CCIT Report #865 July 2014



# On-Line Fountain Codes with Low Overhead

Yuval Cassuto, *Member, IEEE,* and Amin Shokrollahi, *Fellow, IEEE*

**Abstract**

An on-line fountain code is defined as a fountain code for which an optimal encoding strategy can be found efficiently given any instantaneous decoding state. This property is important for data distribution in practical networks. In this paper we formalize the problem of on-line fountain code construction, and propose new on-line fountain codes that outperform known ones in having factor 3-5 lower redundancy overhead. The bounding of the code overhead is carried out using analysis of the dynamics of random-graph processes.

**Index Terms**

Fountain codes, rateless codes, on-line codes, codes with feedback, random graphs.

## I. Introduction

Fountain coding was proposed [3] for efficient distribution of data in lossy networks, with the goal to allow information transmission that is oblivious to losses of individual packets. Fountain codes that meet this goal with negligible overhead were found [10], and later improved in complexity [11]. These works, and others that followed, specify methods to encode (and decode) data blocks at the sender (and receiver) nodes, and prove upper bounds on the average overhead in the case of random losses.

Low overhead is clearly an important code-design objective, but some applications may find it insufficient on its own, with system performance being dominated by other properties of the code. The long code blocks and fixed pre-defined encoding procedures of the aforementioned fountain codes result in high decoding latency, and no way for the receivers to control or even monitor the progress of the data reception. In addition, packet losses in the network may not be all random, further exacerbating the risks of a long batch transmission designed for pure-random losses. A practical fountain code should thus balance low redundancy overhead with an *on-line* ability to adapt the code to instantaneous network conditions.

In the framework developed in this paper, a fountain code is called *on-line* if given an arbitrary decoding state at the receiver, the optimal encoding strategy at the sender can be found efficiently. This is a much stronger property than conventional fountain codes, which only guarantee optimality for the initial state of decoding. The importance of the on-line property is that it guarantees optimal performance even at states that differ significantly from the expected under random losses, e.g. due to an adversary or extremely unfortunate circumstances. As it turns out, there are known fountain codes with the on-line property such as *growth codes* [8] and *real-time oblivious codes* [2], (see also [5]). However, these codes suffer from very high redundancy overheads.

The codes proposed here attain the on-line property with significantly lower overhead, between factor $3$ and factor $5$ lower compared to the previously known codes [8],[2]. This significant improvement is achieved by representing the decoding state as a *uni-partite* graph, and using the graph structure to efficiently find the optimal coding strategy at the current state. The simple graph representation also allows to analyze the coding overhead, building on fundamental results from random-graph theory [6]. The contributions of the paper include the first formalization of the on-line property for fountain codes (Section II), the development of the uni-partite view of fountain codes (Section III), an on-line fountain code construction using an efficient algorithm to find the optimal coding strategy (Section IV), and bounding the overhead of a simplified construction by analyzing the dynamics of random graphs (Section V).

There are many potential applications for on-line fountain codes. One is data distribution in the presence of intermittent adversaries, which bring receivers to arbitrarily bad decoding states and then leave them to recover. Another important application is for distributed storage, where code symbols are distributed among multiple nodes, and the on-line property gives sufficient transparency to control the long-term recoverability of the data given an instantaneous node-failure state.

## II. On-Line Fountain Codes

Block rateless fountain codes, such as LT [10] and Raptor [11] codes, are designed for *batch* transmission of coded symbols. These codes aim at minimizing the number of coded symbols required to successfully decode the entire code block with high probability. Given a current state of code symbols already received at a receiver node, batch fountain codes do not address the minimization of code-symbol transmissions until complete decoding. It has been recognized in prior work that not considering the current state of already received symbols results in sub-optimal performance in many practical scenarios. Several works

Yuval Cassuto is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa Israel (email: ycassuto@ee.technion.ac.il). Part of the research was done while at the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL).

Amin Shokrollahi is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station 14, CH-1015 Lausanne, Switzerland (e-mail: amin.shokrollahi@epfl.ch).

Part of this work was presented at the 2011 Information Theory Workshop, Paraty Brazil [4].