Nano Communication Networks 23 (2020) 100280

Contents lists available at ScienceDirect

Nano Communication Networks

journal homepage: www.elsevier.com/locate/nanocomnet

Hamming–Luby rateless codes for molecular erasure channels

Zhuangkun Wei [a], Bin Li [b], Wenxiu Hu [a], Weisi Guo [c], Chenglin Zhao [b,*]

[a] School of Engineering, University of Warwick, Coventry CV4 7AL, UK
[b] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[c] School of Aerospace, Transport, and Manufacturing, Cranfield University, Milton Keynes, MK43 0AL, UK

ARTICLE INFO

Article history:
Received 19 April 2019
Received in revised form 18 November 2019
Accepted 20 November 2019
Available online 27 November 2019

Keywords:
Molecular communications
Erasure code

ABSTRACT

Nano-scale molecular communications encode digital information into discrete macro-molecules. In many nano-scale systems, due to limited molecular energy, each information symbol is encoded into a small number of molecules. As such, information may be lost in the process of diffusion–advection propagation through complex topologies and membranes. Existing Hamming-distance codes for additive counting noise are not well suited to combat the aforementioned erasure errors. Rateless Luby-Transform (LT) code and cascaded Hamming-LT (Raptor) are suitable for information-loss, however may consume substantially computational energy due to the repeated uses of random number generator and exclusive OR (XOR). In this paper, we design a novel low-complexity erasure combating encoding scheme: the rateless *Hamming–Luby Transform* code. The proposed rateless code combines the superior efficiency of Hamming codes with the performance guarantee advantage of Luby Transform (LT) codes, therefore can reduce the number of random number generator utilizations. We design an iterative soft decoding scheme via successive cancelation to further improve the performance. Numerical simulations show this new rateless code can provide comparable performance comparing with both standard LT and Raptor codes, while incurring a lower decoder computational complexity, which is useful for the envisaged resources constrained nano-machines.

1. Introduction

Recent developments in molecular communications have attracted widespread interest from synthetic molecular system designers to bio-engineering applications. The application areas vary from precision medicine, Internet-of-Nano-Things (IoNT) [1], and military covert signaling in electromagnetically denied environments. Molecular communications modulate information into the physical or chemical properties of molecules, which are then undergo a combination of diffusion and advection propagation. The stochastic nature of the propagation can yield distinctive advantages when faced with complex obstacle channels [2]. Whilst macro-scale prototypes have demonstrated the feasibility of molecular communications using concentration-encoded molecule symbols, energy scarce nano-scale systems are likely to use very few molecules to represent each information symbol, which provides a substitute for the macro-case and can be used for short information (e.g., the control signals) transmission. For the nano-scale systems, there is a high likelihood of both transposition and erasure errors.

1.1. Current state-of-the-art

1.1.1. Molecule-rich macro-system

For the molecule-rich macro-systems, each symbol or bit is represented by a large number of molecules (e.g. concentration shift keying — CSK [3]). Thus, the decoding process is mainly affected by additive counting noise and inter-symbol-interference (ISI). Usually, the counting noise follows a Normal distribution [4, 5]. In such cases, classical Hamming-distance based forward error correction (FEC) codes can be applied [6]. For example, the linear-group Hamming code, low density parity check (LDPC) code and cyclic Reed–Muller (C-RM) code have been applied to combat additive noise sources [7]. Note that, the computation burden in the decoding algorithm has attracted research focus, in order to reduce the burden on bio-molecular machines (i.e., the adenosine triphosphate (ATP) expenditure is used as a unit of logic operation measurement) [8–11].

1.1.2. Molecule-scarce nano-system

For molecule-scarce nano-systems, the nano-devices are far more likely to transmit a small burst of a molecule cluster (due to finite reservoir) and the information is likely to be encoded in a simple (existence or no existence) 1 bit format, or in the composition of the chemical compound (such as DNA [12]) for

* Corresponding author.
E-mail address: clzhao@bupt.edu.cn (C. Zhao).

https://doi.org/10.1016/j.nancom.2019.100280
1878-7789/© 2019 Published by Elsevier B.V.

higher data rate [13]. In either case, the erasure channel applies, where the molecule cluster is either received or not. Such information-loss is due to the physical absorption or entrapment (see porous media bio-membrane example [14]) or due to the biochemical interactions (see bacteria consumption of organic compounds [15] or enzyme reactions [16]). In such a case, which is not extensively researched, the information demodulation will be dominated by each molecule's arrival process. For example, transposition errors may lead to intra- or inter-codeword bit disorder and Hamming-weight codes have been proposed to combat such errors [17,18].

One type of error has not been extensively considered and that is the probability of erasure [19]. A molecule may be *erased* by a number of mechanisms, such as [20,21]: (i) absorption by another unintended bio-receiver (i.e. bacteria), (ii) chemical reaction with another molecule (i.e., nano-scale binding interactions governed by the Lennard-Jones potential), and (iii) trapped in complex porous media structures (common in cell tissues) and other surface deposition effects [22]. If these effects occur, the traditional linear-group FEC codes may perform poorly as the check-matrix cannot combat information erasure effectively.

Rateless erasure codes (or fountain codes) allows the original source symbols k to be recovered from any subset of the codeword k', such that $k/k' \approx 1$ in optimal cases. These codes offer the advantage of overcoming the coupon collector's problem, in that *only any subset* of the encoding symbols k' needs to be received, without disorder. These codes have no fixed rate and Luby transform (LT) codes are the first practical realization of fountain codes, with further performance improvements later on in the form of Raptor codes. Whilst LT codes can guarantee erasure channel performance, it uses a large number of random number generator and exclusive OR (XORs), especially the first of which is at the cost of substantially computational resources for energy-restricted molecular communications. On the other hand, Hamming codes can provide check bits without the uses of the random number generator, but cannot guarantee a rateless performance in the face of excessive erasure. By incorporating both the characteristics of rateless LT codes and Hamming codes, a novel rateless code is therefore designed, which has a lower overhead and is able to combat excessive loss.

1.2. Contribution

In this paper, we focus on a molecule-scarce nano-system and the erasure channel. The code is advantageous for nano-scale systems due to its low complexity, and particularly effective for non-concentration encoded molecular communication channels that face erasure noise. In order to adapt Hamming codes to combat the aforementioned challenge, we proposed a new rateless linear-group FEC code. In general, the main contributions are summarized as follows:

(1) We introduce extra constraint bits into linear-group codes, generated via LT codes. Such extra output bits are directly attached to the Hamming code, and thereby a new rateless code with the more powerful FEC ability is constructed. Also, by incorporating Hamming and LT codes, we reduce the number of the random number generator utilizations.

(2) We design an effective soft decoding scheme, relying on the conception of successive cancelation (SC). Rather than the check matrix based hard-decision decoding, in our new scheme the output bit is recovered successively one by one. In this manner, the information can be decoded via other extra constraint bits, even single molecular messenger is erased.

(3) We evaluate the new rateless code in the presence of erasure diffusion channels. We show by numerical simulations that our new constructed code outperforms the LT and the Raptor codes when the code length is short, and hence is of promise to nano-scale molecular communications, whereby the molecular resources are restricted while the diffusive channels are characterized by certain erasure probabilities.

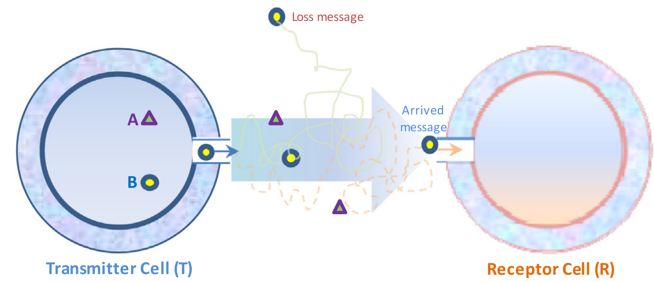

Fig. 1. Molecule-scarce model where each bit from transmitter cell T, towards receptor cell R is represented by single molecule. Two types molecules, *A* and *B*, are assumed to carry different bit information.

The rest of this article is structured as follow. The system model considered by this work is shortly depicted in Section 2. Then, in Section 3 the new rateless code is constructed, including the basic principle of encode and decode. In Section 4, numerical simulations are then provided to validate our designed code. Finally, we conclude this study in Section 5.

2. Channel model

We study the bio-inspired molecule-scarce signaling scenario whereby two cells (T and R) are communicating with each other, via GF(2) data set $\{0, 1\}$. That is to say, sending a smaller number of molecules A represents one bit of 0 (e.g. cell proliferation), whist sending another molecules B represents one bit of 1 (e.g. cell death). Different from the molecule-rich macro-systems, in molecule-scarce scenarios, the molecules emitted from the transmitter T towards the receptor R, may be lost due to the complex diffusive and reactive processes [23]. An illustration of the molecule-scarce scenarios is shown by Fig. 1.

In order to characterize the information-loss scenarios, the binary erasure channel (BEC) model is adopted, where the erasure probability ε specifies the loss probability of one molecule. Such erasure probability is related with both the property of the biological transceivers (e.g., the size of the receptor region), and the channel characteristics (such as the propagation distance, and the reactive substances). In this work, different cases of erasure probabilities are simulated via the Monte-Carlo simulator of the two close-set communicating cells in Fig. 1. Specially, we assume the propagation of each molecular messenger is independent and identically distributed (i.i.d.), as the any two emitted molecules will not interact with each other.

The aim of this paper then can be summarized as the design of a rateless code in order to combat the information-loss, caused by the molecule-scarce communication scenarios with complex and absorbing channel environment. It is also noteworthy that such code implementation should consider the energy-restricted property of the molecular communications, which suggests that most of the soft-decision algorithms such as likelihood computations may by infeasible. In the following, we will elaborate the designs of the Hamming-LT code and the SC decoder.

3. Rateless Hamming-LT codes

3.1. Novelty

As aforementioned, the traditional linear-group code can reduce transmission error in erasure channels to some extent. However, this become less effective in nano-systems due to the absence of higher-layer check mechanisms to cope with the loss of