{i=1}^r$ with partial side information $\{a_i\}_{i \neq j}$. Thus, it may be possible that practical solutions to this problem can find applications in lower complexity network coding proposals, such as [29].

IV. SYSTEMATIC RAPTOR CODING WITH DECODER SIDE INFORMATION

Standard fountain codes are non-systematic by their construction as output symbols are equally important random linear functions on the set of input symbols. However, it is possible to design a systematic fountain code at the expense of the increase of encoding/decoding complexity. Namely, it is necessary to explicitly calculate the vector of *intermediate symbols* $\bar{\mathbf{x}}$, from the input vector \mathbf{x} , such that the input vector will be replicated in the fountain encoding stream. This is performed by solving the equation

$$\mathbf{G}_{LT}^{\{1:k\}} \bar{\mathbf{x}}^T = \mathbf{x}^T, \quad (6)$$

where $\mathbf{G}_{LT}^{\{1:k\}}$ is a predetermined invertible $k \times k$ matrix used as the first k rows of the LT generator matrix. Making Raptor codes systematic also allows some interesting ways of realizing the encoding and decoding process. Namely, both processes perform two basic operations [30]: code constraints processing, which solves a set of constraint equations (6) (at the encoder) or decodes the intermediate symbols from the observed encoding symbols (at the decoder), and LT encoding, which generates the actual output stream (at the encoder) or calculates the input vector based on the intermediate symbols (at the decoder). The computation of the intermediate symbols with Gaussian elimination is generally quadratic in k , unless a special structure of matrix $\mathbf{G}_{LT}^{\{1:k\}}$ is imposed such that the linear system (6) can be solved with the direct elimination of one unknown at a time.

The possibility of a systematic fountain code seems as a natural solution for the problem of coding with partial information. Namely, our setting where both \mathcal{C}_V and \mathcal{C}_A are erasure channels can be reduced to the channel coding problem of reliable transmission over a pair of binary erasure channels, under constraint that the transmission of systematic symbols X^k through \mathcal{C}_V precedes the transmission of non-systematic symbols, and results in the decoder side information Y^k . Hence, a universal (applicable to any erasure channel) systematic fountain code would be sufficient to optimally solve the proposed problem. The application of systematic Raptor design in this setting is demonstrated in Fig. 2. The encoder needs to calculate intermediate symbols \bar{x} for each information sequence via Gaussian elimination, and then proceed with the encoding but from the $(k+1)$ -th row of the generator matrix - as the strategy consists of sending only non-systematic encoding symbols. The decoder directly embeds the decoder side information \mathbf{y} in the decoding graph. The erased symbols in \mathbf{y} are simply ignored, whereas nonerased symbols are used as the output corresponding to the systematic symbols, i.e., the first k symbols of the fountain encoding stream. Upon recovering the intermediate symbols, an additional *encoding* step is performed in order to calculate the (unknown part of the) actual information sequence \mathbf{x} by multiplying intermediate symbols \bar{x} with the first k rows of the LT generator matrix. The universality of (systematic) Raptor codes for erasure channels implies that application of the systematic Raptor codes to the proposed problem will bring a nearly optimal design.

However, our primary aim in this contribution is to analyse and improve the performance of simpler, i.e., non-systematic LT and Raptor codes when applied to the proposed problem of multicasting with decoder side information. The advantage of the non-systematic fountain codes is the simplicity of design and the lower computational complexity. The reduction in computational complexity arises for two principal reasons: (a) systematic Raptor codes require preprocessing to determine an invertible matrix $\mathbf{G}_{LT}^{\{1:k\}}$, solving the system of equations (6) for each block of data, as well as the additional step when decoding: multiplication of intermediate symbols by $\mathbf{G}_{LT}^{\{1:k\}}$ to recover the actual information sequence; (b) decoding of non-systematic fountain codes in multicasting with decoder side information is performed on a significantly smaller decoding graph - if we assume that the number of nodes arising from precoding is negligible, an ideal systematic code performs decoding on a graph with at least $k(1 + \frac{H(X/Y)}{C_{ap}(\mathcal{C}_A)})$ check nodes, whereas an ideal non-systematic code requires slightly more than $k \frac{H(X/Y)}{C_{ap}(\mathcal{C}_A)}$ check nodes. As the decoding time is proportional to the size of the decoding graph, a severalfold decrease in computational complexity is possible.

Example 2: Let \mathcal{C}_V be a BEC with erasure probability $p = 0.1$ and let the transmission channel be noiseless. Ideal systematic fountain code requires $1.1k$ check nodes, whereas ideal non-systematic fountain code requires $0.1k$ check nodes.

V. NON-SYSTEMATIC FOUNTAIN CODING WITH PARTIAL INFORMATION

We have seen that low computational complexity motivates us to study a non-systematic fountain code design for coding

with partial information. However, it is intuitively clear that non-systematic fountain codes designed for standard channel coding problems will be rather inefficient for coding with partial information, since classical output symbol degree distributions will be too sparse to accommodate useful information within output symbols. Namely, since with probability $(1-p)$ an arbitrary input symbol is already known at the decoder, an arbitrary output symbol with degree d is completely useless to the decoder with the probability $(1-p)^d$, which is prohibitively large for the smaller values of d . Thus, it is necessary to modify the degree distributions $\Omega(x)$ and shift them towards higher degrees, while sustaining their compliance with the BP graph pruning algorithm.

Assuming that the transmission of encoding symbols also occurs over a BEC of an undetermined erasure probability, the encoder employs a standard LT code with an output degree distribution $\Phi(x)$ to generate as many encoding packets as necessary. The decoder picks up $t \geq pk$ correctly received packets, where p is the channel erasure probability. Note that pk packets correspond to the optimal compression rate, since $H(X|Y) = p$. The receiver forms the decoding graph and removes the input nodes corresponding to the packets available from the side information, appropriately updating the output nodes. A BP decoding for the erasure channel, i.e., a graph pruning procedure, can then be performed. However, once known input nodes have been removed, the output degree distribution is changed, and exactly this distribution determines the performance of the scheme. Nonetheless, since the source packets are chosen uniformly after the degree of an output node has been selected, one can relate the “starting” degree distribution $\Phi(x)$ to the degree distribution $\Omega(x)$ after removal of the known input nodes from the decoding graph.

Lemma 3: Let $\Phi(x) = \sum_{d=1}^k \Phi_d x^d$ and $\Omega(x) = \sum_{d=1}^k \Omega_d x^d$ be respectively the generating polynomials of the output degree distribution used at the encoder (*incoming* degree distribution) and the output degree distribution after removal of the known source nodes from the decoding graph (*resulting* degree distribution), then:

$$\Omega(x) = \Phi(1 - p + px). \quad (7)$$

Proof: The probability that an arbitrary output node has degree i after removal of the known source nodes conditioned on its degree before removal being $j \geq i$ is given by $\binom{j}{i} (1-p)^{j-i} p^i$. Thus, the relation between the distributions Φ and Ω is given by

$$\Omega_i = \sum_{j=i}^k \Phi_j p^i (1-p)^{j-i}, \quad i = 1, \dots, k, \quad (8)$$

which is equivalent to (7). ■

In [19], the authors independently studied an equivalent problem. They attempted the modification of the LT degree distribution design for erasure channels under the assumption that a fixed number of input packets is already available at the receiver side. The code design was aimed at data synchronization scenarios [31], where each receiver typically has a possibly outdated version of some common database, and needs to recover only a small unknown portion of a large set of data. The authors introduced the shifted robust soliton

distribution, with superior performance over the standard fountain degree distributions, such as robust soliton distributions. The rationale behind the shifted robust soliton distribution is simple. If the original information sequence \mathbf{x} contains k packets and τ of those packets are known at the decoder, each output node of the decoding graph will have (τ/k) -fraction of edges removed prior to executing the BP decoding algorithm. For cases where the number of known input packets is not fixed, but is rather a random variable with a known probability mass function, the authors introduce a distributionally shifted robust soliton distribution.

Definition 2: Let N_l , $l \geq 1$, denote the set $\{1, \dots, l\}$, let $\Omega^{k,c,\delta}(x) = \sum_{d=1}^k \Omega_d^{k,c,\delta} x^d$ denote a robust soliton distribution on N_k with parameters c and δ and let $(\cdot)_{\mathbb{Z}}$ denote rounding to the nearest integer.

(a) The shifted robust soliton distribution (SRSD) $\Phi^{k,c,\delta,\tau}(x) = \sum_{d=1}^k \Phi_d^{k,c,\delta,\tau} x^d$, with parameters c and δ on set N_k shifted by a nonnegative integer $\tau < k$, is $\forall j \in N_k$ given by:

$$\Phi_j^{k,c,\delta,\tau} = \begin{cases} \Omega_i^{k-\tau,c,\delta} & \exists i \in N_{k-\tau} : (\frac{i}{1-\tau/k})_{\mathbb{Z}} = j, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

(b) The distributionally shifted robust soliton distribution (DSRSD) $\Theta^{k,c,\delta,\pi}(x) = \sum_{d=1}^k \Theta_d^{k,c,\delta,\pi} x^d$ with parameters c and δ on set N_k shifted by a distribution $\pi(x) = \sum_{\tau=0}^k \pi_{\tau} x^{\tau}$ on $\{0\} \cup N_k$, is $\forall j \in N_k$ given by:

$$\Theta_j^{k,c,\delta,\pi} = \sum_{\tau=0}^{k-1} \pi_{\tau} \cdot \Phi_j^{k,c,\delta,\tau}. \quad (10)$$

A comparison in Fig. 3 depicts probability mass functions of the robust soliton distribution for $k = 1000$, $c = 0.02$ and $\delta = 0.5$; SRSD with the same parameters for the case when $\tau = 500$ input packets is already known at the decoder; and DSRSD where the number of packets known at the decoder behaves as a binomial random variable on 1000 trials with success probability 0.5, i.e., $\pi(x) = (0.5 + 0.5x)^{1000}$.

The simplicity of the design of SRSD and DSRSD makes them appealing for applications where receivers have access to partial information about the source. However, in the remainder of this section, we will provide a more accurate degree distribution design for such applications. We will start by proving the failure of SRSD and DSRSD to reach Slepian-Wolf limits, as $k \rightarrow \infty$.

A. Penalties of shifted robust soliton distributions

For the sake of simplicity, we will assume that $k/\tau = q \in \mathbb{N}$. In that case, both SRSD shifted by τ and DSRSD where π is a binomial distribution with mean τ converge pointwise to the limiting incoming distribution, as $k \rightarrow \infty$, which is given by:

$$\Phi(x) = \sum_{i \geq 2} \frac{x^{qi}}{i(i-1)}. \quad (11)$$

Note that $\Phi'(x) = -qx^{q-1} \log(1-x^q)$. According to Lemma 3 the limiting resulting distribution is then $\Omega(x) = \Phi(\frac{x+q-1}{q})$.

Now we can apply the vanishing error rate condition (4) at code overhead ε with respect to the Slepian-Wolf limit, i.e., $n = (1+\varepsilon)pk$ encoding packets are observed at the receiver. If

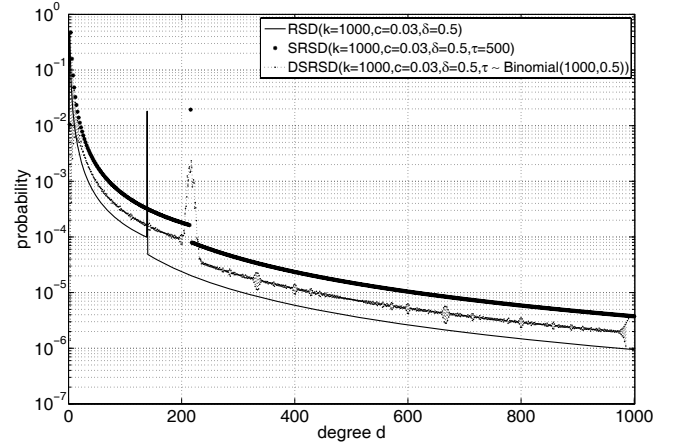


Fig. 3. Comparison of distributions RSD, SRSD and DSRSD.

SRSD and DSRSD should approach Slepian-Wolf limit, $\forall x \in [0, 1)$, $\forall \varepsilon > 0$,

$$-(1+\varepsilon) \left(\frac{x+q-1}{q} \right)^{q-1} \log \left(1 - \left(\frac{x+q-1}{q} \right)^q \right) + \log(1-x) \geq 0. \quad (12)$$

But this cannot be satisfied whenever $q > 0$.

Example 3: Let $q = 2$. Eq. (12) is equivalent to

$$1-x \geq \left(1 - \left(\frac{x+1}{2} \right)^2 \right)^{(1+\varepsilon) \frac{x+1}{2}}. \quad (13)$$

If the condition is to be satisfied $\forall x \in [0, 1)$, i.e., the error rate is vanishing, we require $\varepsilon \geq 0.1227$, whereas as $\varepsilon \rightarrow 0$, the error rate is bounded with $\delta \geq 0.6995$.

Thus, SRSD and DSRSD typically exhibit an "all-or-nothing" decoding behavior. The error rate vanishes after a large threshold code overhead, whereas for smaller code overheads, error rate generally stays large. As we will see, in the design of Raptor-like codes for the partial information setting, better performing distributions can be designed.

B. Optimization of the incoming distribution

Note that, from Lemma 3 we also have $\omega(x) = \phi(1-p+px)$, where $\phi(x) = \frac{\Phi'(x)}{\Phi'(1)}$ is the incoming edge perspective output degree distribution. From here, we can obtain a simple AND-OR tree analysis of the performance of incoming distribution Φ , captured by the following corollary of Lemma 3:

Corollary 1: The packet error rate of an $LT(k, \Phi(x))$ with average input degree α for coding with partial information across the input packets unknown at the decoder prior to transmission, as $k \rightarrow \infty$, is given by $y = \lim_{l \rightarrow \infty} y_l$, where y_l is given by:

$$\begin{aligned} y_0 &= 1, \\ y_l &= \exp(-\alpha \phi(1 - p y_{l-1})). \end{aligned} \quad (14)$$

The packet error rate across the entire information sequence is given by $\hat{y} = p \cdot y$.

If an LT code with output degree distribution $\Phi(x)$ should provide for the decoding of $(1-\delta)t$ unknown source packets, when $k \rightarrow \infty$, where t is the number of unknown source packets, we need to have $\exp(-\alpha \phi(1 - p y)) < y$, $\forall y \in [\delta, 1]$.

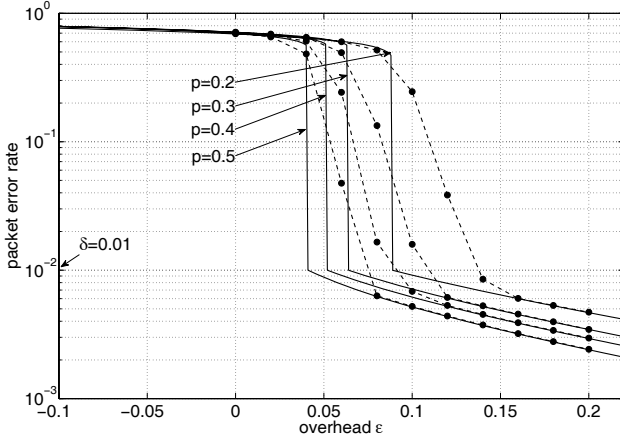


Fig. 4. Asymptotic (full lines) and simulated (dashed lines) packet error rates of degree distributions with desired error rate $\delta = 0.01$ at the minimized code overhead for correlation BEC of probability $p = 0.2, 0.3, 0.4, 0.5$. The block length used in simulations was $k = 6 \cdot 10^4$.

By transforming the above condition in terms of a function linear in variables ϕ_d we obtain the set $LP_2(p, \delta, d_{max}, N)$ of linear programs:

$$LP_2 : \quad \min \frac{1}{p} \sum_{d=1}^{d_{max}} \frac{\phi_d}{d}$$

$$\phi(y_i) \geq -\ln\left(\frac{1-y_i}{p}\right), \quad i \in 1, 2, \dots, N \quad (15)$$

$$\phi_d \geq 0, \quad d \in 1, 2, \dots, d_{max}.$$

where $1-p = y_1 < y_2 < \dots < y_N = 1 - \delta p$ are equidistant points on $[1-p, 1 - \delta p]$. Fig. 4 shows the asymptotic and simulated finite length packet error rates for degree distributions obtained by linear program LP_2 for $d_{max} = 100$, $\delta = 0.01$, and $p \in \{0.2, 0.3, 0.4, 0.5\}$, with grid $y_1 < y_2 < \dots < y_N$ of granularity 0.001. The block length used in simulations was $k = 6 \cdot 10^4$.

Unfortunately, we have discovered that code design with linear programs (15) comes with a code overhead penalty: for fixed δ , ϵ with respect to Slepian-Wolf limit stays bounded above some penalty value $\epsilon^* > 0$ as the maximum degree $d_{max} \rightarrow \infty$. Namely, a simple modification of Lemma 2 from [24] yields the following result:

Proposition 1: Let $m \in \mathbb{N}$ be such that $\delta \geq \frac{1}{p(m+1)}$. There exists a solution $\phi^*(x)$ of $LP_2(p, \delta, d_{max}, N)$, $d_{max} > m$, with $\phi_j^* = 0$ for $j \geq m+1$, Now we can look at the duals of the undiscretized version of $LP_2(p, \delta, m, N)$:

$$\max_{pZ} \mathbb{E}\left[-\ln \frac{1-Z}{p}\right] \quad (16)$$

$$\mathbb{E}[Z^{d-1}] \leq \frac{1}{pd}, \quad d \in 1, 2, \dots, m$$

$$Z \in [1-p, 1 - \delta p],$$

The dual program enables us to lower-bound the code overhead penalty, as for any feasible solution Z of the dual and, the value of the objective function of the dual ψ_{dual} is less than the optimal value of the objective function of the primal $\psi_{primal}^* = 1 + \epsilon^*$. We can obtain a feasible solution of the dual by looking at the discrete distributions with finite

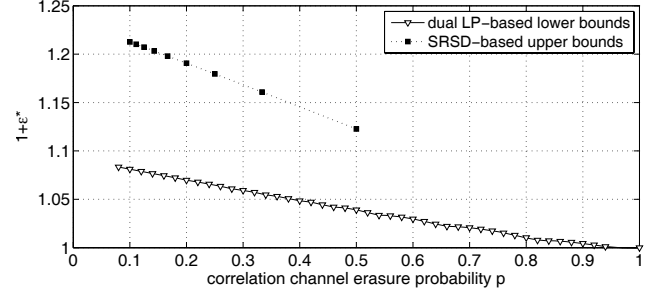


Fig. 5. Code overhead penalty in coding with partial information.

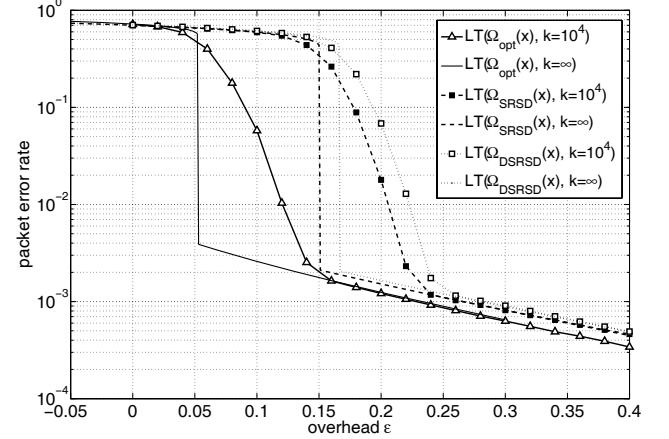


Fig. 6. Comparison of asymptotic and simulated performance of light SRSD, light DSRSD ($d_{max} = 100$) and degree distribution obtained by linear program $LP_2(p = 0.5, \delta = 0.004, d_{max} = 100, N = 500)$.

support, e.g., on a uniform grid. The lower bounds obtained this way, for $\delta = 0.01$ and uniform grid of granularity 0.001 are shown in Fig. 5 and compared with penalties arising from SRSD and DSRSD, calculated as in Example 3. Note that for the case of $p = 1$, or the classical channel coding problem, there is no penalty - namely, an ensemble of LT codes whose degree distributions converge pointwise to a limiting soliton distribution [24] would achieve the Shannon limit, whereas there is a clear gap with $p < 1$. Our results in Fig. 4 demonstrate that these lower bounds are sharp.

We have seen that coding schemes with linear encoding-decoding complexity employ light degree distributions, capped at some maximum degree. Our results indicate that light SRSD and DSRSD distributions perform poorly. In Fig. 6, we present the asymptotic and finite length ($k = 10^4$) performance of the custom degree distribution $\Omega_{opt}(x) = 0.4816x^5 + 0.3916x^6 + 0.0792x^{29} + 0.0051x^{30} + 0.0425x^{100}$ as contrasted to light SRSD and DSRSD of the same maximum degree $d_{max} = 100$. Distribution $\Omega_{opt}(x)$ was chosen as to mimic the error floor of SRSD and DSRSD but at the minimized code overhead. Note that although DSRSD is actually based on the assumption that partial information is an output of a BEC, it actually performs worse than SRSD in this setting, and we attribute this result to a small maximum degree $d_{max} = 100$.

Thus, our study of non-systematic fountain codes for coding with partial information has led to the following conclusions: (a) light SRSD and DSRSD suffer significantly larger penalties on code overhead as $k \rightarrow \infty$, (b) SRSD and DSRSD perform poorly in comparison to optimized degree distributions which

are more applicable for linear complexity and Raptor-like schemes.

C. Raptor-like scheme

By concatenating an LT code with a custom distribution derived by the optimization procedure as described above to a very high-rate hybrid LDPC-Half precode, constructed as in [17], we have implemented a non-systematic Raptor solution to the problem of coding with partial information. The length of the information sequence was set to $k = 4 \cdot 10^4$ and the correlation channel erasure probability was $p = 0.3$. Note that the precoding somewhat changes the optimization procedure since equation (7) now reads

$$\Omega(x) = \Phi(s(1-p) + (1-s+sp)x), \quad (17)$$

where s is the precode rate. Fig. 7 depicts the histogram of the numbers of received packets necessary for successful decoding for 500 transmission trials. On average, 13750 packets were required for full recovery, which is about 34.4% of the length of the information sequence, compared to the optimal 30%, i.e., 12000 packets. This demonstrates that our code design, albeit sub-optimal, can be utilized in a practical robust low complexity data multicast coding scheme with partial information.

VI. SOFT-DECISION DECODING AND NOISY CORRELATION CHANNELS

Let us assume that the binary information source X over the alphabet $\{-1, 1\}$ and the soft side information Y are correlated via $Y = X + N$, where N is a Gaussian random variable of zero mean and variance σ_V^2 . This means that \mathcal{C}_V is a binary input additive white Gaussian noise channel (BIAWGNC) with noise variance σ_V^2 . In this case, $H(X|Y) = 1 - \text{Cap}(\text{BIAWGN}_{\sigma_V})$, where the capacity of BIAWGNC [22] is given by

$$\text{Cap}(\text{BIAWGN}_{\sigma}) = 1 - \frac{1}{2\sqrt{\pi m}} \int_{-\infty}^{\infty} \log_2(1 + e^{-x}) e^{-\frac{(x-m)^2}{4m}} dx, \quad (18)$$

where $m = 2/\sigma^2$.

Fountain codes on general noisy binary input memoryless symmetric (BIMS) channels can be decoded by a BP sum-product algorithm [7], [21]. Every output node f , corresponding to the encoding symbol, has a corresponding channel log-likelihood ratio (LLR) $L(z_f)$, derived based on the channel output z_f . In addition, an input node v may be associated to the side information y_v , if present. This side information can be embedded directly into the sum-product rules as intrinsic soft information, i.e., log-likelihood ratio $L(y_v)$ based on the output of the correlation channel. The sum-product rules are given by:

$$m_{v,f}^{(i)} = \begin{cases} L(y_v), & i = 0 \\ L(y_v) + \sum_{g \neq f} \mu_{g,v}^{(i-1)}, & i > 0 \end{cases} \quad (19)$$

$$\tanh\left(\frac{\mu_{f,v}^{(i)}}{2}\right) = \tanh\left(\frac{L(z_f)}{2}\right) \prod_{u \neq v} \tanh\left(\frac{m_{u,f}^{(i)}}{2}\right), \quad (20)$$

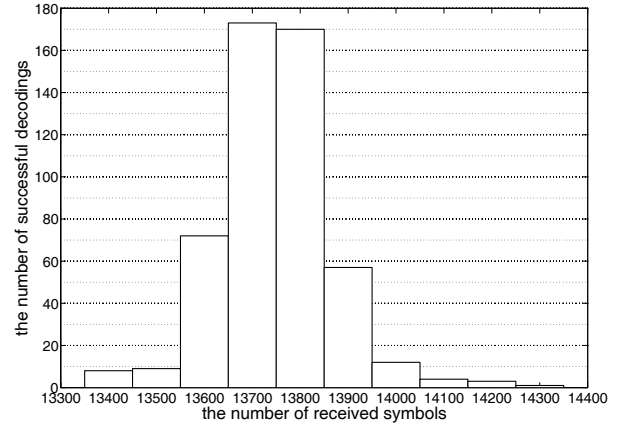


Fig. 7. The histogram of the number of successful recoveries in non-systematic Raptor coding with partial information.

where $\mu_{f,v}^{(i)}$ ($m_{v,f}^{(i)}$) are the messages passed from the output node f to the input node v (from the input node v to the output node f) at the i -th iteration. In Raptor codes, after a fixed number of iterations l , the posterior LLR of the input node v is given by

$$\hat{L}(y_v) = L(y_v) + \sum_g \mu_{g,v}^{(l)}. \quad (21)$$

and can be used as a prior value in additional BP iterations on the static decoding graph of the Raptor code, i.e., the decoding graph of the precode. This additional step has a role of removing any error floor arising from light degree distributions. In [13], a similar design was employed for the joint-source channel coding scenario using non-systematic Raptor codes with a standard Soliton-like output degree distribution. In the rest of this section, we will show how to improve the design of output degree distributions in this setting.

A. Using a systematic Raptor design

Systematic Raptor codes can be applied in the noisy side information scenario similarly as described in Section IV. It has recently come to our attention that the problem of the side information scenario with the correlation channel modelled by a binary symmetric channel (BSC) was independently studied in [10]. The authors employed the systematic Raptor codes and modified the message passing strategy at the decoder to take into account the source Y perfectly known at the decoder as the noisy version of source X and transmit only the non-systematic encoding Raptor symbols, which is a soft-decision version of the systematic Raptor employment described in section 4. However, in addition to the conclusion that systematic Raptor codes can be used in this setting, it was also argued in [10] that non-systematic Raptor codes are not applicable to the problem. However, the authors of [13] consider a similar setting but use standard non-systematic Raptor codes enhanced by a more frequent selection of parity symbols in formation of the encoding symbols. We here show that, by carefully designing their output degree distributions, non-systematic Raptor codes may yield promising performance for coding with noisy side information as well.

B. Density evolution and semi-Gaussian approximation

The BP decoder of LDPC codes is extensively analysed with the set of tools collectively referred to as density evolution (DE) [22]. The density evolution calculates density functions of messages passed during the BP algorithm. This approach is significantly simplified by *Gaussian approximation* [32]. Gaussian approximation models all the messages passed during the decoding algorithm as consistent Gaussian variables, i.e., Gaussian variables whose variance is equal to twice their mean. However, a more accurate analysis is possible with *semi-Gaussian approximation* [33], which was used in the fountain code design for noisy channels [7]. We will also adopt this approach in fountain coding with noisy side information.

The messages passed from the variable nodes are obtained as sums of i.i.d. random variables of finite mean and variance and behave as Gaussian random variables on large scale. However, as argued in [33], the messages passed from the check nodes (especially those with a small degree) exhibit a rather different behaviour. Hence, we will assume that the input→output messages $M_j^{(i)} \sim \mathcal{N}(\nu_i, 2\nu_i)$, $i \geq 0$ are consistent Gaussian variables and explicitly calculate the mean of the output→input messages. The key part of the analysis is the function η which describes how the mean of input→output messages changes in a single iteration of an LT decoder, i.e., $\nu_{i+1} = \eta(\nu_i)$, $i \geq 0$. By taking expectations in (19) and (20), we obtain:

$$\eta(\nu) = \mathbb{E}L(Y) + \alpha \sum_{d=1}^{d_{max}} \omega_d \xi(\nu, d, Z). \quad (22)$$

where $\xi(\nu, d, Z)$ is the mean of the output→input messages passed from an output node of degree d [7], and is given by:

$$\xi(\nu, d, Z) = 2\mathbb{E} \left[\operatorname{atanh} \left(\tanh \left(\frac{Z}{2} \right) \prod_{j=1}^{d-1} \tanh \left(\frac{M_j}{2} \right) \right) \right]. \quad (23)$$

Here, Z is the random variable describing the LLR of the transmission channel and $M_j \sim \mathcal{N}(\nu, 2\nu)$, $j \in \{1, \dots, d-1\}$, are i.i.d. random variables. As suggested in [7], $\xi(\nu, d, Z)$ can be approximated by an empirical mean.

The condition that the BP decoder converges to an all-zero codeword translates to $\eta(\nu) > \nu$ on $\nu \geq \mathbb{E}L(Y)$. In the fountain code design for channel coding, the starting mean of the input→output messages is zero, and thus a corresponding condition becomes too restrictive. This explains poor performance of standard fountain code degree distributions for coding with noisy side information, as reported in [10]. However, incorporating the condition $\eta(\nu) > \nu$ on $\nu \in [\mathbb{E}L(Y), \nu_{max}]$, for some predetermined cut-off LLR ν_{max} , into our code design problem produces a robust way to design non-systematic fountain codes for this problem.

Example 4: Assume that \mathcal{C}_V is a BIAWGNC with noise variance σ_V^2 and \mathcal{C}_A is a BIAWGNC with noise variance σ_A^2 . We obtain the following set $LP_3(\sigma_V^2, \sigma_A^2, \nu_{max}, d_{max}, N)$ of linear programs:

$$LP_3 : \quad \min \frac{\operatorname{Cap}(\mathcal{C}_A)}{1 - \operatorname{Cap}(\mathcal{C}_V)} \sum_{d=1}^{d_{max}} \frac{\omega_d}{d}$$

$$\sum_{d=1}^{d_{max}} \omega_d \xi(\nu_i, d, Z) \geq \nu_i - 2/\sigma_V^2, \quad i \in 1, 2, \dots, N,$$

$$\omega_d \geq 0, \quad d \in 1, 2, \dots, d_{max}, \quad (24)$$

where $2/\sigma_V^2 = \nu_1 < \nu_2 < \dots < \nu_N = \nu_{max}$ are equidistant points on $[2/\sigma_V^2, \nu_{max}]$.

C. Gaussian transmission with partial information

The optimization of degree distributions in the case when \mathcal{C}_V is a BEC of probability p and \mathcal{C}_A is a BIAWGNC of noise variance σ_V^2 follows from similar ideas. It is sufficient to insert the relationship

$$\omega_d = \sum_{i=d}^{d_{max}} \binom{i}{d} (1-p)^{i-d} p^d \phi_d, \quad (25)$$

into condition (22). In this case, the input→output means start at $\nu = 0$ as we track the means at the portion of data unknown apriori, and this portion of data contains no soft information: $L(Y) = 0$. The new design constraints are given by:

$$\sum_{i=1}^{d_{max}} \left(\sum_{d=1}^i \binom{i}{d} (1-p)^{i-d} p^d \xi(\nu, d, Z) \right) \phi_i > \nu, \quad \nu \in [0, \nu_{max}], \quad (26)$$

and can be easily transformed into an appropriate linear program.

D. Simulation results

We compared three different methods for coding with Gaussian side information on information sequence of length $k = 3140$: a systematic Raptor code, a standard non-systematic Raptor code with degree distribution from [3], [10] and the non-systematic Raptor code with degree distribution $\Omega(x) = 0.0954x^5 + 0.1192x^6 + 0.1121x^7 + 0.12938x^8 + 0.1054x^9 + 0.0807x^{10} + 0.1109x^{11} + 0.2470x^{100}$, obtained from LP in (24). The results are presented in Fig. 8. The horizontal axis represents the signal-to-noise ratio (SNR) of the correlation channel, which is related with the channel noise variance by $\text{SNR} = 10 \log_{10} \frac{1}{\sigma^2}$. The vertical axis represents the average joint source-channel code rate necessary for succesful decoding, i.e., t/k , where t is the average number of received encoding symbols at the decoder. The transmission channel was a BIAWGN channel with $\text{SNR} = 3$ dB. The assumed SNR of the virtual channel during this optimization was also set to 3 dB. The systematic Raptor is clearly superior to non-systematic schemes in this setting. However, the non-systematic Raptor code with optimized $\Omega(x)$ does come close to the performance of the systematic Raptor code scheme at the higher region of virtual SNR. This demonstrates that non-systematic Raptor codes with carefully designed degree distributions may nonetheless be an attractive solution for coding with noisy side information. Note, however, that in the lower region of virtual SNR our code design constraints become insufficient to provide low overheads as starting means of the input→output messages are lower than anticipated.

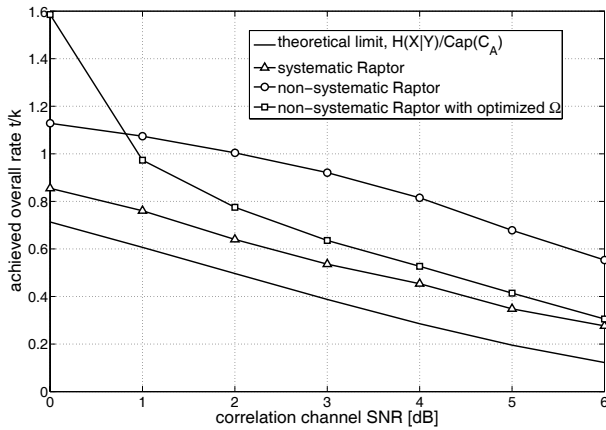


Fig. 8. The comparison of Raptor codes for coding with noisy correlation.

VII. CONCLUSIONS

In this paper, the design of fountain codes for multicast transmission with side information at the receivers has been studied. We have assumed that side information is modelled as the correlation channel output when the original information sequence is its input. We have investigated the instances of problem with correlation channel modelled either as a BEC or as a BIAWGNC. While in both cases a solution based on systematic Raptor codes seems the most advantageous, its higher complexity and the need for preprocessing motivated us to study the performance of the non-systematic fountain codes when applied to these multicast problems. We have shown how to improve their performance by optimizing the output degree distribution for a particular correlation channel model. The results in our contribution indicate that fountain codes are a natural practical coding scheme for multicast transmission with side information and that their design may be tuned to perform close to information theoretic bounds with low computational cost.

ACKNOWLEDGEMENTS

D. Sejdinović and R. Piechocki would like to thank Toshiba Telecommunications Research Laboratory and its directors for supporting this work.

REFERENCES

- [1] D. Sejdinović, R. Piechocki, A. Doufexi, and M. Ismail, "Fountain coding with decoder side information," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008.
- [2] M. Luby, "LT codes," in *Proc. IEEE Symp. Foundations Computer Science (FOCS)*, Vancouver, Canada, Nov. 2002.
- [3] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551-2567, June 2006.
- [4] R. Palanki and J. Yedidia, "Rateless codes on noisy channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Chicago, USA, July 2004.
- [5] H. Jenkač, T. Mayer, T. Stockhammer, and W. Xu, "Soft decoding of LT codes for wireless broadcast," in *Proc. IST Summit 2005*, Dresden, Germany, June 2005.
- [6] J. Castura and Y. Mao, "Rateless coding over fading channels," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 46-48, Jan. 2006.
- [7] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2033-2051, May 2006.
- [8] S. Shamai, E. Telatar, and S. Verdú, "Fountain capacity," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4372-4377, Nov. 2007.
- [9] G. Caire, S. Shamai, A. Shokrollahi, and S. Verdú, "Fountain codes for lossless data compression," *Algebraic Coding Theory Inform. Theory, in DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, A. Ashikhmin, A. Barg, I. Duursma, Eds., vol. 68, pp. 1-20, AMS, 2006.
- [10] M. Fresia and L. Vandendorpe, "Distributed source coding using Raptor codes," in *Proc. IEEE GLOBECOM*, Washington DC, USA, Nov. 2007.
- [11] B. N. Ndzana, A. Shokrollahi, and J. Abel, "Fountain codes for the Slepian-Wolf problem," in *Proc. Allerton Conf.*, Allerton, USA, 2006.
- [12] N. Dütsch, H. Jenkač, T. Mayer, and J. Hagenauer, "Joint source-channel fountain coding for asynchronous broadcast," in *Proc. IST Summit 2005*, Dresden, Germany, June 2005.
- [13] Q. Xu, V. Stanković, and Z. Xiong, "Distributed source-channel coding of video using Raptor codes," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 851-861, May 2007.
- [14] O. Y. Bursalioğlu, M. Fresia, G. Caire, and H. V. Poor, "Lossy joint source-channel coding using Raptor codes," *Int. J. Digital Multimedia Broadcasting*, 2008.
- [15] J. P. Wagner, J. Chakareski, and P. Frossard, "Streaming of scalable video from multiple sources using rateless codes," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Toronto, Canada, July 2006.
- [16] Q. Xu, V. Stanković, and Z. Xiong, "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 17, pp. 901-906, July 2007.
- [17] "Universal Mobile Telecommunications System (UMTS): Multimedia Broadcast/Multicast Service (MBMS), Protocols and Codecs," 3GPP TS 26.346, Ver. 6.3.0, Dec. 2006.
- [18] "Digital Video Broadcasting (DVB): IP-Datcast over DVB-H, Content Delivery Protocols," ETSI TS 102 472 V1.2.1, 2006.
- [19] S. Agarwal, A. Hagedorn, and A. Trachtenberg, "Adaptive rateless coding under partial information," in *Proc. Inform. Theory Applications (ITA) Workshop*, San Diego, USA, Feb. 2008.
- [20] J. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *Proc. ACM SIGCOMM Conf.*, Vancouver, Canada, Sept. 1998.
- [21] F. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498-519, 2001.
- [22] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.
- [23] M. Luby, M. Mitzenmacher, and A. Shokrollahi, "Analysis of random processes via AND-OR tree evaluation," in *Proc. SIAM Symp. Discrete Algorithms (SODA)*, San Francisco, USA, Jan. 1998.
- [24] S. Sanghavi, "Intermediate performance of rateless codes," *IEEE Inform. Theory Workshop 2007*, Tahoe, USA, Sept. 2007.
- [25] R. W. R. Darling and J. R. Norris, "Structure of large random hypergraphs," *Ann. Appl. Probab.*, vol. 15, no. 1A, 2005.
- [26] D. Slepian, J.K. Wolf. "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471-480, July 1973.
- [27] A. W. Eckford and W. Yu, "Rateless Slepian-Wolf codes," in *Proc. Asilomar Conf.*, Pacific Grove, USA, 2005.
- [28] S. Shamai and S. Verdú, "Capacity of channels with uncoded side information," *European Trans. Telecommun.*, vol. 6, pp. 587-600, 1995.
- [29] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf.*, Allerton, USA, 2003.
- [30] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Trans. Broadcasting*, vol. 53, no. 1/2, pp. 235-246, Mar. 2007.
- [31] S. Agarwal, A. Hagedorn, and A. Trachtenberg, "Near optimal update-broadcast of datasets," in *Proc. Int. Workshop Data Intensive Sensor Networks (DISN)*, Mannheim, Germany, May 2007.
- [32] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 657-670, Feb. 2001.
- [33] M. Ardakani and F.R. Kschischang, "A more accurate one-dimensional analysis and design of irregular LDPC codes," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2106-2114, Dec. 2004.



Dino Sejdinović was born in 1984 in Tuzla, Bosnia-Herzegovina, and received a Dipl. Math.-Inf. degree in mathematics and theoretical computer science from the University of Sarajevo, Bosnia-Herzegovina in 2006. He is currently pursuing his Ph.D. degree in electrical and electronic engineering at the University of Bristol, as a recipient of Ph.D. studentship by Toshiba Research Europe Ltd. His research interests include coding and information theory, inference in graphical models, graph theory and wireless communications.



Robert J. Piechocki is currently RCUK Senior Research Fellow at the University of Bristol. He received an M.Sc. degree from the Technical University of Wrocław, Poland in 1997 and a Ph.D. from the University of Bristol in 2002. Robert conducts research in signal processing, coding and networking for wireless communications systems. He has authored and co-authored over 60 journal and conference papers and is a holder of 12 patents.



Angela Doufexi is currently Senior Lecturer in Wireless Networks at the University of Bristol. She graduated from the University of Athens with a B.Sc. in Physics in 1996. She received her M.Sc. in Electronic Engineering from Cardiff University in 1998. She then joined the Centre for Communications Research at the University of Bristol, where she received her Ph.D. in 2002. She has worked in a number of projects including the European IST projects SATURN, ROMANTIK, WCAM and ASTRALS. Her research interests include OFDM and OFDMA systems, multiuser diversity and resource allocation, wireless LANs and WiMax (PHY and MAC), space time coding and MIMO, LTE and 4th generation communications systems, cognitive radio and multimedia transmission. She has published over 90 journal and conference papers in these areas.



Mohamed Ismail received his M.Sc. degree in signal processing from the University of Surrey, United Kingdom, in 1998. Between 1990 and 1999 he was involved in research and development of speech recognition, synthesis and coding technologies. Since 1999, while working for the Telecommunications Research Laboratory (TRL) of Toshiba Research Europe Limited, he has been researching wireless communications. His current interests are in channel coding, especially for high throughput, low power systems.