On the Concatenation of Non-Binary Random Linear Fountain Codes with Maximum Distance Separable Codes

Francisco Lazaro Blasco
Institute of Communications and Navigation
DLR (German Aerospace Center)
Wessling, Germany 82234
Email: Francisco.LazaroBlasco@dlr.de

Gianluigi Liva
Institute of Communications and Navigation
DLR (German Aerospace Center)
Wessling, Germany 82234
Email: Gianluigi.Liva@dlr.de

Abstract—The performance of a novel fountain coding scheme based on maximum distance separable (MDS) codes constructed over Galois fields of order $q = 2^r$ is investigated. Upper and lower bounds on the decoding failure probability under maximum likelihood decoding are developed. Differently from Raptor codes (which are based on a serial concatenation of a high-rate outer block code, and an inner Luby-transform code), the proposed coding scheme can be seen as a parallel concatenation of an outer MDS code and an inner random linear fountain code, both operating on the same Galois field. A performance assessment is performed on the gain provided by MDS based fountain coding over linear random fountain coding in terms of decoding failure probability vs. overhead. It is shown how, for example, the concatenation of a (15, 10) Reed-Solomon code and a linear random fountain code over $\mathbb{F}_{16}$ brings the decoding failure probability 4 orders of magnitude lower than the linear random fountain code for the same overhead in a channel with a packet loss probability of $\epsilon = 5 \cdot 10^{-2}$. Moreover, it is illustrated how the performance of the concatenated fountain code approaches that of an idealized fountain code for higher-order Galois fields and moderate packet loss probabilities. The scheme introduced is of special interest for the distribution of data using small block sizes.

I. INTRODUCTION

Fountain codes were introduced in [1] as an efficient alternative to automatic retransmission query (ARQ) protocols in multicast/broadcast transmission systems. Consider the case where a sender (or source) needs to deliver a file to a set of $N_u$ users. Consider furthermore the case where users are affected by packet losses. In this scenario, the usage of an ARQ protocol can result in large inefficiencies, since users may loose different packets, and hence a large number of retransmissions would crowd the downlink channel. Among the efficient (coded) alternatives to ARQ protocols [2]–[5], we shall focus on fountain codes only. When a fountain code is used, the source file is split in a set of $k$ source packets. The sender, or fountain encoder, computes linear combinations of the $k$ source packets and broadcasts them through the communication medium. After receiving $k$ fountain coded packets, receivers can try to recover the source packets. If they fail to recover the source packets they will try again to decode when they receive additional packets. The efficiency of a fountain code deals with the amount of packets (source+redundancy) that a user needs to collect for recovering the source file. An idealized fountain code would allow the file recovery with a probability of success $P_s = 1$ from any set of $k$ received packets. Real fountain decoders need in general to receive a larger amount of packets, $m = k + \delta$, for achieving a certain success probability. Commonly, $\delta$ is referred to as overhead of the fountain code, and is used to measure its efficiency. More generally a universal fountain code is a code which can recover the $k$ original source symbols out of $k+\delta$ symbols for any erasure channel and $\delta$ small. The first class of universal fountain codes are Luby-transform (LT) codes [6]. One subclass of LT codes are random LT codes or linear random fountain codes (LRFCs) [7]. When a binary LRFC is used [8], [9] the success probability can be accurately modeled as $P_s = 1 - 2^{-\delta}$ for $\delta > 2$ (it can be proved that $P_s$ is actually always lower bounded by $1 - 2^{-\delta}$ [9]). In [9] it was shown that this expression is still accurate for fountain codes based on sparse matrices (e.g., Raptor codes [7]). Moreover, in [9], the performance achievable by performing linear combinations of packets on Galois fields of order greater than 2 was analyzed. For a LRFC performing the linear combinations over $\mathbb{F}_q$, the decoding failure probability $P_e = 1 - P_s$ is bounded by [9]

$$q^{-\delta-1} \leq P_e(\delta, q) < \frac{1}{q - 1}q^{-\delta}$$

where both bounds are tight for increasing $q$. Furthermore, in [9] it was also shown that non-binary Raptor codes can in fact tightly approach the bounds (1) down to moderate error rates. The result is remarkable, considering that for a Raptor code the encoding and decoding costs (defined as the number of arithmetic operations divided by the number of source symbols, $k$) are $O(\log(1/\alpha))$ and $O(k \log(1/\alpha))$ respectively, being $k(1+\alpha)$ the number of output symbols needed to recover the source symbols with a high probability. For a LRFC the encoding cost is $O(k)$ and the decoding cost is $O(k^2)$, and thus it does not scale favorably with the input block size.
However, if the block size is kept small, the decoding cost is still affordable.

The motivation of this paper is the analysis of a further improvement of the approach proposed in [9] for designing fountain codes with good performance for short block sizes. As in [9], in order to achieve the objective non-binary fountain codes are considered. Moreover, maximum distance separable (MDS) codes are introduced in parallel concatenation with the fountain encoder to enhance the performance of the scheme. By doing that, the first \( n \) output symbols of the encoder are the \( n \) output symbols of the MDS code.\(^1\)

In this paper, we illustrate how the performance of \( \text{LRFC} \) in terms of probability of decoding failure can be further improved by such a concatenation. An analytical expression for the decoding failure probability vs. overhead will be derived under the assumption of maximum-likelihood (ML) decoding. We show how, when the packet loss rates are moderate-low, the probability of failure can be reduced by several orders of magnitude, approaching the performance of idealized fountain codes. The simulated performance of schemes based on Reed Solomon (RS) codes are compared with the proposed expressions, confirming the accuracy of the proposed approach. The analysis is developed for the case of \( \text{LRFC} \). We conjecture that similar gains shall be expected also in the case where (non-binary) LRFC codes are employed in the concatenation.

The paper is organized as follows. In Section II the proposed concatenated scheme is introduced. In Section III the performance analysis is provided, while numerical results are presented in Section IV. Conclusions follow in Section V.

II. CONCATENATION OF BLOCK CODES WITH RANDOM LINEAR FOUNTAIN CODES

Without losing in generality, we define the source block \( \mathbf{u} = (u_1, u_2, \ldots, u_k) \) as a sequence of symbols belonging to a Galois field of order \( q \), i.e. \( \mathbf{u} \in \mathbb{F}_q^k \). In the proposed approach, the source block is first encoded via a \((n, k)\) systematic linear block code \( \mathbf{C'} \) over \( \mathbb{F}_q \) with generator matrix \( \mathbf{G'} = ([I] \mathbf{P'}) \), where \( I \) is the \( k \times k \) identity matrix and \( \mathbf{P'} \) is a \( k \times (n - k) \) matrix with elements in \( \mathbb{F}_q \). The encoded block is hence given by \( \mathbf{c'} = \mathbf{uG'} = (c'_1, c'_2, \ldots, c'_n) \), where \( c'_1 = u_1, c'_2 = u_2, \ldots, c'_k = u_k \) and the remaining \( n - k \) symbols of \( \mathbf{c'} \) are the redundancy symbols given by \( (c'_{k+1}, c'_{k+2}, \ldots, c'_n) = \mathbf{uP'} \).

Additional redundancy combinations can be obtained by computing random linear combinations of the \( k \) source symbols as

\[
c_i = c''_i = \sum_{j=1}^{k} g_{j,i} u_j, \quad i = n + 1, \ldots, l
\]

where the coefficients \( g_{j,i} \) are picked from \( \mathbb{F}_q \) with a uniform probability \((1/q)\). The encoded sequence is hence given by \( \mathbf{c} = (c'_1, c'_2, \ldots, c'_n) \). The overall generator matrix has the form

\[
\mathbf{G} = \begin{bmatrix}
g_{1,1} & g_{1,2} & \cdots & g_{1,n} 
g_{2,1} & g_{2,2} & \cdots & g_{2,n} 
\vdots & \ddots & \ddots & \vdots 
g_{k,1} & g_{k,2} & \cdots & g_{k,n} 
g_{1,n+1} & g_{1,n+2} & \cdots & g_{1,l} 
g_{2,n+1} & g_{2,n+2} & \cdots & g_{2,l} 
\vdots & \ddots & \ddots & \vdots 
g_{k,n+1} & g_{k,n+2} & \cdots & g_{k,l}
\end{bmatrix}
\]

where \( \mathbf{G''} \) is the generator matrix of the \( \text{LRFC} \). (Note that, being the \( \text{LRFC} \) rate-less, the number \( l \) of columns of \( \mathbf{G} \) can in principle grow indefinitely.) The encoder can be seen hence as a parallel concatenation of the linear block code \( \mathbf{C'} \) and of a \( \text{LRFC} \) (Fig. 1).

\[
c_1, c_2, \ldots, c_n \quad \text{Block Code} \quad (n, k) \quad u_1, u_2, \ldots, u_k \quad c_1, c_2, \ldots, c_n, c_{n+1} \ldots \quad \text{LRFC}
\]

Fig. 1. Fountain coding scheme seen as a parallel concatenation of a \((n, k)\) linear block code and a linear random fountain code.

III. PERFORMANCE ANALYSIS

Based on the bounds derived in [9], tight upper and lower bounds for the decoding failure probability of the fountain coding scheme can be derived in case of uncorrelated erasures. The decoding failure probability \( P_F = \Pr(F) \), where \( F \) denotes the decoding failure event, is defined as the probability that the source block \( \mathbf{u} \) cannot be recovered out of a set of \( m \) received symbols. In this paper we will focus on the case where the linear block code used in concatenation with the \( \text{LRFC} \) is maximum distance separable (MDS). When binary codes will be used, we will assume \((k + 1, k)\) single-parity-check (SPC) codes. When operating on higher order Galois fields, we will consider (shortened) RS codes.

The encoded sequence is given by \( \mathbf{c} = \mathbf{uG} = (c_1, c_2, \ldots, c_l) \), where the first \( n \) symbols \((c_1, c_2, \ldots, c_n)\) represent a codeword of \( \mathbf{C'} \), and the remaining \( l - n \) symbols are produced by the \( \text{LRFC} \). At the receiver side, a subset of \( m \) symbols is received. We denote by \( J = \{j_1, j_2, \ldots, j_m\} \) the set of the indexes on the symbols of \( \mathbf{c} \) that have been received. The received vector \( \mathbf{y} \) is hence given by

\[
\mathbf{y} = (y_1, y_2, \ldots, y_m) = (c_{j_1}, c_{j_2}, \ldots, c_{j_m})
\]

and it can be related to the source block \( \mathbf{u} \) as \( \mathbf{y} = \mathbf{uG} \). Here, \( \mathbf{G} \) denotes the \( k \times m \) matrix made by the columns of \( \mathbf{G} \) with

\[^1\text{Note that for Raptor codes the output of the precode is further encoded by a LRFC code. Hence the first } n \text{ output symbols of the fountain encoder are not the output of the precode.}\]

\[^2\text{We will assume a MDS linear block code constructed on the same field } (\mathbb{F}_q) \text{ of the fountain code.}\]

\[^3\text{Repetition codes are not considered here, since they would lead to a trivial fountain scheme where the source block is given by 1 symbol only.}\]
The recovery of \( u \) reduces to solving the system of \( m = k + \delta \) linear equations in \( k \) unknowns
\[
\tilde{G}^T u^T = y^T, \tag{3}
\]
e.g., via Gaussian elimination. The solution is possible if and only if \( \text{rank}(\tilde{G}) = k \).

Assuming \( C' \) being \textbf{MDS} the system is solvable with probability 1 if, among the \( m \) received symbols, at least \( k \) have indexes in \( \{1, 2, \ldots, n\} \), i.e. if at least \( m' \geq k \) symbols produced by the linear block encoder have been received.

Let’s consider the less trivial case where \( m' < k \) among the \( m \) received symbols have indexes in \( \{1, 2, \ldots, n\} \). We can partition \( \tilde{G}^T \) as
\[
\tilde{G}^T = \begin{pmatrix} \tilde{G}'^T \\ \tilde{G}''^T \end{pmatrix} = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,m'} \\
g_{2,1} & g_{2,2} & \cdots & g_{2,m'} \\
\vdots & \vdots & \ddots & \vdots \\
g_{k,1} & g_{k,2} & \cdots & g_{k,m'} \\
g_{1,m'+1} & g_{2,m'+1} & \cdots & g_{k,m'+1} \\
g_{1,m'+2} & g_{2,m'+2} & \cdots & g_{k,m'+2} \\
\vdots & \vdots & \ddots & \vdots \\
g_{1,m'} & g_{2,m'} & \cdots & g_{k,m'} \end{pmatrix}. \tag{4}
\]

The \textbf{MDS} property of \( C' \) assures that \( \text{rank}(C') = m' \), i.e. the first \( m' \) rows of \( \tilde{G}^T \) are linearly independent. Note that the \( m'' \times k \) matrix \( \tilde{G}''^T \) (with \( m'' = m - m' \)) is a random matrix whose entries are picked with uniform probability in \( \mathbb{F}_q \). It follows that the system defined by \( \tilde{G}^T \) can be put (via column permutations over \( \tilde{G}^T \) and row permutations/combinations over \( \tilde{G}''^T \)) in the form
\[
\tilde{G}^T = \begin{pmatrix} 1 & A \\ 0 & B \end{pmatrix}, \tag{5}
\]
where \( I \) is the \( m' \times m' \) identity matrix, \( 0 \) is a \( m'' \times m' \) all-0 matrix, and \( A, B \) have respective sizes \( m' \times (k - m') \) and \( m'' \times (k - m') \). Note that the lower part of \( \tilde{G}^T \) given by \( (0 \ B) \) is obtained by adding to each row of \( \tilde{G}''^T \) a linear combination of rows from \( \tilde{G}^T \), in a way that the \( m' \) leftmost columns of \( \tilde{G}''^T \) are zeroed-out. It follows that the statistical properties of \( \tilde{G}''^T \) are inherited by the \( m'' \times (k - m') \) sub-matrix \( B \), whose entries are hence picked with uniform probability in \( \mathbb{F}_q \). The system is solvable if and only if \( B \) is full rank, i.e. if and only if \( \text{rank}(B) = k - m' \).

Suppose now that the encoded symbols \( e \) are sent to a receiver over an erasure channel which erasure probability of \( 1 - e \). The probability that at least \( k \) symbols out of the \( n \) symbols produced by the linear block code encoder are received is given by
\[
Q^*(\epsilon) = \sum_{i=k}^{n} \frac{n}{i} (1 - e)^i e^{n-i}. \tag{6}
\]

Hence, with a probability \( P^*(\epsilon) = 1 - Q^*(\epsilon) \) the receiver would need to collect symbols encoded by the \text{LRFC} encoder to recover the source block. Assuming that the user collects \( m = k + \delta \) symbols, out of which only \( m' < k \) have been produced by the linear block encoder, the conditional decoding failure probability can be expressed as
\[
\Pr(F|m', m' < k, \delta) = \Pr(\text{rank}(B) < k - m'). \tag{7}
\]

Note that \( B \) is a \( m'' \times (k - m') \) \( (k - m') \times (k - m') \) random matrix, i.e. a random matrix with \( \delta \) equations in excess w.r.t. the number of unknowns. We can thus replace \( (1) \) in \( (7) \), getting the bounds
\[
q^{-\delta-1} \leq \Pr(F|m', m' < k, \delta) \leq \frac{1}{q-1} q^{-\delta}. \tag{8}
\]

We remark that, thanks to the independency of the bounds in \( (1) \) from the size of the random matrix (i.e., the bounds depend only on the overhead), we can remove the conditioning on \( m' \) from \( (8) \), leaving
\[
q^{-\delta-1} \leq \Pr(F|m' < k, \delta) \leq \frac{1}{q-1} q^{-\delta}. \tag{9}
\]

The final failure probability can be written as
\[
P_F(\delta, \epsilon) = \Pr(F|m' < k, \delta) \Pr(m' < k) + \Pr(F|m' \geq k, \delta) \Pr(m' \geq k), \tag{9}
\]
where \( \Pr(F|m' \geq k, \delta) = 0 \) and \( \Pr(m' < k) = P^*(\epsilon) \).

It results that
\[
P^*(\epsilon) q^{-\delta-1} \leq P_F(\delta, \epsilon) < P^*(\epsilon) \frac{1}{q-1} q^{-\delta}. \tag{10}
\]

From an inspection of \( (1) \) and \( (10) \), one can note how the bounds on the failure probability of the concatenated scheme are scaled down by a factor \( P^*(\epsilon) \), where \( P^*(\epsilon) = \sum_{i=0}^{k-1} \binom{n}{i} (1 - \epsilon)^i \epsilon^{n-i} \) is a monotonically increasing function of \( \epsilon \). It follows that, when the channel conditions are \textit{bad} (i.e., large \( \epsilon \) \( P^*(\epsilon) \to 1 \), and the bounds in \( (10) \) tend to coincide with the bounds in \( (1) \). When the channel conditions are \textit{good} (i.e., small \( \epsilon \)), most of the time \( m' \geq k \) symbols produced by the linear block encoder are received, leading to a decoding success (recall the assumption of \textbf{MDS} code). In these conditions, \( P^*(\epsilon) \ll 1 \), and according to the bounds in \( (10) \) the failure probability may scale down even of several orders of magnitude.

Fig. 2 shows the probability of decoding failure as a function of the number of overhead symbols for a concatenated code built using a \( (11, 10) \) \text{SPC} code in \( \mathbb{F}_2 \). It can be observed how, for lower erasure probabilities, the performance gain in terms of probability of decoding failure increases. For \( \epsilon = 0.01 \) the decoding failure probability is more than 2 orders of magnitude lower. Fig. 3 shows the probability of decoding failure vs. the number of overhead symbols for the concatenation of a \( (15, 10) \) \text{RS} and a \text{LRFC} over \( \mathbb{F}_{10} \). The performance of the concatenated code is compared with that of the \text{LRFC} built on the same field for different erasure probabilities. In this case the decrease in terms of probability of decoding failure is bigger than in for the previously presented code in \( \mathbb{F}_2 \). For
a channel with an erasure probability $\epsilon = 0.05$, the probability of decoding failure of the concatenated scheme is 4 orders of magnitude lower than for the LRFC.

The analysis provided in this section is also valid if the LRFC is substituted by a LT or Raptor code. In order to calculate the performance of such a concatenated code one has to substitute in (9) the term $\Pr(F|m' < k, \delta)$ by the probability of decoding failure of the LT or Raptor code. Again the failure probability of the concatenated scheme is scaled down by a factor $P^*(\epsilon)$, where $P^*(\epsilon) \leq 1$.

IV. Numerical Results

Fig. 4 shows the results of simulations together with the bounds calculated using (10). In this case a (15, 10) RS was concatenated with a LRFC in $\mathbb{F}_{16}$, and a channel with an erasure probability $\epsilon = 0.1$ was used. It can be seen how the simulation results match the analytical results down to a probability of decoding failure of $10^{-7}$. Fig. 5 shows the simulation results for a concatenated code using a (11, 10) parity check code in $\mathbb{F}_2$, and a channel with an erasure probability $\epsilon = 0.1$. It can be seen how the simulation results match the analytical results again. However, in $\mathbb{F}_2$ the bounds are less tight than in higher order Galois fields.

An assessment the performance of the concatenated scheme in a system with a high number of users has been performed, assuming a system in which a transmitter sends a source block to a set of $N$ receivers. We considered the erasure channels from the transmitter to the receivers to be independent, with an identical erasure probability $\epsilon$. Furthermore, we assumed that the receivers send an acknowledgement to the transmitter when they have successfully decoded the block. Ideal (error-free) feedback channels have been considered. When all receivers have sent an acknowledgement, the transmitter stops encoding redundant symbols for the source block.

If $k + \Delta$ (where $\Delta$ denotes the transmitter overhead) symbols have been transmitted, the probability that a specific receiver gathers exactly $m$ symbols is:

$$P_R\{k + \Delta, m\} = \binom{k + \Delta}{m}(1 - \epsilon)^m \epsilon^{k + \Delta - m}$$

(11)

The probability of decoding failure at the receiver given that the transmitter has sent $k + \Delta$ symbols is hence

$$P_e = \sum_{m=0}^{k-1} P_R\{k + \Delta, m\} + \sum_{m=k}^{k+\Delta} P_R\{k + \Delta, m\} P_F\{\delta = m - k, \epsilon\}.$$

The probability that at least one user has not decoded successfully is thus

$$P_E(N, \Delta, \epsilon) = 1 - (1 - P_e)^N$$

(12)

Using the bounds in (10) $P_E(N, \Delta, \epsilon)$ can also be bounded. In the following we provide an example to assess the performance of the new scheme in comparison with LRFC codes and also with an idealized fountain code. We assume a system with $N = 10^4$ users and a channel with an erasure probability $\epsilon = 0.01$. The performance of LRFC codes over $\mathbb{F}_2$ and $\mathbb{F}_{16}$ is shown as well as that of two concatenated schemes: a concatenation of a (11, 10) SPC code with a LRFC code in $\mathbb{F}_2$, and a concatenation of a (15, 10) RS code and a LRFC code over $\mathbb{F}_{16}$. It can be seen how the concatenated scheme in $\mathbb{F}_2$ outperforms the LRFC constructed on the same Galois field. For example, for $P_E = 10^{-4}$ the concatenated scheme in $\mathbb{F}_2$ needs only $\Delta = 20$ overhead symbols whereas the LRFC needs 27 (Fig. 6). In the case of the fountain codes operating in $\mathbb{F}_{16}$, the concatenated code shows a performance very close to that of an idealized fountain code.

V. Conclusions

A novel fountain coding scheme has been introduced. The scheme consists of a parallel concatenation of a MDS block
code with a LRFC code, both constructed over the same field, $\mathbb{F}_q$. The performance of the concatenated fountain coding scheme has been analyzed through derivation of tight bounds on the probability of decoding failure as a function of the overhead. It has been shown how the concatenated scheme performs as well as LRFC codes in channels characterized by high erasure probabilities, whereas they provide failure probabilities lower by several orders of magnitude at moderate/low erasure probabilities.

VI. ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Enrico Paolini, Dr. Francesco Rossetto and Giuliano Garrammone for the useful discussions.

REFERENCES

[1] J. Byers, M. Luby, and M. Mitzenmacher, “A digital fountain approach to reliable distribution of bulk data,” IEEE Journal on Selected Areas in Communications, vol. 20, no. 8, pp. 1528–1540, Oct. 2002.
[2] J. Metzner, “An improved broadcast retransmission protocol,” IEEE Transactions on Communications, vol. 32, pp. 679–683, June 1984.
[3] G. Liva and C. Kissling, “Verfahren zur broadcasting-Übertragung von in datenpaketen angeordneten informationen,” Deutsches Patent- und Markenamt, Patent application 10 2008 003 588.2, Jan. 2008.
[4] J. K. Sundararajan, D. Shah, and M. Medard, “ARQ for Network Coding,” in ISIT’08: Proceedings of the 2009 IEEE International Conference on Information Theory, July 2008, pp. 1651–1655.
[5] S. Sorour and S. Valaee, “Adaptive network coded retransmission scheme for wireless multicast,” in ISIT’09: Proceedings of the 2009 IEEE International Conference on Information Theory, 2009.
[6] M. Luby, “LT codes,” in Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on, 2002, pp. 271 – 280.
[7] M. Shokrollahi, “Raptor codes,” IEEE Transactions on Information Theory, vol. 52, no. 6, pp. 2551–2567, June 2006.
[8] D. J. C. MacKay, “Fountain codes,” IEE Proc.-Commun., vol. 152, pp. 1062–1068, 2005.
[9] G. Liva, E. Paolini, and M. Chiani, “Performance versus overhead for fountain codes over $\mathbb{F}_q$,” IEEE Comm. Letters., vol. 14, no. 2, pp. 178–180, 2010.