Hamming-Luby Rateless Codes for Molecular Erasure Channels

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

[a]*School of Engineering, University of Warwick, UK*
[b]*School of Information and Communication Engineering, BUPT, China*

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 maybe lost in the process of diffusion-advection propagation through complex topologies and membranes. Existing work has considered Hamming-distance codes for additive counting noise and Hamming-weight codes for transposition errors. Current linear-group codes are not well suited to combating the aforementioned erasure errors. 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. We design an iterative soft decoding scheme via successive cancelation to further improve the performance. Numerical simulations show this new rateless code can outperform both standard LT and cascaded Hamming-LT (Raptor) codes, while incurring a lower decoder computational complexity, which is useful for the envisaged resources constrained nano-machines.

Keywords: Molecular communications, erasure code.

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 signalling 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].

---

*Corresponding author
Email address: Zhuangkun.wei@warwick.ac.uk (Zhuangkun Wei)

September 7, 2020

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. In the latter case of using a small number of molecules to convey information, 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, 9, 10, 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 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 realisation of fountain codes, with further complexity improvements later on in the form of Raptor codes. Whilst LT codes can guarantee erasure channel performance, it comes at the cost of a relatively large overhead. On the other hand, Hamming codes can perform equally well with a smaller overhead, 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 summarised 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.

(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 can outperform other codes (even the cascaded Hamming and LT Raptor code), and hence is of promise to nano-scale molecular communications, whereby the molecular resources are restricted while the diffusive channels are characterised by certain erasure probabilities.

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

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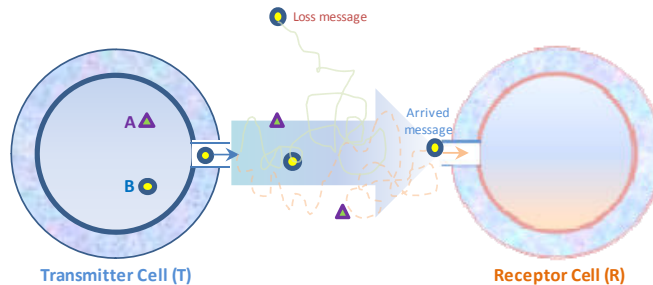
Figure 1: Molecule-scarce model where each bit from transmitter cell `T`, towards receptor cell `R` is represented by single molecule. Two types molecules, A and B, are assumed to carry different bit information.

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 characterise 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 monto-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 information. In this work, we design a new code scheme to improve the
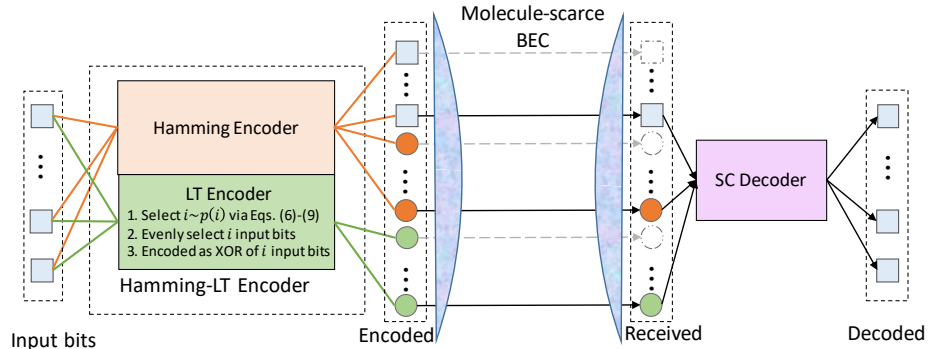
Figure 2: Illustration of the proposed Hamming-LT scheme for the molecule-scarce communication, in which the input bits are encoded by the Hamming-LT encoder, which are then transmitted via the BEC channel. The receiver cell will receive the molecules, and the input information can be decoded via the SC decoder.

transmission performance in erasure channels, known as the rateless Hamming-LT code. To be specific, the new rateless Hamming-LT code is a constructive combination of the Hamming and LT code, whereby extra constraint bits are generated via the LT code and then directly attached to a Hamming code, see the structure in Fig. 2.

It is noteworthy that such direct combination of the proposed Hamming-LT code gives rise to two merits for molecule-scarce communication applications. First, the proposed Hamming-LT code can combat the information-loss in molecular-scarce communication scenarios, attributed to the combining advantage of the Hamming and LT erasure code. We will illustrate it in Figs. 6-7. Second, the proposed Hamming-LT scheme is able to save the energy by reducing the overheads when comparing with the traditional LT and the Raptor code. Such reduced overheads leads to the less computation energy used for both the encoding and the decoding processes, as the lower number of check-bits requires less number of operations (e.g., the random number generator, and the XOR operator). This is significant for the designs of the man-made molecular transceivers where the energy/resources are restricted, since the use of random number generator will cost large amounts of energy. We will analyze the overhead in detail in Section 3.2.3.

3.2. Encoding Scheme

The encoding process illustrated in Fig. 2 can be pursued via the generator matrix $\mathbf{G}_{HL}$, i.e.,

$$\mathbf{G}_{HL} = \begin{bmatrix} \mathbf{G}_{ham} \mid \mathbf{G}_{LT} \end{bmatrix}, \tag{1}$$

where $\mathbf{G}_{ham}$ and $\mathbf{G}_{LT}$ denote the generator matrices of Hamming and LT codes, and will be elaborated in Section 3.2.1 and Section 3.2.2 respectively. As such,

5

given $\mathbf{u}$ as the input bit vector, the Hamming-LT encoder can be implemented as:

$$\mathbf{x}_{ham-lt} = \mathbf{G}_{ham} \cdot \mathbf{u}. \tag{2}$$

3.2.1. Sub-Hamming Generator Matrix

The Sub-Hamming generator matrix is derived from the Hamming linear-group FEC whose code length is $N = 2^M - 1$, with M check bits and $K = N - M$ input bits. The check matrix, denoted by $\mathbf{H}_{ham} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_N]^T$ whose elements $\mathbf{h}_j, j \in \{1, 2, ..., N\}$ disorderly enumerate all M-dimension $GF(2)$ vectors but $\mathbf{0}$, can be expressed as:

$$\mathbf{H}_{ham} = \begin{bmatrix} \mathbf{C}_{M \times K} | \mathbf{I}_{M \times M} \end{bmatrix}^T, \tag{3}$$

where $\mathbf{I}_{M \times M}$ is an $M \times M$ identity matrix, and $\mathbf{C}_{M \times K}$ is an $M \times K$ matrix consists of all M-dimension $GF(2)$ vectors whose supports are larger than 1. The generator matrix of the (N, K) Hamming code is then given as:

$$\mathbf{G}_{ham} = \begin{bmatrix} \mathbf{I}_{K \times K} | \mathbf{C}_{M \times K}^T \end{bmatrix}^T, \tag{4}$$

where $\mathbf{C}_{M \times K}^T$ is the transpose of matrix $\mathbf{C}_{M \times K}$.

3.2.2. Sub-LT Generator Matrix

The sub-LT generator matrix is derived from the encoder process of the LT code. The main advantage of the LT code is attributed to its flexible code length, and the randomness of the checking packets that embrace the information bits. Such randomness contains two aspects. First, the selected number of input bits (denoted as i) for each packet is random. Second, the selection of the input bits is also random. For the number of selected bits (denoted as i), it follows the *Robust Soliton Degree distribution*: [24]:

$$i \sim p(i) = \frac{\rho(i) + \tau(i)}{\sum_{i=1}^{K} [\rho(i) + \tau(i)]}, \tag{5}$$

where $\rho(i)$ is the ideal Soliton distribution of i [24], i.e.

$$\rho(i) = \begin{cases} \frac{1}{K} & i = 1, \\ \frac{1}{i(i-1)} & i = 2, 3, ..., K, \end{cases} \tag{6}$$

and $\tau(i)$ is designed to improve the robustness of an ideal Soliton distribution [24], i.e.,

$$\tau(i) = \begin{cases} \frac{R}{K} \frac{1}{i}, & i = 1, 2, ..., \frac{K}{R} - 1, \\ \frac{R}{K} \ln\left(\frac{R}{\delta}\right), & i = \frac{K}{R}, \\ 0, & i > \frac{K}{R}, \end{cases} \tag{7}$$

with

$$R = v\sqrt{K} \ln K / \delta. \tag{8}$$

In Eqs. (5)-(8), δ is the allowable failure probability of the decoder that can fully recover the total input bits for a given number $K \cdot \sum_{i=1}^{K} [\rho(i) + \tau(i)]$ of encoded signals, and R is referred as the expected *ripple size*, i.e., the average number of different 1-degree signals before $K - R$ input bits have been decoded. v is a small constant.

After the determination of degree distribution from above Eqs. (5)-(8), the generator matrix can be derived as $\mathbf{G}_{LT} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_\Omega]^T, \Omega > K$, of which $\mathbf{g}_n$ $(n = 1, 2, ..., \Omega)$ is a K-dimension vector and each position of its 1-elements represents the corresponding selected index of one input bit.

It is noteworthy that given the input length as K, the proposed Hamming-LT code may have limitless code length (also known as overheads), due to the flexible LT sub-code length Ω. In the context of the molecule-scarce and energy-limited communication scenarios, the absence of understanding how long such overhead can ensure a high recovery performance may lead to the waste of the energy utilization on the redundancy encoding. This is extremely significant for the sub-LT code, as the random number generator for degree generator will cost large amounts of computation and storage resources. This raises the necessity of the overhead analysis for the proposed Hamming-LT code in the following.

3.2.3. Overhead Analysis

The definition of overhead can be described as the **minimum** redundances of Hamming check bits (M) and LT extra constraint bits (Ω), in order to recover the total input bits with a relatively high probability. Intuitively, the larger this overhead, the higher the probability of successful recovery, but at the expense of higher computation resources. For the Hamming sub-code in our rateless Hamming-LT code, it is easily understood that the K input bits can be completely recovered if N Hamming signals have been totally received. This indicates that the overhead for Hamming sub-code is thereby

$$M = \mathrm{O}(\log K), \tag{9}$$

according to its redundances $M = N - K$.

We then analyze the overheads for the sub-LT code. The probability of successful recovery taken from sub-LT code depends on the number of received 1-degree signals, after m $(m = 0, 1, ..., K - 1)$ input bits have been successfully decoded. Denote $s(\Omega', m)$ as the number of 1-degree signals, where $\Omega' \leq \Omega$ is the total number of received LT signals, and m is the number of decoded input bits. Considering the common iterative decoding scheme, $s(\Omega', m)$ can be recursively computed by its predecessor $s(\Omega' - 1, m)$ as follows:

$$s(\Omega', m) = s(\Omega' - 1, m) + \Delta(m), \tag{10}$$

where the additive term

$$\Delta(m) = \sum_{i=1}^{m+1} p(i) \cdot \frac{C_m^{i-1} \cdot (K - m)}{C_K^i}, \tag{11}$$

7

can be interpreted as an increment of the number of input bits from a new received LT encoded signal. Consider $m = K - 1$, this increment is $\Delta(K-1) = \sum_{i=1}^{K} p(i) \cdot i / K$, and $s(\Omega', K-1)$ can be re-written as:

$$
\begin{aligned}
s(\Omega', K-1) &= \Omega' \cdot \frac{1}{K} \sum_{i=1}^{K} p(i) \cdot i, \\
&= \Omega' \cdot \frac{\ln K / \delta}{K}, \qquad (12)
\end{aligned}
$$

in the sense that $s(0, K-1) = 0$. Note that, $s(\Omega', K-1)$ should be no lesser than 1, and $\mathbb{E}(\Omega') = (1-\varepsilon)\mathbb{E}(\Omega)$ with $\mathbb{E}$ the expectation, hence the redundancy of LT sub-code Ω should meet:

$$
\Omega = O\left(\frac{K}{\ln K}\right) \qquad (13)
$$

Finally, by taking the above two sub-codes together, we can compute the overhead of the proposed rateless Hamming-LT code, denoted as Θ, and prove such overhead is lesser than that of the traditional LT and Raptor code, i.e.,

$$
\Theta = M + \Omega = O\left(\ln K + \frac{K}{\ln K}\right) < O\left(K \ln K\right). \qquad (14)
$$

In Eq. (14), the smaller overheads indicates the less energy utilization for the proposed Hamming-LT scheme, and thereby makes it a potential candidate for the molecule-scarce and energy-limited communication scenarios.

3.3. SC Decoding Algorithm

In nano-scale molecular communication scenarios where the complicated mathematical operations (such as division and exponent computation) are infeasible, in terms of computation and energy consumption [25]. Thus, most likelihood-based soft-decision decoding algorithms, e.g. belief propagation (BP), may become less attractive in such applications.

One intuitive solution is to resort to the hard-decision decoding algorithm, which uses only the simple XOR operations and leverages on the minimal Hamming distance, i.e. d_{min}. However, such decoder puts strict requirements on the errors in received signals, whereby the maximum length of FEC is $(d_{min}-1)/2$. As far as the random diffusion propagation is concerned, most existing linear-group codes may become insufficient to correct the error signals, leading to the high failure probability. To be specific, given an erasure probability ε (< 0.5), the failure probability of hard-decision decoding of the Hamming-LT code, denoted as ζ_{hd}, is equivalent with the probability of losing no lesser than 2 encoded signals (or messengers), i.e.,

$$
\zeta_{hd} = \sum_{l=2}^{N_s} C_{N_s}^{l} \varepsilon^l (1-\varepsilon)^{N_s - l}. \qquad (15)
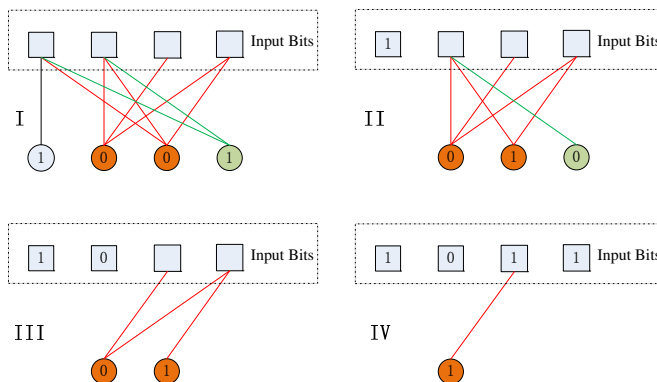$$

8

Figure 3: SC decoding algorithm, where one iterative step firstly finds 1-degree signal and directly decodes its correspondingly input bit, secondly cancels this decoded 1-degree signal, thirdly operates XOR on this input bit with signals that link with this input bit.

As such, when applied to the concerned BEC channel, the hard-decision decoding scheme will be greatly degraded by the missing of information bits.

To cope with this challenge, we design a novel iterative decode algorithm by fully considering the erasure nature of molecular channels. The main concept of our SC decoding scheme is to successively cancel the constraint relation with output bits, i.e. recovering the information successively one by one. Taking the simple $K = 4$ Hamming-LT code for instance (as in Fig.3), as receiving $\mathbf{x}_{HL} = [1, 0, 0, 1]^T$, which are respective encoded by u_1, $u_2 \oplus u_3 \oplus u_4$, $u_1 \oplus u_2 \oplus u_4$ and $u_1 \oplus u_2$, where $\oplus$ is denoted as XOR operator, input bits can be fully decoded as $\mathbf{u} = [1, 0, 1, 1]^T$.

In general, this flow of SC decoding process with regards to the generator matrix $\mathbf{G}_{HL}$ is given by **Algorithm.1**. The inputs of the SC-decoder are received signal $\mathbf{x}_{HL}$ as well as its corresponding $\mathbf{G}_{HL}$, i.e., supposing N_r signals have been received to recover K input bits, we then rewrite the generator matrix as $\mathbf{G}_{HL} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{N_r}]^T$, which involves only the rows corresponding to the N_r encoded signals. Steps 1-3 are designed to identify and decode one input bit in encoded signal of 1-degree. Steps 4-6 aim to remove the constraint links between such recovered input bit and the remaining encoded signals. Step 7 is to judge whether the input bits can be fully recovered.

We then compare the theoretical decoding performance between the designed SC decoder with hard-decision one. The failure probability of the SC decoding algorithm will have a loose upper-bound. That is to say, the failure probability

Algorithm 1 SC decoding algorithm
___
Require: $\mathbf{x}_{HL} = [x_1, x_2, ..., x_{N_r}]^T$, $\mathbf{G}_{HL} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{N_r}]^T$
1: Find one vector $\mathbf{g}_\gamma$ whose support is 1 from $\mathbf{G}_{HL}$
2: Directly recover one input bit u_c from received signals $\mathbf{x}_{HL}$ as $u_c = x_\gamma$, where c is the index of 1-element from $\mathbf{g}_\gamma$
3: Update $\mathbf{G}_{HL}$ by deleting this row $\mathbf{g}_\gamma$
4: Find from $\mathbf{G}_{HL}$ rows $\mathbf{g}_{e_1}, \mathbf{g}_{e_2}, ..., \mathbf{g}_{e_j}$, of which the cth element equals to 1
5: Update $\mathbf{x}_{HL}$ by replacing these signals $x_{e_1}, x_{e_2}, ..., x_{e_j}$ with $x_{e_1} \oplus u_c, x_{e_2} \oplus u_c, ..., x_{e_j} \oplus u_c$ respectively, where $\oplus$ is denoted as XOR operator.
6: Update $\mathbf{G}_{HL}$ by assigning cth element of each $\mathbf{g}_{e_1}, \mathbf{g}_{e_2}, ..., \mathbf{g}_{e_j}$ as 0
7: Reprocess 1-6 until $\mathbf{u}$ has been totally recovered, or no row vector from updated $\mathbf{G}_{HL}$ is 1-support.

Ensure: u
___

of the SC decoding algorithm can be expressed as:

$$
\begin{aligned}
\zeta \quad &\overset{(a)}{<} \quad C_{N_s}^2 \varepsilon^2 (1-\varepsilon)^{N_s-2}[1 - C_{N_s-K}^2 / C_{N_s}^2] \\
&\quad + \sum_{l=3}^{N_s} C_{N_s}^l \varepsilon^l (1-\varepsilon)^{N_s-l}, \\
&< \quad \sum_{l=2}^{N_s} C_{N_s}^l \varepsilon^l (1-\varepsilon)^{N_s-l} = \zeta_{hd}.
\end{aligned}
\tag{16}
$$

In the above inequality, (a) holds for the fact that the total K input bits can be recovered successfully, if none of the two lost bits are K information bits from the Hamming sub-code. So, we theoretically prove the designed SC decoder indeed obtains a higher recovery probability than the hard-decision schemes in diffusive channels characterised by random erasure probability.

Another advantage of the designed SC-decoder lies in its avoidance of the statistics and likelihood computations. This is important for the molecular communications, as such statistical operators are difficult to be implemented via the biological circuit. Instead, as is illustrated in Algorithm 1, the designed SC-decoder relies on only the XOR operators, which can be generated via protein phosphorylation/de-phosphorylation cycle, i.e., the cascade cycle [10], illustrated in Fig. 4(a). Similar to the response of the electrical transistor, the cascade cycle responds to the increasing concentration of the input, thereby capable of transiting the states and can be used for the design of logical circuit. Fig. 4(b) is the designed biological XOR circuit that uses the cascade cycles. In our future work, such biological circuit will be used for the implementations of the proposed Hamming-LT schemes for the communications in porous media bio-membrance example [14].
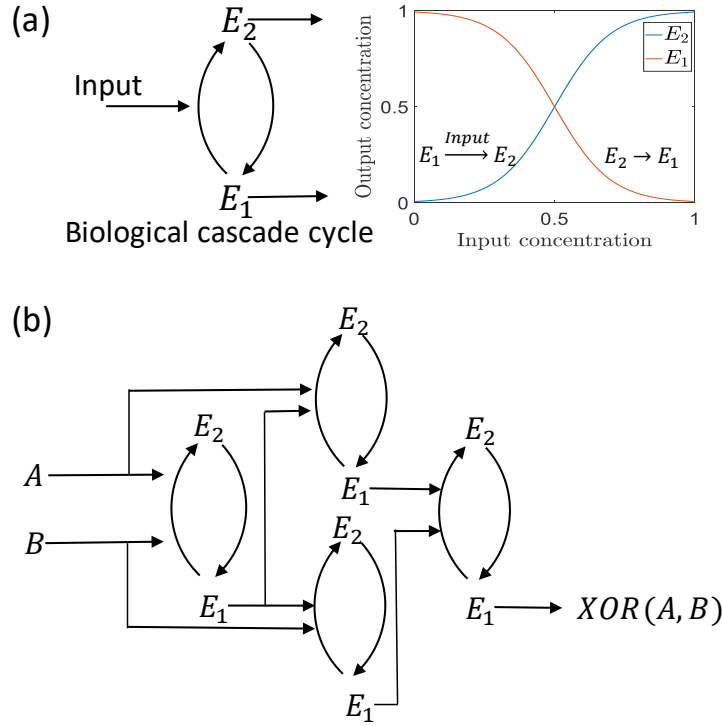
Figure 4: The design of biological circuit where (a) is the cascade cycle that can transit one state to another when an input is injected, and (b) illustrates the designed biological XOR operator leveraged on such cascade circuit.

4. Experimental Simulations

In this section, we evaluate the performance of the proposed rateless Hamming-LT code in molecule-scarce communication scenarios. Various cases of the erasure probabilities are generated by the Monte-carlo simulator via Brownian motion. The two spherical communicating cells with radius 2.25×10^{-7}m are deployed in a 3-D plot. The diffusion rate is assigned as $D = 5 \times 10^{-9}(\text{m})^2/\text{s}$. Then, by varying the communication distances, we create different levels of erasure probabilities, which can be statistically computed via the ratio of the number of received molecules.

Given the different levels of erasure probabilities in molecule-scarce communication, we compare (1) the decoding performance of the novel SC decoding algorithm and the hard-decision scheme, (2) the bit error rate (BER) of the rateless Hamming-LT code and its counterparts, i.e. rateless LT code, and (3) the BER of the rateless Hamming-LT code with the Raptor code (i.e., a cascaded code where Hamming code is employed as the inner code, and LT code is used as the outer code).
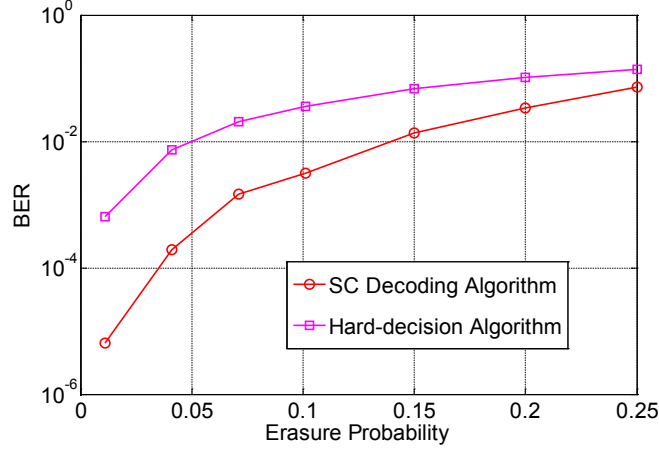
11

Figure 5: BER comparison between SC decoding algorithm and hard-decision algorithm for a $K = 11$, $M = 4$, $\Omega = 4$ Hamming-LT code.

4.1. SC Decoding vs Hard-decision Decoding

For the first experiment, the $(19, 11)$ Hamming-LT code is studied (i.e. $K = 11$, $M = 4$, $\Omega = 4$), respectively using the proposed SC decoding algorithm and the hard-decision decoding scheme. Parameters of the code scheme are configured to $v = 0.15$ and $\delta = 0.8043$. We configure the erasure probability to be a small value, ranging from 0 to 0.25, i.e. $\varepsilon \in (0, 0.25)$. It is shown that, at the lower region of erasure probabilities, the SC decode scheme can significantly outperform the classical hard-decision decoder. For example, for a small erasure probability $\varepsilon = 0.04$, the BER of a hard-decision decoder is around 7×10^{-3}, while the BER of our new scheme can be decreased to 2×10^{-4}. Note that, however, for larger erasure probabilities (e.g. $\varepsilon > 0.2$), these two decoder may become comparable with each other. Thus, numerical simulations further validate our new decoding schemes, which can indeed offer the better BER performance.

Then, we give a rough complexity analysis of the designed SC-decoder. Here, it is noteworthy that such rough analysis considers only the number of XOR operator, as other steps in Algorithm 1 are dependent with the specific biological system and circuits, therefore out of the scope of this work. For the hard-decision decoder, the required number of XOR operations is $M \cdot (N - 2^{M-1})$. For the SC-decoder, the required XOR operations will be also related with the average degree of encode signals, whose upper-bound is $M \cdot (N - 2^{M-1}) + \Omega \cdot \ln K / \delta$. According to the numerical simulations (with 100,000 input bits), the average number of XOR operations is around 2.54×10^5 for the hard-decision scheme. In comparison, such a required number will be slightly increased to 3.50×10^5 as our new SC-decoder is concerned. That is to say, the FEC performance can be remarkably enhanced by our new SC-decoder at the expense of the slightly increased complexity.
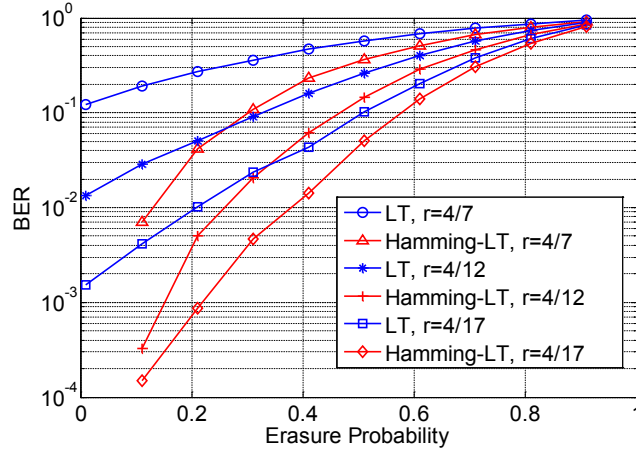
Figure 6: BER comparison among Hamming-LT codes, LT codes and Hamming code, where r is referred as code rate.

4.2. Hamming-LT vs LT Code

In the second simulation, we first consider the special case of our Hamming-LT code, i.e. the rateless Hamming-LT code of $K = 4$ and $\Omega = 0$ may degenerate to a $(7, 4)$ Hamming code. Parameters of the code scheme are configured to $v = 0.3$ and $\delta = 0.5019$. From Fig. 5, we note that the Hamming code will be superior to the LT code, when the SC-decoding scheme and the same code rate (i.e. $r = 4/7$) are used. We then fix the input code length to $K = 4$, and evaluate the decoding performance under various code rates. We can observe from Fig. 5 that the decoder performance of both LT and the new rateless Hamming-LT code will be enhanced by the smaller code rate. More importantly, we note that, as the code rate increases, the decoder BER of our new code scheme will be always smaller than that of an LT code.

4.3. Hamming-LT vs Raptor Code

In Fig. 6, we further compare our new rateless Hamming-LT code with Raptor code, i.e. cascading Hamming code and LT code. In this simulation, parameters of the Hamming-LT scheme are configured to $v = 0.3$ and $\delta = 0.5019$. The coding rate of both Hamming-LT code and Raptor code is $r = 4/17$. We can observe from numerical results that, at the lower region of erasure probability, the Hamming-LT scheme can remarkably outperform the cascaded Raptor scheme. E.g. for a small erasure probability $\varepsilon = 0.2$, the BER of Raptor code is around 2×10^{-2}, while the BER of the Hamming-LT code is decreased to 8×10^{-4}. Note that, in the presence of larger erasure probability region, (e.g. $\varepsilon > 0.6$), however Raptor code and Hamming-LT code are comparable with each other. Thus, the numerical simulation further validate our novel Hamming-LT code, which indeed achieves the better BER performance in diffusive erasure channels.
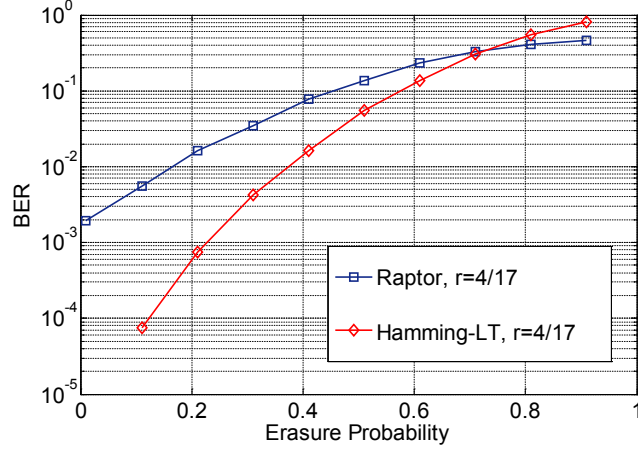
13

Figure 7: BER comparison between Hamming-LT code, and Raptor code (i,.e. the cascaded Hamming code and LT code, where r is referred as code rate.

We also note that the complexity of our new rateless Hamming-LT code is much lower than that of the Raptor code. Here, given that the soft-decision leveraging on the likelihood and posterior computations cannot be easily implemented via the biological circuits, we use the designed SC-decoder for Raptor code in order to make a fair complexity analysis. Measured by the total number of XOR operations, the encoding and decoding implementation complexities of the Raptor code will be $M \cdot (N - 2^{M-1}) + (N + \Omega) \cdot \ln K/\delta$. For the new Hamming-LT code, in constrast, the required XOR operations is $M \cdot (N - 2^{M-1}) + \Omega \cdot \ln K/\delta$. According to the numerical simulations (with 100,000 input bits), the average number of XOR operations is around 1.11×10^6 for the Raptor code, whilst this number can be decreased to 7.45×10^5 for the new Hamming-LT code. That is, compared with the cascaded Hamming and LT code, our proposed new rateless code is superior both in terms of the decoding performance and lower computational complexity.

5. Conclusion

Future bio-inspired nano-systems that communicate using a small number of molecules will require low complexity coding schemes to combat a variety of error sources. In complex in vivo environments, when a small number of molecules are used to represent information, erasure errors can arise due to molecular deposition/binding/entrapment. This raises the demand of a new family of FEC codes. In this study, we designed a novel rateless linear-group code, combining the efficiency benefits of Hamming codes with the performance guarantee of LT codes, in the process enabling flexible output length and low computational complexity. Rather than simply cascading the aforementioned

14

two codes together, our new proposed *Hamming-Luby Transform* code introduces extra constraint bits to combat the missing information bits in decoding input. Furthermore, we designed an iterative successive cancelation (SC) decoding scheme for this rateless linear-group code, which, in contrast to the check-matrix based hard-decision decoders, will recover the information bit one by one relying on the energy-efficient XOR operations. In this manner, even if the input errors surpass the maximum correcting length, successful decoding can be still achieved. Simulation results presented in the paper validate our expectations for the proposed FEC code, and demonstrate that it is very suitable to erasure molecular communication channels and requires only the simple decoding operations. 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 our future work, we will realize the designed Hamming-LT code via the biological circuits, and use it to combat the information-loss in porous media bio-membrane communication applications.

References

[1] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, Y. Koucheryavy, The internet of bio-nano things, IEEE Communications Magazine 53 (3) (2015) 32–40.

[2] W. Guo, C. Mias, N. Farsad, J. L. Wu, Molecular versus electromagnetic wave propagation loss in macro-scale environments, IEEE Transactions on Molecular, Biological and Multiscale Communications 1 (1) (2015) 18–25.

[3] W. Guo, T. Asyhari, N. Farsad, H. B. Yilmaz, B. Li, A. Eckford, C. B. Chae, Molecular communications: channel model and physical layer techniques, IEEE Wireless Communications 23 (4) (2016) 120–127.

[4] S. K. Tiwari, P. K. Upadhyay, Maximum likelihood estimation of snr for diffusion-based molecular communication, IEEE Wireless Communications Letters 5 (3) (2016) 320–323.

[5] Y. Sun, M. Ito, K. Sezaki, Adaptive code width protocol for mitigating intersymbol interference in diffusion-based molecular communication with mobile nodes, in: 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), IEEE, 2016, pp. 1–6.

[6] M. S. Leeson, M. D. Higgins, Error correction coding for molecular communications, in: 2012 IEEE International Conference on Communications (ICC), IEEE, 2012, pp. 6172–6176.

[7] Y. Lu, M. D. Higgins, M. S. Leeson, Comparison of channel coding schemes for molecular communications systems, IEEE Transactions on Communications 63 (11) (2015) 3991–4001.

[8] C. Bai, M. S. Leeson, M. D. Higgins, Minimum energy channel codes for molecular communications, Electronics Letters 50 (23) (2014) 1669–1671.

[9] M. Ş. Kuran, H. B. Yilmaz, T. Tugcu, B. Öerman, Energy model for communication via diffusion in nanonetworks, Nano Communication Networks 1 (2) (2010) 86–95.

[10] H. M. Sauro, B. N. Kholodenko, Quantitative analysis of signaling networks, Progress in Biophysics and Molecular Biology 86 (1) (2004) 5–43.

[11] M. Osgood, K. Ocorr, A. Lehninger, The absolute, ultimate guide to lehninger principles of biochemistry: Study guide and solutions manual.

[12] J. Wang, B. Yin, M. Peng, Diffusion based molecular communication: principle, key technologies, and challenges, China Communications 14 (2) (2017) 1–18.

[13] D. T. McGuiness, S. Giannoukos, A. Marshall, S. Taylor, Experimental results on the open-air transmission of macro-molecular communication using membrane inlet mass spectrometry, IEEE Communications Letters 22 (12) (2018) 2567–2570.

[14] Y. Fang, W. Guo, M. Icardi, A. Noel, N. Yang, Molecular information delivery in porous media, IEEE Transactions on Molecular, Biological and Multi-Scale Communications 4 (4) (2018) 257–262.

[15] S. Qiu, N. Farsad, Y. Dong, A. Eckford, W. Guo, Under-water molecular signalling: A hidden transmitter and absent receivers problem, in: 2015 IEEE International Conference on Communications (ICC), 2015, pp. 1085–1090.

[16] A. Noel, K. C. Cheung, R. Schober, Improving receiver performance of diffusive molecular communication with enzymes, IEEE Transactions on NanoBioscience 13 (1) (2014) 31–43.

[17] P. Shih, C. Lee, P. Yeh, K. Chen, Channel codes for reliability enhancement in molecular communication, IEEE Journal on Selected Areas in Communications (JSAC) 31 (12) (2013) 857–867.

[18] W. Haselmayr, S. Aejaz, A. Asyhari, A. Springer, W. Guo, Transposition errors in diffusion-based mobile molecular communication, IEEE Communications Letters 21 (9) (2017) 1973–1976.

[19] H. Egashira, J. Suzuki, J. Mitzman, T. Nakano, H. Fukuda, Robust directional-diffusive hybrid molecular communication with parity-check erasure coding, in: 2014 IEEE Soft Computing and Intelligent Systems, IEEE, 2017, pp. 1–6.

[20] Kolodny, M. Gerald, Cell-to-Cell Communication, Plenum Press, 1987.

[21] T. Maruyama, Cell-to-cell communication properties of auerbach's plexus cells in culture, Brain Research 336 (2) (1985) 368–71.

[22] M. Icardi, G. Boccardo, D. Marchisio, T. Tosco, R. Sethi, Pore-scale simulation of fluid flow and solute dispersion in three-dimensional porous media, Physical Review E 90.

[23] R. Albert, Scale-free networks in cell biology, Journal of Cell Science 118 (Pt 21) (2005) 4947–4957.

[24] M. Luby, Lt codes, in: The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings., IEEE, 2002, pp. 271–280.

[25] B. Li, M. Sun, S. Wang, W. Guo, C. Zhao, Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications, IEEE Transactions on Communications 64 (5) (2016) 2079–2091.