Channel Codes for Reliability Enhancement in Molecular Communication

Po-Jen Shih, *Student Member, IEEE,* Chia-Han Lee, *Member, IEEE,* Ping-Cheng Yeh, *Member, IEEE,* and Kwang-Cheng Chen, *Fellow, IEEE*

***Abstract*—Molecular communications emerges as a promising scheme for communications between nanoscale devices. In diffusion-based molecular communications, molecules as information symbols diffusing in the fluid environments suffer from molecule crossovers, i.e., the arriving order of molecules is different from their transmission order, leading to intersymbol interference (ISI). In this paper, we introduce a new family of channel codes, called *ISI-free codes*, which improve the communication reliability while keeping the decoding complexity fairly low in the diffusion environment modeled by the Brownian motion. We propose general encoding/decoding schemes for the ISI-free codes, working upon the modulation schemes of transmitting a fixed number of identical molecules at a time. In addition, the bit error rate (BER) approximation function of the ISI-free codes is derived mathematically as an analytical tool to decide key factors in the BER performance. Compared with the uncoded systems, the proposed ISI-free codes offer good performance with reasonably low complexity for diffusion-based molecular communication systems.**

***Index Terms*—Molecular communications, diffusion, intersymbol interference (ISI), channel coding.**

I. Introduction

With the advance in nanotechnology, nanomachines have received much attention for applications in such as biological systems. Since a single nanomachine has limited computational capabilities, a group of nanomachines are usually assigned a specific task, and that requires communications between those nanomachines. Communications between nanomachines is a challenging task—the traditional way of communication through electromagnetic waves is mostly impractical and suffers from significant attenuation in fluid environments [1]. Molecular communications, using molecules as an information carrier, has been considered a promising solution [2]. Recently, molecular communications has been studied by researchers from different disciplines, including nanotechnology, biotechnology, and communication theory [3]. Research efforts on evaluating and bounding achievable information rates, i.e., channel capacity, of molecular communications under different system designs have been conducted [3]–[6].

Due to the effect of channel and noise, channel coding has been widely considered in wireless communications in order to ensure the communication quality. To approach the theoretical limits of bit error rate (BER), it is widely agreed that complicated channel coding schemes and codes with large block size need to be applied. However, complex coding schemes are impractical to be implemented in molecular communications, where the size of components and the encoding/decoding complexity are strictly constrained. Additionally, whether popular channel codes for modern wireless systems still perform well under the environment of molecular communications remains an open question.

The purpose of this paper is to introduce a new family of channel codes for diffusion-based molecular communications. One-dimensional Brownian motion with drift is considered to model the physical propagation process of molecules [7], [8]. The transmitter of the diffusion-based molecular communication encodes information by defining different types of molecules as different information, and every time it releases one or more molecules as an information symbol. At the receiving end, those molecules are gathered and the information sequence is decoded according to the types of arrival molecules. Due to diffusion, the traveling time of molecules is a random variable and the molecules may not arrive in time order. This may result in *intersymbol interference* (ISI), leading to wrong information decoding. In this paper, we propose a new family of channel codes, called *ISI-free codes*, to mitigate ISI in the diffusion-based molecular communication system. Theoretical approximations of the BER performance of the proposed ISI-free codes are derived.

A. Related Work and Contribution

In the history of channel code development, the error rate performance has always been the main concern, and good error correcting codes usually require complex decoding. Although some reduced or low complexity decoding algorithms, such as the syndrome decoding of some linear block codes (e.g., BCH codes) [9], the Viterbi decoding of convolutional codes [9], the iterative decoding of Turbo codes [10], [11] and graph-based codes (e.g., low-density parity-check (LDPC) codes) [12], [13], and the Berlekamp-Massey algorithm [9] and the list decoding [14] of Reed-Solomon codes, have been developed, we believe that these decoding algorithms are still too complex for the applications of nano-communications, in which the nanomachines have very limited computational capability. On the other hand, in spite of the simple decoding rules of Hamming codes and repetition codes [9], these codes do not necessary provide good error correcting performance under

Manuscript August 2, 2012; revised October 30, 2012.

P.-J. Shih, P.-C. Yeh, and K.-C. Chen are with the Dept. of EE, National Taiwan University, Taipei 10617, Taiwan (e-mail: bonoshi@gmail.com; {pcyeh, chenkc}@cc.ee.ntu.edu).

C.-H. Lee is with the Research Center for Information Technology Innovation (CITI), Academia Sinica, Taipei, Taiwan.

Digital Object Identifier 10.1109/JSAC.2013.SUP2.12130018

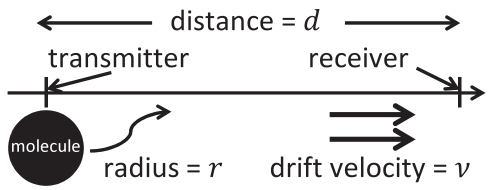


Fig. 1. System model.

the diffusion channel.

The philosophy of the research in this paper is that, when designing channel codes for the different channel, we bear in mind the principles of low decoding complexity. The main contribution of this paper is thus the proposal of families of channel codes that have good performance and low decoding complexity for diffusion-based molecular communications. In [15], we have introduced the ISI-free $(4, 2, 1)$ code. In this paper, two families of ISI-free codes are constructed. Properties of crossover probability are discussed in order to give a design guideline for good channel codes for diffusion channel. In addition, coded-modulation and the improved codeword assignments are discussed.

B. Organization

The remaining sections are organized as follows. The system model is described in Section II. In Section III, ISI-free codes are reviewed and the theoretical BER approximations of the ISI-free codes are derived. Then in Section IV, families of ISI-free codes are introduced. Section V shows how to improve the performance of ISI-free codes by optimizing codeword assignments. In Section VI, decoding complexity is discussed. In Section VII, numerical examples are shown. Finally, conclusions are brought out in Section VIII.

II. System Model

In this section, the Brownian motion channel model describing the diffusion process of molecules is reviewed, and the relations between the process and the inverse Gaussian distribution are examined. The molecular communication scheme adopted throughout this article is also introduced.

A. Brownian Motion Diffusion Channel

The system model is depicted in Fig. 1 [15]. We consider the scheme that the molecules diffuse in an infinite one-dimension dilute fluid environment. The model of Brownian motion with drift is adopted to describe this physical propagation process of molecules [7], [8]. We assume that there is only a pair of transmitter and receiver with zero volume, and they are separated by a distance d with locations fixed. The molecules released from the transmitter diffuse with a drift velocity $\nu > 0$ toward the receiver. We also assume that all the molecules are of the same radius r, never deteriorate, are perfectly absorbed and removed from the fluid once they reach the receiver. No background noise due to other molecules is assumed to happen.

From fluid mechanics, the diffusion coefficient D in dilute solutions is [16]

$$D = \frac{k_B T_a}{6\pi\eta r}, \tag{1}$$

where k_B is the Boltzmann constant, T_a the absolute temperature, and η the viscosity constant, whose value depends on the liquid type and its temperature. The traveling time of molecules from the transmitter to the receiver is the *first hitting time*, whose probability density function (PDF) is [17]

$$f_d(t) = \begin{cases} \frac{d}{\sqrt{4\pi D t^3}} \cdot e^{-\frac{(d-\nu t)^2}{4Dt}} & , t > 0, \\ 0 & , t \leq 0. \end{cases} \tag{2}$$

The random variable t in (2) follows the *inverse Gaussian distribution* $IG(\mu, \lambda)$, where μ is the mean and λ is the scaling parameter. Its PDF is defined as

$$f_{IG}(t; u, \lambda) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right), t > 0, \tag{3}$$

and its cumulative distribution function (CDF) is known as [18]

$$\begin{aligned} F_{IG}(t; u, \lambda) = {} & Q\left(\sqrt{\frac{\lambda}{t}}\left(1 - \frac{t}{\mu}\right)\right) \\ & + \exp\left(\frac{2\lambda}{\mu}\right) Q\left(\sqrt{\frac{\lambda}{t}}\left(1 + \frac{t}{\mu}\right)\right), t > 0, \end{aligned} \tag{4}$$

where $Q(\cdot)$ is the Q-function defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u. \tag{5}$$

Comparing (2) and (3), we obtain $\mu = d/\nu$ and $\lambda = d^2/(2D)$. From (4), the CDF of the first hitting time distribution at the receiver is therefore

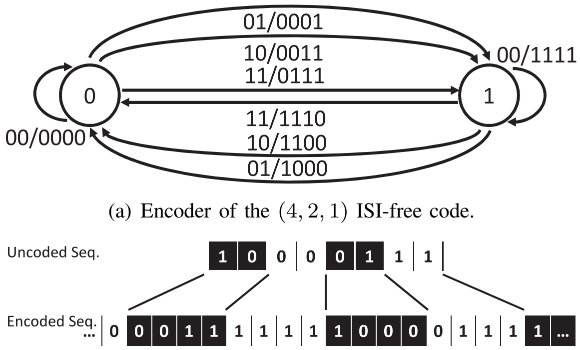
$$F_d(t) = \begin{cases} Q\left(\frac{d-\nu t}{\sqrt{2Dt}}\right) + e^{\frac{d\nu}{D}} Q\left(\frac{d+\nu t}{\sqrt{2Dt}}\right) & , t > 0, \\ 0 & , t \leq 0. \end{cases} \tag{6}$$

B. Molecular Communication Scheme

Different diffusion-based molecular communication schemes have been proposed by carrying information on the number, the type, the inter-transmission duration of molecules, and the combinations of the above schemes [4], [7], [19], [20]. In this paper, we use two distinguishable kinds of molecules to represent information bit 0 and bit 1 respectively. At the beginning of every fixed transmission interval, an odd number m of molecules of the same type, depending on its input being bit 0 or bit 1, are transmitted. We call the system modulation-m (M-m) when referring to transmitting m molecules at a time. In an uncoded system, the receiver gathers m molecules and takes the majority vote to determine which bit has been sent, and outputs an information sequence of 0's and 1's. We consider the scheme of $m = 1$, i.e., M-1, everywhere in this paper except in Section IV-C.

III. ISI-free (n, k, ℓ) Codes

In the molecular communication system that we consider, different types of molecules represent different information symbols. When the arriving order of molecules is different from their transmission order, we call this phenomenon *crossover*. For the M-1 system, only one molecule is sent at a



(a) Encoder of the $(4, 2, 1)$ ISI-free code.

(b) An example of the $(4, 2, 1)$ ISI-free code sequence. The underlined parts indicate the same bits.

Fig. 2. Encoding of the ISI-free $(4, 2, 1)$ code [15].

fixed transmission interval, and a level-ℓ crossover is defined to be the crossover of two molecules sent ℓ transmission intervals apart. The intersymbol interference (ISI) resulting from the crossovers may lead to bit errors. In [15], we have proposed a new class of channel codes, namely, *ISI-free codes*, to remove ISI caused by any number of crossovers up to level-ℓ. An ISI-free (n, k, ℓ) code maps a k-bit information to an n-bit codeword, where ℓ is called the *ISI-free level*. The ISI-free $(4, 2, 1)$ code was explicitely shown in [15] to give an example.

In this section, first the ISI-free $(4, 2, 1)$ code and the theoretical approximations of BER performance are reviewed. Then, more ISI-free (n, k, ℓ) codes are introduced and more mathematical treatments about the BER approximations are provided.

A. Code Construction

The crossovers can be classified into two categories: the crossovers from the contiguous codewords and the crossovers within a codeword. To eliminate the crossovers up to level-ℓ between the contiguous codewords, codewords in the ISI-free (n, k, l) codes have ℓ identical bits in the beginning and ℓ identical bits in the end. Each k-bit information is assigned two different codewords, one starting with ℓ 0's and another starting with ℓ 1's. According to the last bit in the previous codeword, one of the two codewords representing the outgoing information is transmitted to make every contiguous two codewords connected by either 2ℓ 0's or 2ℓ 1's. To eliminate the crossovers within a codeword, the permutation sets from every codeword under all possible crossovers up to level-ℓ should be disjoint. To give an example, an ISI-free $(4, 2, 1)$ code has been introduced in [15] with the codeword assignments shown in Table I. The encoder of the ISI-free $(4, 2, 1)$ code is best illustrated by a state diagram as shown in Fig. 2(a). To choose which codeword to send, the next state is determined by the last bit of the previous codeword. An illustrating sequence is shown in Fig. 2(b): starting with state 0, the information sequence {10, 00, 01, 11} is encoded as {0011, 1111, 1000, 0111}. Since the last bit of the codeword "0011" is "1", the encoder chooses "1111" from the two possible codewords ("0000" and "1111") of the information bits "00". One can easily check why "1000", instead of "0001", is chosen as

TABLE I
CODEWORD ASSIGNMENTS OF THE ISI-FREE $(4, 2, 1)$ CODE AND THEIR LEVEL-1 PERMUTATIONS.

Information bits	Codeword starting with 0	Level-1 permutation	Codeword starting with 1	Level-1 permutation
00	0000	0000	1111	1111
01	0001	0001, 0010	1000	0100, 1000
10	0011	0101, 0011	1100	1100, 1010
11	0111	0111, 1011	1110	1101, 1110

the codeword of the next information sequence "01". The decoding rule of the ISI-free $(4, 2, 1)$ code is as follows: calculate the remainder of the number of 1's in the received codeword divided by 4, i.e., mod(number of 1's, 4), and then convert it to binary expression. This decoding rule is very simple and suitable for molecular communications due to the low complexity requirements.

It is possible to construct ISI-free codes with higher code rates by modifying the construction rules to allow the permutation sets from the two codewords of the same information bits to be *not* disjoint. We define two codewords as a *complementary pair* (CP) if the intersection of their permutation sets is not empty. The qualified ISI-free codes are the ISI-free $(4, 2, 1)$ code, the ISI-free $(5, 2, 2)$ code, the ISI-free $(7, 4, 1)$ code, and so on. The codeword assignments of the ISI-free $(5, 2, 2)$ code when CPs are applied, for example, are shown in Table II. The decoding rule for the ISI-free $(5, 2, 2)$ code involves only the counting of the number of 1's in the received codeword. Denote the number of 1's in the received codeword as a. If $a \leq 3$, the information bits are a (in decimal); otherwise, they are $5 - a$ (in decimal).

The codeword assignments of the ISI-free $(7, 4, 1)$ code are provided in Table III, and the decoding rule of the ISI-free $(7, 4, 1)$ code is shown in Algorithm 1.

The proposed decoder for the ISI-free $(7, 4, 1)$ code is constructed manually in order to make the decoder as simple as possible to implement (but with decent performance). The decoding procedure is explained as follows. First of all, the seven bits of the received sequence are separated in time order into five parts with sizes 2, 1, 1, 1, and 2, respectively. In each part, the decoder counts the number of 1's and records them as a_1, a_2, a_3, a_4, and a_5. For example, if the receiver receives the sequence "0100111", a_1, a_2, a_3, a_4, and a_5 will be 1, 0, 0, 1, and 2, respectively. Also, denote i as the decoded result (information bit in decimal). In line 1 of the decoding algorithm, "a" is the number of bit 1 in the sequence. In lines 3-4, if a is larger than 3, a_c, a_1, a_2, a_3, a_4, and a_5 are modified into their maximum possible values minus their current values, which are $7-a_c$, $2-a_1$, $1-a_2$, $1-a_3$, $1-a_4$, and $2 - a_5$, respectively; otherwise, they remain unchanged. Since every non-CP is composed of its codeword starting with 0 and its bit-complement, the subtractions at line 4 serve to decode a codeword and its bit-complement the same. Note that since line 4 also affects CP's, the value "a" may be needed. In the following, we show how the permutation set of every codeword is decoded. For the codewords "0000000" and "1111111", their permutation sets contain only themselves. Since $a_c = 0$ for this case, $i = 0$ (line 5). For the codewords with information 1, 2, and 3 in decimal, they contain only

Algorithm 1 DECODING RULE OF ISI-FREE $(7, 4, 1)$ CODE

Input: Separate seven bits of the received codeword in time order into five parts with sizes 2, 1, 1, 1, and 2, respectively. In each part, count the number of 1's and record them as $a_1, a_2, \cdots, a_5$.

Output: i as information bits in decimal.

1: $a \leftarrow a_1 + a_2 + a_3 + a_4 + a_5$
2: $a_c \leftarrow a$ // subscript c means "complementary" here
3: **if** $a > 3$ **then**
4: $a_c \leftarrow 7 - a_c$; $a_1 \leftarrow 2 - a_1$; $a_2 \leftarrow 1 - a_2$; $a_3 \leftarrow 1 - a_3$; $a_4 \leftarrow 1 - a_4$; $a_5 \leftarrow 2 - a_5$
5: **end if**
6: **if** $a_c = 0$ **then**
7: $i \leftarrow 0$
8: **else if** $a_c = 1$ **then**
9: $i \leftarrow 1 + 2a_1 + a_5$ // $i =$ 1, 2, or 3
10: **else if** $a_c = 2$ **then**
11: $i \leftarrow 6 - (a_4 + a_5)$ // $i =$ 4, 5, or 6
12: **if** $a_4 + a_5 \neq 2$ **and** $a > 3$ **then**
13: $i \leftarrow i + 2$ // shift $i = 5$ to $i = 7$ and $i = 6$ to $i = 8$
14: **end if**
15: **else** // i.e., $a_c = 3$
16: **if** $a_1 + a_2 = 0$ **then**
17: $i \leftarrow 9$
18: **else**
19: **if** $a_4 = 1$ **and** $a_5 = 1$ **then**
20: $i \leftarrow 12$
21: **else**
22: $i \leftarrow 10 + a_5$ // $i =$ 10, 11, or 12
23: **end if**
24: **if** $a > 3$ **then** // i.e., $a = 4$
25: $i \leftarrow i + 3$ // shift $i = 10$ to 13 and $i = 12$ to 15
26: **end if**
27: **end if**
28: **end if**
29: **return** i

one or six 1's, and thus $a_c = 1$. By observing that the level-1 permutation set of the codeword "0001000" contains "0010000", "0001000", and "0000100", we obtain $(a_1, a_5) = (0, 0)$; for the codeword "1110111", $(a_1, a_5) = (0, 0)$ as well. Similarly, for the codewords with information 2 and 3, we obtain $(a_1, a_5) = (0, 1)$ and $(a_1, a_5) = (1, 0)$. Therefore, we conclude that $i = 1{+}2a_1{+}a_5$ (line 8). For the codewords with information bits 4, 5, 6, 7, and 8 in decimal, all the elements in their permutation sets have $a_c = 2$. It can be shown that the corresponding values of $a_4 + a_5$ are now 2, 1, 0, 1, and 0, respectively. Thus, $i = 6 - (a_4 + a_5)$ (line 10) for the cases with information bits 4, 5, and 6 in decimal. For the codewords with information bits 7 and 8 in decimals, it is easy to observe that their permutation sets both lead to $a = 5$. By the conditions in line 11 and line 12, i is increased by 2, turning $i = 5$ and $i = 6$ into $i = 7$ and $i = 8$, and the decoding is completed. For the codewords with information 9, 10, 11, 12, 13, 14, and 15, their permutation sets all lead to $a_c = 3$. The same trick can be applied for the rest part of the algorithm.

TABLE II
CODEWORD ASSIGNMENTS OF THE ISI-FREE $(5, 2, 2)$ CODE.

Information bits	Codeword starting with 0	Level-2 permutation	Codeword starting with 1	Level-2 permutation
00	00000	00000	11111	11111
01	00100	$\cdots$	11011	$\cdots$
10 (CP)	00011	01010,$\cdots$	11000	01010, $\cdots$
11 (CP)	00111	10101, $\cdots$	11100	10101, $\cdots$

TABLE III
CODEWORD ASSIGNMENTS OF THE ISI-FREE $(7, 4, 1)$ CODE.

Information bits	Codeword starting with 0	Codeword starting with 1
0000	0000000	1111111
0001	0001000	1110111
0010	0000001	1111110
0011	0111111	1000000
0100	0000011	1111100
0101 (CP)	0100001	1000001
0110 (CP)	0110000	1010000
0111 (CP)	0111110	1011110
1000 (CP)	0101111	1001111
1001	0000111	1111000
1010 (CP)	0111000	1011000
1011 (CP)	0110001	1010001
1100 (CP)	0100011	1000011
1101 (CP)	0100111	1000111
1110 (CP)	0101110	1001110
1111 (CP)	0111100	1011100

B. Theoretical Approximations of BER

Now let us investigate the BER performance of the ISI-free codes. Denote T_1 and T_2 as two independent random variables of the first hitting time. Assume two molecules are sent at a time interval t apart, the probability that a crossover happens between these two molecules is:

$$\begin{aligned} P_c(t) = \mathbb{P}[T_2 + t < T_1] &= \int_0^\infty \mathbb{P}[T_1 > u + t] f_d(u) \mathrm{d}u \\ &= \int_0^\infty \overline{F_d}(u + t) f_d(u) \mathrm{d}u, \end{aligned} \tag{7}$$

where $\overline{F_d}(t) = 1 - F_d(t)$, which is the complementary CDF (CCDF) of the first hitting time. The bit error rate (BER) of the ISI-free codes is approximated as BER $\approx$ $\mathrm{BER}_{\mathrm{pre}} + \mathrm{BER}_{\mathrm{next}} + \mathrm{BER}_{\mathrm{in}}$,where $\mathrm{BER}_{\mathrm{pre}}$, $\mathrm{BER}_{\mathrm{next}}$, and $\mathrm{BER}_{\mathrm{in}}$ are the BER accounting for only crossovers from the previous codeword, from the next codeword, and within the same codeword, respectively [15].

In [15], the throughput criterion has been established to compare BER of channel codes with different code rate. Different channel codes are compared under the same throughput, rather than the same inter-transmission time. This is because if the inter-transmission time of molecules is the same, channel codes with lower code rate naturally enjoy better BER performance at the cost of less throughput. Let T_b denote the fixed period to send an information bit. Since the most likely bit error happens due to a single level-$(\ell + 1)$ crossover, the

BER of an ISI-free (n,k,ℓ) code (for the M-1 scheme) is [15]:

$$\text{BER} \approx (e_{\text{pre}}+e_{\text{next}}+e_{\text{in}})P_c((\ell+1)RT_b) = e_{\text{sum}}P_c((\ell+1)RT_b), \tag{8}$$

where $R=k/n$ is the code rate, e_{pre}, e_{next} and e_{in} are the coefficients of the terms $P_c((\ell+1)RT_b)$ in BER_{pre}, BER_{next} and BER_{in}, respectively, and $e_{\text{sum}} = e_{\text{pre}}+e_{\text{next}}+e_{\text{in}}$ is defined as the *ISI-free error coefficient.* As will be studied later, $P_c(\cdot)$ is a function with an exponentially-decaying term in its product. Therefore, for an ISI-free (n,k,ℓ) code, $(\ell+1)R$, defined as the *ISI-free index* ξ, dominates the relative performance between different ISI-free codes under the same information bit rate (i.e., throughput $1/T_b$) when T_b is large. The ISI-free indices of the ISI-free $(4,2,1)$ code, the ISI-free $(5,2,2)$ code, and the ISI-free $(7,4,1)$ code are 1, $\frac{8}{7}$, and $\frac{6}{5}$, respectively.

To determine the ISI-free error coefficient of an ISI-free code, one should derive the probability that a given codeword is wrongly detected. Take the ISI-free $(4,2,1)$ code as an example. Denote $C_{i,j}$ as the codeword which starts with bit i and represents the information bits in the j-th row in Table I, where $i\in\{0,1\}$ and $j\in\{0,1,2,3\}$. If we use C_j to represent both the codewords $C_{0,j}$ and $C_{1,j}$, then $\mathbb{P}[C_j\to C_{j'}]$, $j\neq j'$, denotes the probability that C_j is detected as $C_{j'}$. Note that this probability only accounts for the bit errors due to a single level-$(\ell+1)$ crossover. From the BER analysis in [15], it can be shown that $\mathbb{P}[C_0\to C_1]=\frac{1}{8}P_c(T_b)$, $\mathbb{P}[C_0\to C_3]=\frac{1}{8}P_c(T_b)$, $\mathbb{P}[C_1\to C_0]=\frac{1}{8}P_c(T_b)$, $\mathbb{P}[C_1\to C_2]=\frac{5}{8}P_c(T_b)$, $\mathbb{P}[C_2\to C_1]=\frac{1}{8}P_c(T_b)$, $\mathbb{P}[C_2\to C_3]=\frac{1}{8}P_c(T_b)$, $\mathbb{P}[C_3\to C_0]=\frac{1}{8}P_c(T_b)$, and $\mathbb{P}[C_3\to C_2]=\frac{5}{8}P_c(T_b)$. When C_0 is detected as C_1, the information bits "00" are decoded as "01", and half the bits are wrongly decoded. This means that the ratio of information bits being wrongly detected, denoted as $\gamma(C_0\to C_1)$, is $\frac{1}{2}$. Similarly, $\gamma(C_0\to C_3)=1$, $\gamma(C_1\to C_0)=\frac{1}{2}$, $\gamma(C_1\to C_2)=1$, $\gamma(C_2\to C_1)=1$, $\gamma(C_2\to C_3)=\frac{1}{2}$, $\gamma(C_3\to C_0)=1$, and $\gamma(C_3\to C_2)=1$. Since the probability of having C_j is $1/2^k$, the BER of the ISI-free $(4,2,1)$ code is approximately

$$\frac{1}{2^k}\sum_{j\neq j'}\gamma(C_j\to C_{j'})\mathbb{P}[C_j\to C_{j'}] = \frac{3}{8}P_c(T_b). \tag{9}$$

The ISI-free error coefficient of the ISI-free $(4,2,1)$ code is therefore $\frac{3}{8}$. Following similar procedures, the ISI-free error coefficients of the ISI-free $(5,2,2)$ code and the ISI-free $(7,4,1)$ code can be calculated as $\frac{9}{16}$ and $\frac{1167}{2048}$, respectively.

C. Crossover Probability

The further study of $P_c(t)$ is needed since the crossover probability is the key factor in the BER approximations of the ISI-free codes. In this subsection, we first derive an asymptotic equivalence of $\overline{F_d}(t)$, which will then be used to approximate $P_c(t)$ and show the important role of the ISI-free index.

1) Asymptotic Equivalence of $\overline{F_d}(t)$: Two functions $f(x)$ and $g(x)$ are said to be asymptotically equivalent if the ratio $f(x)/g(x)$ tends to unity when $x\to\infty$. From (2), the CCDF $\overline{F_d}(t)$ of the first hitting time at the receiver is

$$\overline{F_d}(t) = 1-F_d(t) = Q\left(\frac{\nu t-d}{\sqrt{2Dt}}\right) - \exp\left(\frac{d\nu}{D}\right)Q\left(\frac{\nu t+d}{\sqrt{2Dt}}\right). \tag{10}$$

The Q-function $Q(x)$ has the tight bounds[1] when $x>0$:

$$\frac{1}{\sqrt{2\pi}x}\left(1-\frac{1}{x^2}\right)\exp\left(-\frac{x^2}{2}\right) < Q(x) < \frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}\right). \tag{11}$$

However, neither of the two bounds can be used to get an asymptotic equivalence of $\overline{F_d}(t)$ because the two terms in (11) are asymptotic equivalence of each other[2]. Therefore, a different approach must be taken to obtain the asymptotic equivalence of $\overline{F_d}(t)$. By definition, we have

$$\begin{aligned}\overline{F_d}(t) &= \int_t^\infty f_d(x)\mathrm{d}x < \int_t^\infty \frac{x}{t}f_d(x)\mathrm{d}x\\ &= \frac{1}{t}\int_t^\infty \frac{d}{\sqrt{4\pi Dx}}\exp\left[-\frac{(d-\nu x)^2}{4Dx}\right]\mathrm{d}x\\ &= \frac{d}{\nu t}\int_{\sqrt{t}}^\infty \frac{\nu}{\sqrt{\pi D}}\exp\left[-\frac{(d-\nu x^2)^2}{4Dx^2}\right]\mathrm{d}x.\end{aligned} \tag{12}$$

The integral part of the last expression in (12) is equal to [21]

$$\int_0^{\frac{1}{t}} f_{IG}\left(x;\frac{\nu}{d},\frac{\nu^2}{2D}\right)\mathrm{d}x = F_{IG}\left(\frac{1}{t};\frac{\nu}{d},\frac{\nu^2}{2D}\right). \tag{13}$$

Thus, we have

$$\overline{F_d}(t) < \frac{d}{\nu t}\left[Q\left(\frac{\nu t-d}{\sqrt{2Dt}}\right)+\exp\left(\frac{d\nu}{D}\right)Q\left(\frac{\nu t+d}{\sqrt{2Dt}}\right)\right]. \tag{14}$$

By substituting $Q(x)$ with its upper bound in (11) when $t>d/\nu$, it becomes

$$\begin{aligned}\overline{F_d}(t) &< \frac{2d\sqrt{Dt/\pi}}{\nu^2t^2-d^2}\exp(-\frac{(\nu t-d)^2}{4Dt})\\ &= \frac{4Dt^2}{\nu^2t^2-d^2}f_d(t) \approx \frac{4D}{\nu^2}f_d(t),\end{aligned} \tag{15}$$

which is an asymptotic equivalence of $\overline{F_d}(t)$.

2) Approximation of $P_c(\cdot)$: We are interested in the behavior of $P_c(t)$ when t is very large. Now we approximate the crossover probability in (7) as

$$\int_0^{t'} f_d(u)\overline{F_d}(u+t)\mathrm{d}u, \tag{16}$$

where t' is chosen to be large enough to approximate the integration well but still much smaller than t. For a large t with $u\in(0,t')$, by the following approximations,

$$\frac{d}{\sqrt{(u+t)^3}} \approx \frac{d}{\sqrt{t^3}}, \tag{17}$$

$$\exp\left[-\frac{d^2}{4D(u+t)}\right] \approx \exp\left(-\frac{d^2}{4Dt}\right) \approx 1, \tag{18}$$

[1]The bounds are both asymptotic equivalence of $Q(x)$ because the ratio of the two bounds tend to unity when $x\to\infty$.

[2]Consider, for example, the subtraction of $x/(x^2-1)$ and $x/(x^2+1)$, and $1/(x-1)$ and $1/(x+1)$ are their tight bounds, respectively. It can be shown that $x/(x^2-1)-x/(x^2+1)=2x/(x^4-1)$ is apparently not an asymptotic equivalence of $2x/(x^2-1)$, the subtraction of $1/(x-1)$ and $1/(x+1)$, when $x\to\infty$.

we obtain

$$\begin{aligned} P_c(t) &\approx \int_0^{t'} f_d(u)\frac{4D}{\nu^2} f_d(u+t)\mathrm{d}u \\ &\approx \frac{4D}{\nu^2} f_d(t) \int_0^{t'} f_d(u) \exp\left(-\frac{u\nu^2}{4D}\right)\mathrm{d}u \\ &\approx \frac{4D}{\nu^2} \exp\left[-\frac{(\sqrt{2}-1)d\nu}{2D}\right] f_d(t) \\ &\times \int_0^{t'} f_{IG}\left(u; \frac{d}{\sqrt{2}\nu}, \frac{d^2}{2D}\right)\mathrm{d}u \\ &\approx \frac{4D}{\nu^2} \exp\left[-\frac{(\sqrt{2}-1)d\nu}{2D}\right] f_d(t). \end{aligned} \tag{19}$$

Thus, we prove that ISI-free codes with larger ISI-free index, or larger $(l+1)R$ in (8), have better BER performance when T_b is large. For example, by the order of their ISI-free indices, we see that the ISI-free $(5,2,2)$ code is better than the ISI-free $(7,4,1)$ code and the ISI-free $(7,4,1)$ code is better than the ISI-free $(4,2,1)$ code in terms of BER performance, regardless of the relations between their ISI-free error coefficient, e_{sum}. Similar relations between the probabilities of multiple crossovers and $f_d(t)$ and the corresponding ISI-free indices will be brought out in Sec. IV-B.

IV. Families of ISI-free Codes

The ISI-free codes presented in the previous section can be generalized to three families whose ISI-free indices can be arbitrarily high. By reducing the number of the identical ℓ bits in the beginning of codewords, a new class of ISI-free codes is discussed.

A. *ISI-free (n,k,ℓ) Code Family*

1) Structure of ISI-free (n,k,ℓ) codes: It has been shown in Section III-B that the ISI-free index $(\ell+1)R$ is the decisive factor of the performance of ISI-free (n,k,ℓ) codes. When the ISI-free level ℓ is fixed, a higher code rate R results in larger ISI-free indices. The ISI-free $(7,4,1)$ code could be regarded as an attempt to find the ISI-free (n,k,ℓ) codes with higher code rate at $\ell=1$. However, compared to the ISI-free $(4,2,1)$ code, the ISI-free $(7,4,1)$ code has the following disadvantages: (a) the codewords have less regularity, making it difficult to find generalized ISI-free codes, and (b) the decoding rules may still be too complex for molecular communications since some codewords assigned to different information bits have the same number of 1's (or 0's). Therefore, to have larger ISI-free indices and simple decoding rules, we propose the structures for the ISI-free (n,k,ℓ) codes as follows.

For an ISI-free code with code rate k/n, there must have 2^k codewords starting with ℓ 0's and another 2^k codewords starting with ℓ 1's. Consider only the 2^k codewords with ℓ 0's in the beginning. We differentiate those 2^k codewords into three groups, each having n_0, n_1, and n_b members such that $n_0+n_1+n_b=2^k$. The members of the first group are the codewords called *balanced* codewords, within which the numbers of bit 0 and bit 1 are equal. n_b denotes the number of codewords in this group, and n_0 and n_1 are the numbers of *unbalanced* codewords that end with ℓ 0's and end with ℓ 1's, respectively. The notation $[n_0,n_1,n_b]$ is used here to denote the structure of an ISI-free (n,k,ℓ) code. Note that we let the 2^k codewords have different numbers of 1's to make the decoding rules simple. We can see that the ISI-free $(4,2,1)$ code and the ISI-free $(5,2,2)$ code have the structures $[1,2,1]$ and $[2,2,0]$, respectively. As will be shown later, possible structures are $[m,m,0]$, $[m,m+1,1]$, and $[m,m+2,2]$, which are designated as family-I, family-II, and family-III of the ISI-free (n,k,ℓ) codes.

2) Families of ISI-free (n,k,ℓ) codes: We first give an example of the code construction of family-I with the structure $[m,m,0]$. When $k=2$, the structure is $[2,2,0]$. First, to have two codewords both starting and ending with ℓ 0's, we make their numbers of 1's as small as possible by choosing one of them to be zero and the other to be one. With the fact that the second codeword has one 1 and at least 2ℓ 0's, the codeword length n cannot be less than $2\ell+1$. Secondly, to have two codewords starting with ℓ 0's and ending with ℓ 1's, their numbers of 1's must be at least ℓ and $\ell+1$, respectively. Since the four codewords have different numbers of 1's, ℓ must be greater than one. To make n as small as possible, we choose $\ell=2$, and the only possible four codewords starting with ℓ 0's are "00000", "00100", "00011", and "00111". The other four codewords starting with 1's are assigned the 1's complement of the four codewords starting with 0's: "11111", "11011", "11100", and "11000". Then we arrange the eight codewords into pairs through putting together any two codewords either with the same number of 1's or with the sum of their numbers being 5 ($n=5$). Finally, every pair of codewords is assigned to the information bits whose decimal form is equal to the smaller number of 1's out of the two codewords in the pair. The resulting construction is the ISI-free $(5,2,2)$ code shown in Table II.

For $k=2$ and $\ell=3$, one will get the ISI-free $(7,2,3)$ code with its four codewords starting with ℓ 0's as "0000000", "0001000", "0000111", and "0001111". Note that to reduce the BER of the ISI-free codes, for the codewords ending with ℓ 0's, we should put all their 1's in the middle; for the codewords ending with ℓ 1's, we should put all their 1's in the end. For the ISI-free $(7,2,3)$ code, its ISI-free index is $\frac{8}{7}$, which is less desirable than the ISI-free index $\frac{6}{5}$ from the ISI-free $(5,2,2)$ code. Therefore, we should always choose the code with a smaller ℓ.

In general, the following relation holds for the ISI-free (n,k,l) code family-I:

$$(n,k,\ell) = (3\cdot 2^{k-1}-1, k, 2^{k-1}). \tag{20}$$

Example members of family-I are $(2,1,1)$, $(5,2,2)$, $(11,3,4)$, $(23,4,8)$, $(47,5,16)$, and so on. Note that although $(2,1,1)$ is derived from (20), the ISI-free $(2,1,1)$ code does not follow the structure $[1,1,0]$. The decoding rule for the family-I codes is: If $a \le 2^k-1$, the information bits are a in decimal; otherwise, they are $n-a$.

Following similar procedures, the ISI-free (n,k,ℓ) code family-II with the structure $[m,m+1,1]$ can be constructed with

$$(n,k,\ell) = (3\cdot 2^{k-1}-2, k, 2^{k-1}-1). \tag{21}$$

Example members of family-II are $(4,2,1)$, $(10,3,3)$, $(22,4,7)$, $(46,5,15)$, and so on. Note that when $k=1$, by

TABLE IV
CODEWORD ASSIGNMENTS OF THE ISI-FREE $(8, 3, 2)$ CODE.

Information bits	Codeword starting with 0	Codeword starting with 1
000	00000000	11111111
001	00001000	11110111
010	00000011	11000000
011	00000111	11100000
100	00001111	11110000
101	00011111	11111000
110	00111111	11111100
111	00111100	11000011

(21) one will get $(1, 1, 0)$, which is an uncoded system and not an ISI-free code. The ISI-free $(4, 2, 1)$ code is shown in Table I. The decoding rule of the family-II codes is exactly the same as that of the family-I codes.

For the family-III ISI-free (n, k, ℓ) codes following the structure $[m, m+2, 2]$, there are two balanced codewords starting with bit 0. We make one of the codewords have all its $n/2$ 1's exactly in the middle, and the other codeword have all its 1's in the last $n/2$ bits. Following the similar procedure as in the family-I, except that the pair of balanced codewords with $n/2$ identical bits in the center is assigned to the binary expression of $2^k - 1$ as its information bits, we obtain the relation that holds for the family-III codes:

$$(n, k, \ell) = (3 \cdot 2^{k-1} - 4, k, 3 \cdot 2^{k-3} - 1). \tag{22}$$

Family-III includes the member codes of $(8, 3, 2)$, $(20, 4, 5)$, $(44, 5, 11)$, and so on. The ISI-free $(8, 3, 2)$ code, as an example, is shown in Table IV. Note that due to the two balanced codewords, the ISI-free level ℓ of the family-III codes is limited to $n/4$, which is less desirable compared to the family-I and the family-II codes. The decoding rule for the ISI-free code family-III should cope with different information bits that have $n/4$ 1's in their codewords. Let the number of 1's in the first half of the received codeword as a_1, and the number of 1's in the last half as a_2, where $a = a_1 + a_2$. If $a > 2^k - 2$, the information bits in decimal is $n - a$. On the other hand, if $a \le 2^k - 2$, two cases need to be considered: if $a = n/2$ and $|a_1 - a_2| < n/4$, then the information bits in decimal is $2^k - 1$; otherwise, it is a.

Besides the aforementioned structures, one can check that the structures $[m+1, m, 1]$ and $[m+2, m, 2]$ are not desirable since they will incur lower ISI-free indices than the structures of $[m, m+1, 1]$ and $[m, m+2, 2]$.

In summary, the family-I and the family-II of the ISI-free (n, k, ℓ) codes have the ISI-free indices respectively equal to

$$\xi_{\mathrm{I}} = \frac{k}{3 \times 2^{k-1} - 1} \times (2^{k-1} + 1) \tag{23}$$

and

$$\xi_{\mathrm{II}} = \frac{k}{3 \times 2^{k-1} - 2} \times 2^{k-1}, \tag{24}$$

with the family-I codes having slightly larger ISI-free indices than those of the family-II codes with the same k. When 2^k is large, the ISI-free indices of the two families can be both approximated as $\frac{k}{3}$. For the family-III codes, the ISI-free index, derived as

$$\xi_{\mathrm{III}} = \frac{3k \times 2^{k-3}}{3 \times 2^{k-1} - 4}, \tag{25}$$

is approximated by $\frac{k}{4}$ when 2^k is large. We can see that the ISI-free indices of all the three families can be arbitrarily high. While the family-III codes have smaller ISI-free indices and therefore poorer BER performance, the family-I codes enjoy the best performance among the three families.

B. ISI-free (n, k, ℓ, s) Code Family

In this subsection, we consider multiple crossovers between codewords, and conclude that in the previous level-ℓ ISI-free codes, it is unnecessary to have ℓ identical bits in the beginning of a codeword. By redefining ISI-free levels and ISI-free indices, we obtain the more powerful ISI-free (n, k, ℓ, s) codes.

1) Multiple Molecule Crossovers: When there are three molecules m_1, m_2, and m_3 released subsequently at every interval t, the probability that m_3 arrives first at the receiver is

$$\int_0^\infty \overline{F_d}(u+t)\overline{F_d}(u+2t) f_d(u) \mathrm{d}u. \tag{26}$$

As in (19), it can be shown that (26) can be approximated by

$$\left(\frac{4D}{\nu^2}\right)^2 f_d(t) f_d(2t) \exp\left[-\frac{(\sqrt{3}-1)d\nu}{2D}\right]. \tag{27}$$

In general, when the molecules m_1, m_2,$\cdots$, m_N are transmitted one by one at every interval t, the probability that m_N arrives first is

$$P_c^N(t) = \int_0^\infty f_d(u) \prod_{i=1}^N \overline{F_d}(u+it) \mathrm{d}u \tag{28}$$

$$\approx \left(\frac{4D}{\nu^2}\right)^N \exp\left[-\frac{(\sqrt{N+1}-1)d\nu}{2D}\right] \prod_{i=1}^N f_d(it). \tag{29}$$

Let $\Theta(\cdot)$ be the big Theta notation. Since $f_d(t)$ is in $\Theta\left(t^{-1.5}\exp\left(-\frac{\nu^2 t}{4D}\right)\right)$, we have that $P_c^N(t)$ belongs to $\Theta\left(t^{-1.5N}\exp\left(-\frac{\nu^2 N(N+1)t}{8D}\right)\right)$. The exponential term $\exp\left(-\frac{\nu^2 N(N+1)t}{8D}\right)$ in $\Theta(\cdot)$ is more significant than the term $t^{-1.5N}$ when t is large. Therefore, the expression $P_c^{N_1}(\alpha_1 t) + P_c^{N_2}(\alpha_2 t)$ can be approximated by $P_c^{N_1}(\alpha_1 t)$ if $\alpha_2 N_2(N_2+1)$ is larger than $\alpha_1 N_1(N_1+1)$.

2) Definitions of ISI-free (n, k, ℓ, s) codes: Different from what we discussed in Section III, here a level-ℓ ISI-free code guarantees that ℓ is the maximum number such that a single level-ℓ crossover in a sequence will not lead to decoding error. This class of ISI-free codes have less than ℓ identical bits in the beginning of the codewords, and the notation (n, k, ℓ, s) is used to denote that the codeword starts with (at least) s identical bits and ends with (at least) ℓ identical bits.

To explain why the ISI-free codes could have less than ℓ identical bits in the beginning of their codewords, the ISI-free $(5, 2, 2)$ code is taken here as an example. Let us consider the scenario that a coded sequence $\{\cdots 100|001 \cdots\}$ is changed to $\{\cdots 000|101 \cdots\}$ and leads to decoding error due to a level-3 crossover, which happens with probability $P_c^1(1.2T_b)$. Note that the symbol "|" denotes the boundary of the consecutive codewords. Now consider another case that the coded

TABLE V
CODEWORD ASSIGNMENTS OF THE ISI-FREE $(4, 2, 2, 1)$ CODE.

Information bits	Codeword starting with 0	Codeword starting with 1
00	0000	1111
01	0100	1000
10	0011	1100
11	0111	1011

sequence $\{\cdots 100|001\cdots\}$ is changed to $\{\cdots 101|000\cdots\}$, where the last bit 1 arrives earlier at the receiver than the three bit 0's transmitted before it. This happens with probability $P_c^3(0.4T_b)$ according to (29). The term $P_c^3(0.4T_b)$ can be omitted compared to the term $P_c^1(1.2T_b)$ when T_b is large. As a result, we can remove the first bit of all the codewords in Table II and rearrange the codewords into pairs by putting together any two codewords either with the same number of 1's or the summation of the numbers of 1's being 4 ($n = 4$) to obtain the ISI-free $(4, 2, 2, 1)$ code shown in Table V. The increase in R from 0.4 to 0.5 reduces the dominant term of BER from $P_c^1(1.2T_b)$ to $P_c^1(1.5T_b)$, but boosts the previously-ignored term $P_c^3(0.4T_b)$ to $P_c^2(0.5T_b)$, which is now as significant as $P_c^1(1.5T_b)$ since they have the same exponential term $\exp\left(-\frac{1.5\nu^2 T_b}{4D}\right)$. Nevertheless, it can be shown that the approximated BER of the ISI-free $(4, 2, 2, 1)$ code is

$$\frac{3}{4}(P_c^1(1.5T_b) + P_c^2(0.5T_b)), \tag{30}$$

which is much smaller than $\frac{3}{8}P_c(T_b)$ of the ISI-free $(4, 2, 1)$ code or $\frac{9}{16}P_c(1.2T_b)$ of the ISI-free $(5, 2, 2)$ code when T_b is large. The decoding rule of the ISI-free $(4, 2, 2, 1)$ code is the same as that of the ISI-free $(4, 2, 1)$ code.

3) ISI-free (n, k, ℓ, s) Code Family: We have shown that the family-I ISI-free (n, k, ℓ) codes are the best among the three families in terms of BER performance. Similar to the ISI-free $(4, 2, 2, 1)$ code, we construct a family of the ISI-free (n, k, ℓ, s) codes by removing the first $\ell - s$ identical bits in the family-I of ISI-free (n, k, ℓ) codes. Hence, $(n, k, \ell, s) = (2^k - 1 + s, k, 2^{k-1}, s)$. The optimal number of the identical bits which should be removed from the ISI-free (n, k, ℓ, s) codes can be estimated. By defining the ISI-free index of the ISI-free (n, k, ℓ, s) code family as

$$\xi(s) = \min\left(\frac{R(s+1)(s+2)}{2}, (\ell+1)R\right), \tag{31}$$

the optimal s, namely the remaining number of identical bits, is given by

$$\arg\max_s \xi(s) = \arg\max_s \ \min\left(\frac{(s+1)(s+2)}{2(2^k-1+s)}, \frac{\ell+1}{2^k-1+s}\right). \tag{32}$$

When $(s+1)(s+2)/2$ is larger than $\ell + 1$, the term $\min\left(\frac{(s+1)(s+2)}{2(2^k-1+s)}, \frac{\ell+1}{2^k-1+s}\right)$ becomes smaller with the increase in s. Therefore, (32) becomes

$$\arg\max_{s \le \text{ceil}\left(\frac{-3+\sqrt{9+8\ell}}{2}\right)} \frac{(s+1)(s+2)}{2(2^k-1+s)}. \tag{33}$$

The solution to this maximization problem is $\text{ceil}\left(\frac{-3+\sqrt{9+8\ell}}{2}\right)$ since the term in $\arg\max(\cdot)$ has extreme values only at $s < 0$ when $k \geq 1$. Therefore, the family of the ISI-free (n, k, ℓ, s) codes includes $(4, 2, 2, 1)$, $(9, 3, 4, 2)$, $(18, 4, 8, 3)$, $(36, 5, 16, 5)$, and so on. The decoding rule of the ISI-free (n, k, ℓ, s) family is the same as that of the ISI-free (n, k, ℓ) code family-I and family-II.

C. Coded-modulation with the ISI-free Code Families

The ISI-free codes can also be implemented upon modulation schemes M-m with $m > 1$. For the ISI-free (n, k, ℓ) code families and the ISI-free (n, k, ℓ, s) code family, the transmitter sends nm molecules for every codeword. The receiver gathers nm molecules and counts the number of the molecules standing for bit 1. This number is divided by n and round to the closest decimal. By applying the decoding rule for the ISI-free code family with $m = 1$ (the M-1 scheme), the information bits are obtained. Take the ISI-free $(4, 2, 2, 1)$ code upon M-3 as an example. When the number of 1's lies in the closed intervals $[0, 1]$, $[2, 4]$, $[5, 7]$, $[8, 10]$ and $[11, 12]$, the coded-modulation in effect will decode the information bits as "00", "01", "10", "11", and "00", respectively.

For different M-m systems, one should also take into account the number of molecules spent in unit time to make the BER performance comparison fair. Define E_c as the molecules spent in unit time at the transmitter, then $E_c = \frac{m}{RT_b}$. Under the same throughput $1/T_b$, m is chosen to make sure that systems with different channel code settings have approximately the same E_c. For example, to compare the BER performance between an uncoded system and a system coded by the ISI-free $(4, 2, 2, 1)$ code, one should use M-$(2m+1)$ (or M-$(2m-1)$) and M-m in the respective systems.

V. IMPROVED CODEWORD ASSIGNMENTS

Inspired by the conventional Gray-coded symbol assignments, the mapping from information bits to codewords can be improved to get better BER performance for the ISI-free codes. Take the ISI-free $(4, 2, 1)$ code as an example. The optimal codeword assignments are to let information bits "00", "01", "11", and "10" be the pairs C_0, C_1, C_2, and C_3, respectively. Then, when any pair C_j is detected wrongly as another pair $C_{j'}$ due to a level-2 crossover, only half of the information bits is flipped. Therefore, the BER of the ISI-free $(4, 2, 1)$ code is reduced to $\frac{1}{8}\times(\frac{1}{4}P_c(T_b)\times 6+\frac{5}{4}P_c(T_b)\times 2)\times\frac{1}{2} = \frac{1}{4}P_c(T_b)$, with approximately 33% in reduction. Notice that the reassignments of codewords will change their decoding rules.

Information bits could be manually assigned to codewords with the knowledge of probability $\mathbb{P}[C_j \to C_{j'}]$, which accounts for the most likely scenario of bit error. For example, suppose that $k = 4$ and there are 16 C_j's, $j \in \{0, 1, \cdots, 15\}$. The two-dimensional 4-bit Gray-coding map is shown at the left hand side of Fig. 3, where the four bits in each grid are information bits to be assigned to C_j's. Note that for information bits in neighboring grids, their Hamming distance is only one. Also note that the 4-bit Gray-coding map should be extended to include more neighboring grids as shown at the right hand side of Fig. 3. If all the pairs of C_j's with probability $\mathbb{P}[C_j \to C_{j'}]$ are neighbors in the Gray-coding map, then all the possible wrong decisions of C_j's only lead

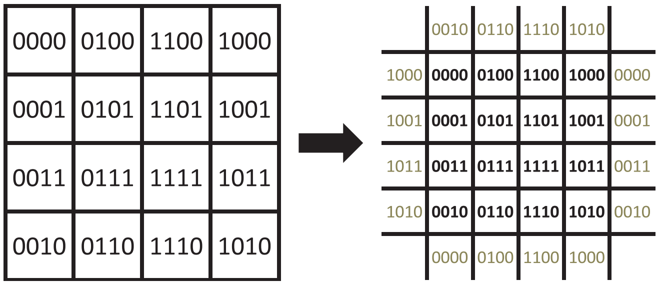


Fig. 3. A 4-bit Gray coding map.

to one bit error. Therefore, the optimal codeword assignments are achieved.

Generally speaking, for an ISI-free code, not all the pairs of C_j's with probability $\mathbb{P}[C_j \to C_{j'}]$ could be assigned as neighbors in Gray-coding maps. Then, the pairs with smaller error probability should be assigned to have larger Hamming distance. It is interesting to develop efficient algorithms to guarantee the optimality of the mapping between information bits and codewords, which is left as future works.

VI. Decoding Complexity

Since this work is motivated by the limited computational capability of nanomachines, we compare the decoding complexity of the proposed ISI-free codes with some popular channel codes in this section.

The decoding of the ISI-free (n, k, l, s) code family takes only an accumulator to count the numbers of 1's, summing up to at most n. Whenever the receiver receives a molecule, the accumulator only has to make a decision: plus one or remain the same value. At the end of a codeword, depending on the number of 1's, the decoder may perform only a subtraction.

In [15], we compare the ISI-free $(4, 2, 1)$ code with the $[133, 171]$ convolutional code and show that the ISI-free $(4, 2, 1)$ code outperforms the $[133, 171]$ convolutional code when T_b is small. Since the proposed ISI-free (n, k, l, s) codes far outperform the ISI-free $(4, 2, 1)$ code, the ISI-free (n, k, l, s) codes also outperform convolutional codes. Therefore, the $[133, 171]$ convolutional code is used for the decoding complexity comparison. The Viterbi decoder for convolutional codes is suggested to use five times of its encoding memory for each path (called survivors). For the decoding of the $[133, 171]$ convolutional code, it has 26 trellis paths, each with at least 30 information bits, and thus overall 960 bits should be stored. Also, each time it should perform 27 additions and 27 comparisons to maintain the metrics of 26 survivors. In conclusion, the decoding complexities of the ISI-free codes are far less than that of the convolutional codes.

The decoding complexity of some other popular channel codes is generally larger than the proposed ISI-free codes. The computational complexity of BCH decoders is $O(n^2 \log(n)^2)$ [22] and the complexity of the syndrome-based decoding of Reed-Solomon codes is $O(n(\log_2 n)^2)$ [23], where n is the code length. Usually, large n is required in order to achieve good performance, which makes the complexity high.

The decoding complexity using belief propagation for LDPC codes can be calculated as $O(2Jp + 4rq)$, where J is the number of rows, r is the number of columns in the parity-check matrix with each row consisting of p 1's and each column consisting of q 1's ([9], page 877). The decoding complexity of LDPC codes is thus prohibitively high for nanomachines. This shows the superiority of the proposed ISI-free codes in terms of decoding complexity.

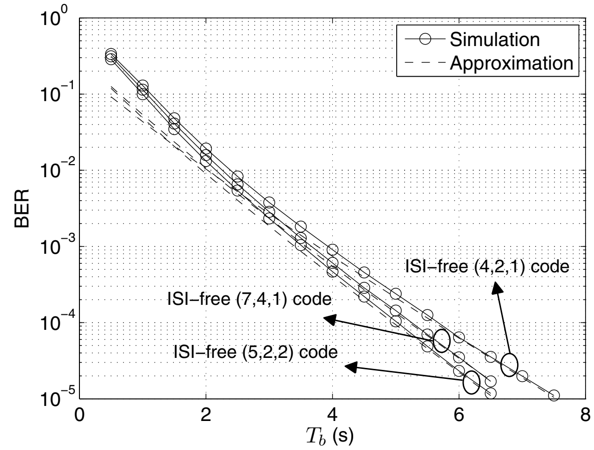


Fig. 4. Comparison of simulations and theoretical approximations for the ISI-free $(4, 2, 1)$, $(7, 4, 2)$, and $(5, 2, 2)$ codes.

VII. Numerical Examples

In this section, numerical examples of the diffusion-based molecular communications with and without coding are presented. Information bits are transmitted randomly and the simulation parameters are as follows: the temperature is set to 298 K, the viscosity of water is 0.894 mPa·s, the molecule radius is 10 nm, the distance between the transmitter and the receiver is 10 μm, the diffusion coefficient is 2.44038×10^{-11} m^2/s, and the drift velocity is 10 μm/s.

First, let us investigate the accuracy of the theoretical BER approximations of the ISI-free codes. In Fig. 4, the BER curves of simulations and theoretical approximations using (8) for the ISI-free $(4, 2, 1)$ code, the ISI-free $(7, 4, 1)$ code, and the ISI-free $(5, 2, 2)$ code upon M-1 are shown. Their theoretical results agree with the simulations very well, and the BER performance is better with the increase in the ISI-free index. Also, we see that the ISI-free codes with larger ISI-free index have better BER performance.

The simulation results of the ISI-free code families upon M-1 with Gray-coding are shown in Fig. 5. We compare the BER performances of the Gray-coded family-I and family-II ISI-free (n, k, ℓ) codes with different k. As expected, larger k results in lower BER due to the increase of the ISI-free index. Also, as we have predicted, the family-I codes have lower BER than the family-II codes, but the gap is smaller with large k. In Fig. 6, the BER performances of the ISI-free (n, k, ℓ, s) code family and the ISI-free (n, k, ℓ) code family-I are compared. One can see that the BER performance of the ISI-free (n, k, ℓ, s) codes is enhanced significantly by removing the first $\ell - s$ bits from the codewords of the ISI-free (n, k, ℓ) code family-I. For example, at $T_b = 2$ s and $T_b = 6.5$ s, the BER of the ISI-free $(4, 2, 2, 1)$ code is only about 56% and 12% of the BER of the ISI-free $(5, 2, 2)$ code.

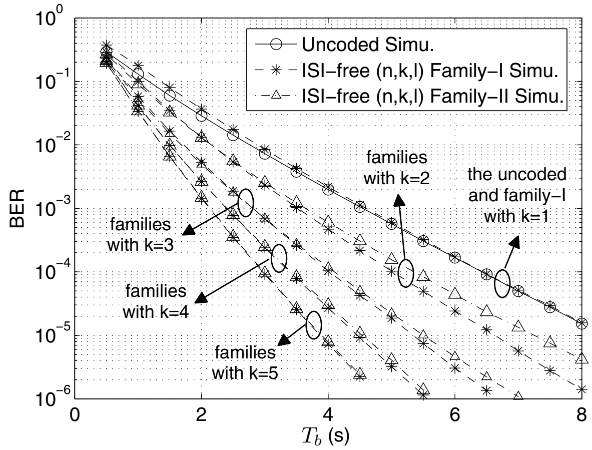


Fig. 5. Performance comparison of the Gray-coded family-I and family-II ISI-free (n, k, ℓ) codes.

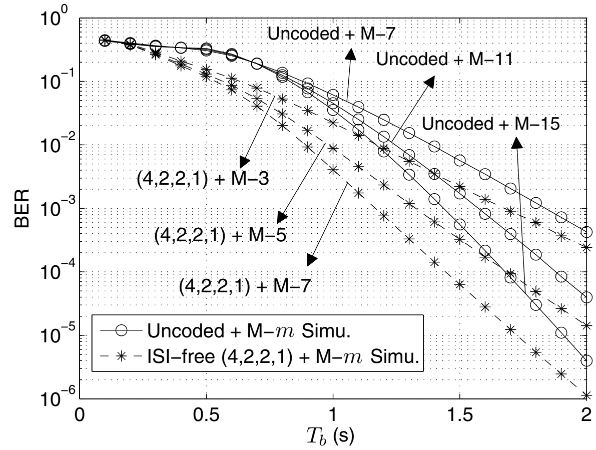


Fig. 7. Performance comparison of the uncoded and the ISI-free-coded system under different m.

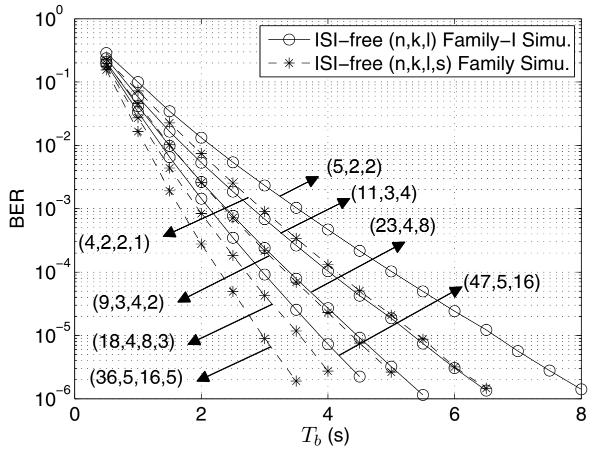


Fig. 6. Performance comparison of the ISI-free (n, k, ℓ, s) code family and the family-I of the ISI-free (n, k, ℓ) codes, both with Gray-coding.

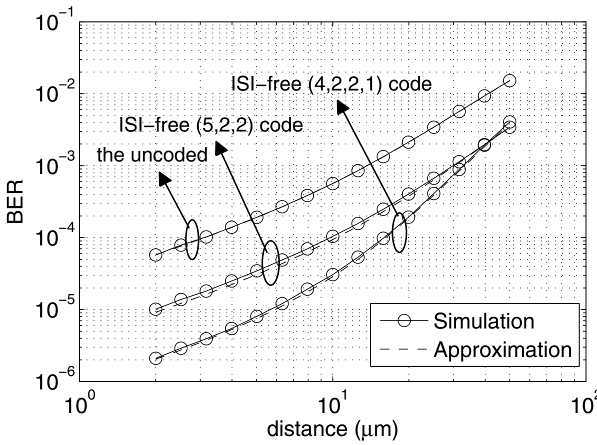


Fig. 8. Performance comparison of the uncoded and the ISI-free-coded system under different distance.

In Fig. 7, the BER performances of the uncoded communication systems and the ISI-free-coded systems are compared with different M-m schemes. One can see that systems coded with the ISI-free $(4, 2, 2, 1)$ code upon M-m are better than the uncoded systems upon M-$(2m + 1)$ when $T_b < 2$ for cases of $m = 2$, $m = 3$, and $m = 4$. Even though the uncoded systems use more molecules in transmission, under the same throughput $1/T_b$, systems coded with the ISI-free $(4, 2, 2, 1)$ code turn out to be more desirable in terms of BER.

Finally, Fig. 8 shows the BER performance due to various distances between the transmitter and the receiver. The figure is obtained using (30) and T_b is set to 5 s. The codeword assignments of the ISI-free $(5, 2, 2)$ code follow Table II and the codeword assignments of the ISI-free $(4, 2, 2, 1)$ code follow Table V.

VIII. Conclusions

In this paper, we have introduced the ISI-free codes and their BER approximations, and then both aspects have been explored further. On the one hand, the approximation functions of the crossover probability have been studied mathematically, and their exponentially-decaying properties justify that the ISI-free index is the dominant factor of the BER performance of ISI-free codes. Also, the BER approximations have been shown to match well with the simulation results of the ISI-free codes. On the other hand, the ISI-free (n, k, ℓ) code families and the ISI-free (n, k, ℓ, s) code family have been proposed, and both of them enjoy arbitrarily high ISI-free indices and hence communication reliability under the same throughput $1/T_b$. Furthermore, those code families have been realized upon the modulation schemes M-m. By applying proper M-m schemes, we have shown that the systems coded with the ISI-free $(4, 2, 2, 1)$ code outperform the uncoded systems even though the uncoded ones use more molecules in transmission.

The effort in this paper serves as an early attempt to design practical channel codes in diffusion-based molecular communications, and the proposed ISI-free codes have proven to be desirable. The principles in evaluating the BER of different channel codes are useful guidelines for researchers to explore promising coding schemes and for designers to analyze performance-cost tradeoff between different systems. As shown in this paper, we have applied the proposed ISI-free codes to the M-m schemes, paving the path of encompassing more type-based schemes in molecular communications. Moreover, the ISI-free codes can be modified to utilize more than two distinguishable kinds of molecules as information carriers, which is left for future work.

REFERENCES

[1] T. Suda, T. Nakano, M. Moore, and K. Fujii, "Biologically inspired approaches to network systems," in *Grid Enabled Remote Instrumentation*, F. Davoli and N. Meyer, Eds., Springer US, 2009, part II, pp. 99–113.

[2] T. Nakano and M. Moore, "Molecular communication paradigm overview," *J. Next Generation Inf. Technol.*, vol. 2, no. 1, pp. 9–16, 2011.

[3] B. Atakan and O. Akan, "An information theoretical approach for molecular communication," in *Proc. IEEE 2nd Bio-Inspired Models Netw., Inf. Comput. Syst.*, Dec. 2007, pp. 33–40.

[4] A. Eckford, "Achievable information rates for molecular communication with distinct molecules," in *Proc. IEEE 2nd Bio-Inspired Models Netw., Inf. Comput. Syst.*, Dec. 2007, pp. 313–315.

[5] T. Nakano and J. Liu, "Design and analysis of molecular relay channels: An information theoretic approach," *IEEE Trans. NanoBiosci.*, vol. 9, no. 3, pp. 213–221, Sep. 2010.

[6] M. Moore, T. Suda, and K. Oiwa, "Molecular communication: Modeling noise effects on information rate," *IEEE Trans. NanoBiosci.*, vol. 8, no. 2, pp. 169–180, June 2009.

[7] A. Eckford, "Nanoscale communication with Brownian motion," in *Proc. IEEE 41st Conf. Inf. Sci. Syst.*, Mar. 2007, pp. 160–165.

[8] S. Kadloor, R. Adve, and A. Eckford, "Molecular communication using Brownian motion with drift," *IEEE Trans. NanoBiosci.*, vol. 11, no. 2, pp. 89–99, June 2012.

[9] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2004.

[10] E. Boutillon, C. Douillard, and G. Montorsi, "Iterative decoding of concatenated convolutional codes: Implementation issues," *Proc. IEEE*, vol. 95, no. 6, pp. 1201–1227, June 2007.

[11] J. Woodard and L. Hanzo, "Comparative study of turbo decoding techniques: An overview," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2208–2233, Nov. 2000.

[12] G. D. Forney, "Codes on graphs: Normal realizations," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 520–548, Feb. 2001.

[13] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[14] R. Koetter and A. Vardy, "Algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2809–2825, Nov. 2003.

[15] P.-J. Shih, C.-H. Lee, and P.-C. Yeh, "Channel codes for mitigating intersymbol interference in diffusion-based molecular communications," in *Proc. IEEE GLOBECOM*, Dec. 2012.

[16] K. R. Sharma, *Principles Of Mass Transfer*. Prentice-Hall Of India Pvt. Limited, 2007.

[17] S. Karlin and H. Taylor, *A First Course in Stochastic Processes*. Academic Press, 1975.

[18] V. Seshadri, *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Springer, 1999.

[19] Y.-P. Hsieh, P.-J. Shih, Y.-C. Lee, P.-C. Yeh, and K.-C. Chen, "An asynchronous communication scheme for molecular communication," in *Proc. IEEE Int. Workshop Molecular Nanoscale Commun. (ICC/MoNaCom)*, June 2012, pp. 7755–7760.

[20] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. F. Akyildiz, "A diffusion-based binary digital communication system," in *Proc. IEEE ICC*, June 2012, pp. 6558–6562.

[21] J. Shuster, "On the inverse Gaussian distribution function," *J. American Statistical Assoc.*, pp. 1514–1516, Dec. 1968.

[22] B. Bajoga and W. Walbesser, "Decoder complexity for BCH codes," *Proc. Institute Electrical Engineers*, vol. 120, no. 4, pp. 429–431, Apr. 1973.

[23] N. Chen and Z. Yan, "Complexity analysis of Reed-Solomon decoding over GF(2m) without using syndromes," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, pp. 16:1–16:11, Jan. 2008.

Po-Jen Shih received his B.S. degree in electrical engineering from National Taiwan University in 2012. From 2011 to 2012, he worked for the Communications and Networks Science (CoWorkS) Laboratory at National Taiwan University and published several papers on channel coding in molecular communications. After graduation, he became more interested in the social dimensions of science and technology, and will pursue related graduate studies. He is currently working as a research assistant under the project Empowerment Network for Social Innovation Talents (ENSIT) in Taiwan.

Chia-Han Lee received his B.S. degree from National Taiwan University, Taipei, Taiwan, in 1999, his M.S. degree from the University of Michigan, Ann Arbor, in 2003, and his Ph.D. from Princeton University in 2008, all in electrical engineering. From 1999 to 2001, he served in the R.O.C. army as a missile operations officer. From 2008 to 2009, he was a postdoctoral research associate at the University of Notre Dame, USA. Since 2010, he has been with Academia Sinica, Taipei, Taiwan, as an Assistant Research Fellow. His research interests include wireless communications, wireless networks, and signal processing.

Ping-Cheng Yeh received his B.S. degree in mathematics and M.S degree in electrical engineering from National Taiwan University, in 1996 and 1998, respectively. In 2005, he received his Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, USA. He joined the Department of Electrical Engineering and the Graduate Institute of Communication Engineering at National Taiwan University in August 2005. His research interests include molecular communications, wireless multimedia transmissions, physical layer security, cooperative communications, cross-layer design in wireless networks, and online education platform design. Dr. Yeh has received various awards in the past, including the EECS Outstanding GSI Award (2002), the University of Michigan Outstanding GSI Award (2003), the NTU Excellence in Teaching Award (2008 & 2009), and the NTU Distinguished Teaching Award (2010). He is currently the Associate Director of the Center for Teaching and Learning Development at National Taiwan University.

Kwang-Cheng Chen received the B.S. from National Taiwan University in 1983, and the M.S. and Ph.D from the University of Maryland, College Park, USA, in 1987 and 1989, respectively, all in electrical engineering. From 1987 to 1998, Dr. Chen worked with SSE, COMSAT, the IBM Thomas J. Watson Research Center, and National Tsing Hua University, in mobile communications and networks. Since 1998, Dr. Chen has been with National Taiwan University, Taipei, Taiwan, ROC, and is the Distinguished Professor and Deputy Dean in Academic Affairs for the College of Electrical Engineering and Computer Science.

Dr. Chen is actively involved in organizing various IEEE conferences as General/TPC chair/co-chair. He has served as editor with a few IEEE journals and many international journals, and has served in various positions for IEEE. Dr. Chen also actively participates in, and has contributed essential technology to, IEEE 802, Bluetooth, and 3GPP wireless standards. He has authored and co-authored over 250 technical papers and more than 20 granted US patents. He co-edits (with R. DeMarca) the book *Mobile WiMAX* published by Wiley in 2008, and the book *Principles of Communications* published by River in 2009. He is co-author (with R. Prasad) of another book *Cognitive Radio Networks* published by Wiley in 2009. Dr. Chen is an IEEE Fellow and has received a number of awards including the 2011 IEEE COMSOC WTC Recognition Award and has co-authored a few award-winning papers published in IEEE ComSoc journals and conferences. Dr. Chen's research interests include wireless communications and network science.