Probabilistic Constellation Shaping for Molecular Communications

Yuankun Tang, Fei Ji, Miaowen Wen, *Senior Member, IEEE*, Qianqian Wang, *Member, IEEE*, Chan-Byoung Chae, *Fellow, IEEE*, and Lie-Liang Yang, *Fellow, IEEE*

***Abstract*—Molecular communication (MC) is an emerging field aiming at realizing information exchange via chemical signals between nanomachines in nanonetworks. Information transmission that is energy-efficient and relies on relatively low-complexity transceiver techniques is of practical importance for MC systems. Based on the fact that constellation shaping can improve energy efficiency, in this paper, we propose a molecular shell mapping (MSM) scheme to implement the probabilistic constellation shaping for MC. The MSM method is designed to exploit the concentration sequences with the lowest sum sequence weight, which results in an energy-efficient signal constellation. Furthermore, we propose an algorithm for selecting and sorting the concentration sequences to mitigate inter-symbol interference. For information detection, we design a genie-aided maximum-likelihood (ML) detector and a realistic ML detector to leverage the constructive effect of intra-sequence interference, as well as derive their bit error rates and achievable rates. Additionally, for the applications using large blocklength sequences, a low-complexity ML detection method is proposed. Numerical simulation results confirm that the shaped signaling using the MSM method is more energy-efficient than the conventional equiprobable signaling, achieving shaping gains of up to 1.5 dB at an ultra-short blocklength of 4.**

***Index Terms*—Molecular communication, probabilistic constellation shaping, inter-symbol interference, concentration shift keying.**

I. Introduction

MOLECULAR communication (MC) is a bio-inspired communication paradigm that involves chemical signals as information carriers [1]. Compared with electromagnetic signaling, MC signals are energy-efficient and biocompatible [2], [3]. These benefits endow MC with a viable communication method for conveying information between nanomachines and constructing the Internet of Bio-Nano Things [4], [5]. In literature, several MC propagation mechanisms have been studied, such as MC via diffusion (MCvD) [2], [6], molecular motor-based MC [7], and bacterial-based MC [8]. MCvD has been regarded as the most prevailing propagation scheme among them, since the energy supplies of nanomachines are limited while MCvD does not need external energy supplies [2]. While energy-saving techniques have been explored in relation to information carriers [1], [2] and propagation methods [2], [6], to the best of our knowledge and based on a survey of modulation techniques for MCvD [9], no research has been done on energy-efficient techniques from a *modulation* perspective for MCvD systems considering a single communication link.

In MC, information can be modulated into various chemical characteristics, including concentration [10], [11], type [12], [13], timing [14], space [15]–[17], or combinations of some of them [18]–[20]. In this paper, we focus on *concentration shift keying (CSK)*, which conveys information via the concentration levels of information molecules of a single type, since it can avoid the energy consumption for the synthesis and release of multiple types of molecules [9]. Generally, the concentration symbol of CSK is equiprobable to be transmitted. This method can yield the maximum transmission rate; however, it does not take into account the energy consumption of the various concentration symbols [9]–[11]. If the low-energy cost symbols are chosen more often than the high-energy cost symbols, at the same average energy, the saved energy can enlarge the minimum distance between the constellation points, and hence improve the reliability of communications.

Probabilistic constellation shaping (PCS), which has attracted significant attention in various communication scenarios, including wireless communications [21], optical communications [22], [23], and underwater communications [24], develops modulated signaling based on a non-equiprobable input distribution, as shown in Fig. 1. PCS adjusts the *probability of occurrence* of the signal points to approach a capacity-achieving distribution. For example, a Gaussian distribution is required to achieve the channel capacity when communicating over additive white Gaussian noise channels [25]. Such non-equiprobable signaling can obtain a shaping gain, which is

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021B1515120067, in part by the open funding of Guangdong Provincial Key Laboratory of Short-Range Wireless Detection and Communication, in part by the National Research Foundation of Korea (NRF) grant through the MSIT, Korea Government (RS-2023-00208922, 2022R1A5A1027646), and in part by the Engineering and Physical Sciences Research Council (EPSRC) under the project EP/X01228X/1. *(Corresponding author: Miaowen Wen.)*

Y. Tang is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China, and also with the Next Generation Wireless Research Group, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: eeyktang@mail.scut.edu.cn).

F. Ji is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: eefeiji@scut.edu.cn).

M. Wen is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China, and also with Guangdong Provincial Key Laboratory of Short-Range Wireless Detection and Communication, China (e-mail: eemwwen@scut.edu.cn).

Q. Wang is with the College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China, and also with Guangdong Provincial Key Laboratory of Short-Range Wireless Detection and Communication, China (e-mail: qianqian.wang@nwnu.edu.cn).

C.-B. Chae is with the School of Integrated Technology, Yonsei University, Seoul 03722, South Korea (e-mail: cbchae@yonsei.ac.kr).

L.-L. Yang is with the Next Generation Wireless Research Group, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lly@ecs.soton.ac.uk).

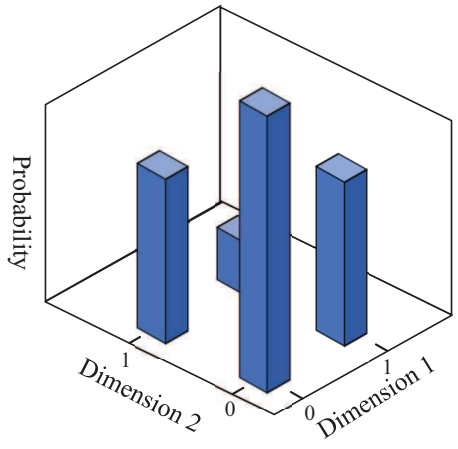

Fig. 1. Constellation of shaped signal for PCS.

defined as the ratio of the energy required by the equiprobable signaling to the energy required by the shaped signaling that achieves the same rate and error performance [22]. For computing the capacity of MC systems, a number of mathematical models were proposed in the literature under different specific constraints [3], [26], [27]. In [27], the authors computed the upper and lower bounds of the capacity over a Poisson channel with finite-state memory. However, the optimal input distribution achieving the capacity does not have a closed-form expression, while it can be found by the numerical Blahut-Arimoto algorithm [27].

For the implementation of PCS, a key module is required to transform the uniformly distributed input bits into the shaped signaling with an approximate target distribution in an invertible manner, where the shaped signaling is a sequence of concentration symbols. Consequently, we call this module the *concentration shaper* for MC systems. Referring to the amplitude shaper studied in optical and wireless communications [21], [22], [28]–[30], the candidate strategies for constructing a concentration shaper can be classified into the *direct* approach and *indirect* approach. As a representative of the former, distribution matching [22], [28] directly generates the concentration sequences having a one-dimensional nonuniform empirical distribution to match the target distribution. Constant composition distribution matching (CCDM) was initially proposed for optical communications [28]. The CCDM outputs the sequences having a constant composition of transmitted symbols, such that it can achieve the empirical distribution to approach the target distribution by adjusting the constant composition. However, the rate loss performance of the CCDM deteriorates as the blocklength decreases. The sphere shaping (SpSh) [21], [29], [30], belonging to the indirect approach, aims at designing an energy-efficient multidimensional constellation to approach the target distribution indirectly. Among the various SpSh methods, enumerative sphere shaping (ESS) lexicographically orders all sequences whose sequence weights are less than a maximum sequence weight constraint, and then outputs the sequences with smaller indices. Shell mapping (SM) [29], [30] is another SpSh method that selects the concentration sequences with the lowest sum sequence cost as the shaping set, and thus constructs the most energy-efficient signal space. Moreover, the SM scheme provides desirable rate loss performance, especially at a short blocklength, which is critical for complexity-limited MC systems. Note that although the properties of CCDM and SM in terms of rate loss performance mentioned above are concluded in the researches of wireless communications or optical communications, these properties are also available in MC. This result can be explained by the fact that the rate loss is caused by the shaping operation at the transmitter, and hence is independent of the channel model. In addition, the design strategies of the above concentration shapers do not take into account *inter-symbol interference (ISI)*.

In this paper, to inherit the merit of SM while mitigating ISI, we propose *molecular shell mapping (MSM)* to implement the PCS for CSK (PCS-CSK). In our MSM, for the sequences with the same weight, those with high-energy cost symbols appearing in the first few positions are sorted first. By doing so, the MSM method is capable of reducing the ISI imposed by the previously emitted concentration sequences, which is referred to as *inter-sequence interference*. Furthermore, we design three maximum-likelihood (ML) detectors to make use of the constructive effect of *intra-sequence interference*, which represents the interference between the symbols within the same sequence. In addition, we derive the bit error rate (BER) and achievable rate (AR) of the proposed MSM method in MCvD systems. BER and AR performance analysis and the corresponding numerical simulations demonstrate that the proposed MSM scheme outperforms the equiprobable signaling schemes for MCvD systems. The main contributions of this paper can be summarized as follows:

- We introduce and compare several design methods of concentration shaping from the aspects of rate loss and blocklengths. In addition, we show that SM is more suitable than other concentration shapers for complexity-limited MC systems. Moreover, we find an appropriate cost function of SM for MC and prove that the SM with this cost function is an informational divergence optimal concentration shaper.
- We propose MSM to implement PCS-CSK to combat ISI in MCvD systems. A genie-aided ML (GML) detector assuming the ideal knowledge of ISI is firstly designed to show the potential of MSM. Then, we design a realistic ML (RML) detector based on the statistics of ISI. Furthermore, we propose a low-complexity ML detection method for operating with large blocklength sequences.
- While the equiprobable multi-level CSK signaling shows worse BER performance than the on-off keying (OOK) in both the uncoded [10] and coded systems [31], our simulation results reveal that the proposed multi-level PCS-CSK exhibits better BER performance than the OOK. Moreover, the MSM assisted signaling can achieve a shaping gain of up to 1.5 dB when compared with the equiprobable signaling.

The rest of this paper is organized as follows. Section II sketches the channel model and optimizes the distribution of

shaped signaling for the MCvD system. Section III analyzes the rate loss of various concentration shapers and proposes the MSM to implement PCS. In Section IV, we design the ML detectors and derive their ARs and BERs. Numerical results are shown in Section V. Finally, conclusions are drawn in Section VI.

Notations: In this paper, we use $\mathbb{E}[\cdot]$ to represent expectation. $\mathbb{I}(\cdot;\cdot)$ and $\mathbb{H}(\cdot|\cdot)$ denote the mutual information and the conditional entropy, respectively. $\mathbb{D}(\cdot\|\cdot)$ is the Kullback-Leibler divergence between two distributions. $\mathcal{O}(\cdot)$ represents the required computational complexity. $|\cdot|$ is the cardinality of a set. The N-symbol sequence $[x_1,\ldots,x_n,\ldots,x_N]$ and the corresponding sequence weight $\sum_{n=1}^{N} x_n$ are shown by $x^{(N)}$ and $W^{(N)}$, respectively. $\mathbf{x}_u$ denotes the N-symbol sequence transmitted in the u-th sequence interval. $\%$ and $\lfloor\cdot\rfloor$ represent the modulo operation and the floor operation, respectively. $n!$ is the factorial of n, and $Q(\cdot)$ denotes the Q-function.

II. System Model

A. Channel Model

The considered MCvD system consists of a point-like transmitter, a channel, and a passive spherical receiver in a 3-D unbounded environment. Since the PCS-CSK is built on the CSK modulation, let us introduce the principle of the CSK modulation first. The CSK conveys information by the number of information molecules of a single type, denoted by $N_{\rm tx}X$, where X is assumed to be a discrete random variable subject to the uniform distribution defined on a symbol set $\mathcal{X}=\{0,1,\ldots,M-1\}$ consisting of M equidistant signal points. The value of $N_{\rm tx}$ is chosen to satisfy the average energy constraint of $\mathbb{E}[N_{\rm tx}X]=E$. The average energy E is defined as the average number of emitted molecules per channel use for MC systems.

For the PCS-CSK, at the beginning of a sequence interval, K input bits are mapped to a sequence compromising N concentration symbols with an equal emission interval $T_{\rm e}$ based on the MSM codebook, which will be shown in detail in Section III. The MSM algorithm chooses 2^K sequences of N concentration symbols with the lowest average weight to form a set $\mathcal{C}_{\rm MSM}=\{\mathbf{s}_1,\ldots,\mathbf{s}_j,\ldots,\mathbf{s}_{2^K}\}$. We assume that a number of sequences denoted by $x_0^{(N)},x_1^{(N)}\ldots,x_u^{(N)},\ldots$ are successively transmitted, where $x_u^{(N)}=[x_u[1],\ldots,x_u[n],\ldots,x_u[N]]\in\mathcal{C}_{\rm MSM}$ denotes the N-symbol sequence transmitted in the u-th sequence interval and $x_u[n]\in\mathcal{X}=\{0,1,\ldots,M-1\}$. To simplify notations, let the vector $\mathbf{x}_u$ denote the N-symbol sequence $x_u^{(N)}$. Next, the transmitter releases $\Delta N_{\rm tx}x_u[n]$ molecules at the beginning of the w-th emission interval to convey the concentration symbols $x_u[n]$, where $w=uN+n$. The parameter Δ is a constellation scaling factor to make the energy consumption satisfy the average energy constraint.

Following Fick's second law of diffusion and uniform concentration assumption [32], the probability of observing an information molecule, which is emitted by the transmitter at time $t_0=0$, inside the receiver at time $t>0$ can be calculated by [2]

$$p(t)=\frac{V_{\rm r}}{(4\pi Dt)^{\frac{3}{2}}}\exp\left(-\frac{d^2}{4Dt}\right), \tag{1}$$

where d is the distance between the transmitter and the center of receiver, $V_{\rm r}=\frac{4}{3}\pi r^3$ is the volume of the receiver with radius r, and D is the diffusion coefficient of information molecules. The receiver is assumed to sample after a time interval of $t_{\max}=d^2/6D$ from the emission instant to maximize the probability of observing the information molecules, as implied in (1). For the sequential transmissions with an emission interval $T_{\rm e}$, the MCvD channel can be characterized by the channel coefficients $h[\ell]=p(\ell T_{\rm e}+t_{\max})$ for $\ell=0,1,\ldots,L_{\rm e}$, where $L_{\rm e}$ denotes the length of channel memory, which results in ISI, if $L_{\rm e}>0$. The channel coefficient $h[\ell]$ indicates the probability of observing an information molecule that is emitted at time $t_0=0$ and received at time $t=\ell T_{\rm e}+t_{\max}$.

At the receiver, the observations corresponding to the N-sequence $\mathbf{x}_u$ are denoted by $\mathbf{y}_u=[y_u[1],\ldots,y_u[n],\ldots,y_u[N]]$, where $y_u[n]$ represents the number of molecules observed at the w-th emission interval. Let $\mathbf{x}^L$ represent the previously transmitted L number of N-sequences $\mathbf{x}_{u-L},\mathbf{x}_{u-L+1},\ldots,\mathbf{x}_{u-1}$ for brevity, where $L=\lfloor L_{\rm e}/N\rfloor$ is the channel memory in terms of the blocklength N. When taking into account the memory of MCvD channel and diffusion noise [33], for a given $\mathbf{x}^L$, $y_u[n]$ approximately follows a Poisson distribution, i.e., $y_u[n]\sim\mathcal{P}\big(\Lambda_n(u)\big)$ [11], with

$$\Lambda_n(u)=\Delta N_{\rm tx}x_u[n]h[0]+\Lambda_{\mathrm{I}_{\rm intra}|\mathbf{x}_{\rm u}}+\Lambda_{\mathrm{I}_{\rm inter}|\mathbf{x}^{\rm L}}+\lambda_{\rm env}, \tag{2}$$

where the first term is the expected number of molecules resulting from the current transmission of $x_u[n]$; $\Lambda_{\mathrm{I}_{\rm intra}|\mathbf{x}_{\rm u}}$ represents the intra-sequence interference imposed by the previously transmitted symbols within the u-th sequence on the current observations for a given $\mathbf{x}_u$; $\Lambda_{\mathrm{I}_{\rm inter}|\mathbf{x}^{\rm L}}$ represents the inter-sequence interference due to the L number of N-sequences $\mathbf{x}^L$ transmitted before $\mathbf{x}_u$; $\lambda_{\rm env}$ is the Poisson parameter for the distribution of the environment noise resulting from the other MC systems.

B. Customizing the Distribution of Shaped Signaling

The key step in constructing a PCS-CSK system is to choose a target source distribution P_X on $\mathcal{X}$ such that this distribution can maximize the mutual information $\mathbb{I}(X;Y)$ under the average energy constraint. However, this optimal input distribution does not have a closed-form expression. Therefore, in this subsection, we choose the maximum entropy distribution as a sub-optimal input distribution leading to a closed-form expression.

Compared with the equiprobable signaling, the non-equiprobable signaling can enlarge the minimum distance between constellation points; it will, however, reduce the entropy of the transmitter output, and hence the transmission rate. To minimize the transmission rate loss of non-equiprobable signaling, we choose the maximum entropy distribution as the target source distribution. The energy consumption in MCvD systems is proportional to the number of emitted molecules [34], which is determined by the concentration symbols for a given $N_{\rm tx}$. Consequently, the maximum entropy distribution

of X with a given average energy is [35, Sec. 12.2]

$$P_{X_\lambda}(x) = \frac{e^{\lambda x}}{\sum_{x' \in \mathcal{X}} e^{\lambda x'}}. \tag{3}$$

Here, λ is non-positive, since the low energy cost symbols are chosen more often than the high energy cost symbols. To make the shaped symbols comply with the average energy constraint, we scale X by a constellation scaling factor Δ, which results in

$$\mathbb{E}\left[\Delta X\right] = \frac{E}{N_{\mathrm{tx}}} = \frac{M-1}{2}. \tag{4}$$

Note that (4) is compatible with the average energy $\mathbb{E}\left[N_{\mathrm{tx}} X\right] = E$ of the uniformly-distributed concentration symbols, since $\mathbb{E}\left[N_{\mathrm{tx}} X\right] = E$ is a special case of (4) when $\Delta = 1$. Given a constellation scaling factor Δ, the input distribution can be chosen as

$$P_{X_\Delta}(x) = P_{X_\lambda}(x), \text{ with } \lambda : \mathbb{E}\left[\Delta X_\lambda\right] = (M-1)/2. \tag{5}$$

The value of λ, for which the condition of (5) is fulfilled, can be efficiently calculated by the bisection method, since $\mathbb{E}\left[X_\lambda\right]$ is a strictly monotonically increasing function of λ. Based on (5) we can know that for any Δ, the distribution $P_{X_\Delta}(x)$ meets the average energy constraint. Now the objective is to find the optimum Δ, denoted by Δ^*, that maximizes the mutual information over all the possible input distributions $P_{X_\Delta}(x)$, i.e.,

$$\Delta^* = \arg\max_{\Delta} \left\{ \mathbb{I}(X_\Delta; Y) \right\}. \tag{6}$$

Note that the mutual information $\mathbb{I}(X_\Delta; Y)$ is a unimodal function of Δ. Therefore, the value of Δ^* can be calculated by the golden section method.

III. Probabilistic Constellation Shaping

A. Principles of Concentration Shaper

In this subsection, we first show the principles of three concentration shapers with the input bits and output sequences having fixed lengths. They map K uniformly distributed input bits to the blocklength N sequences of concentration symbols $\tilde{X}^N \in \mathcal{X}^N$ with the output distribution $P_{\tilde{X}^N}$. Then, we prove that the SM with an appropriate cost function is the optimal concentration shaper that minimizes its informational divergence to the target distribution.

1) Constant Composition Distribution Matching: In this category, we introduce the CCDM, which is widely used in optical communications, as a representative of the distribution matching methods [28]. The CCDM algorithm requires as inputs the symbol set $\mathcal{X}$, blocklength N, and target distribution P_X. The idea of CCDM is to design the set of N-symbol sequences with a fixed empirical distribution $P_{\bar{X}}$ to emulate the desired distribution P_X. To this end, a constant composition constraint is imposed on the output sequences $\tilde{X}^N$, such that all sequences have the same concentration composition, denoted by $n_x \approx N P_X(x),\ x \in \mathcal{X}$, which is the number of times that the symbol x appears in any one sequence. Consequently, the target distribution can be quantized as an empirical distribution as

$$P_{\bar{X}}(x) = \frac{n_x}{N}. \tag{7}$$

The length K of input bits is calculated as $K = \left\lfloor \log_2 \frac{N!}{\prod_{x \in \mathcal{X}} n_x!} \right\rfloor$. Furthermore, the transmission rate of CCDM with N-symbol sequences is

$$R_{\mathrm{t,CCDM}} = \frac{K}{N} = \frac{1}{N} \left\lfloor \log_2 \frac{N!}{\prod_{x \in \mathcal{X}} n_x!} \right\rfloor, \tag{8}$$

in bits per channel use (bpcu). The empirical distribution can converge to the target distribution asymptotically as the blocklength N increases, which will be explained in Fig. 2. However, when N is large, it is infeasible to store the codebook of CCDM. To implement the CCDM effectively, arithmetic coding can be used to perform the mapping from data bits to concentration sequences online [28].

2) Enumerative Sphere Shaping: Unlike the CCDM that emulates the target distribution by a one-dimensional distribution, SpSh algorithms aim at constructing an energy-efficient N-dimensional signal space by designing the set of output sequence comprising N concentration symbols [21], [29], [30]. The N-symbol sequence is denoted as $x^{(N)} = [x_1, \ldots, x_n, \ldots, x_N]$. The SpSh-based scheme starts with a transmission rate given by

$$R_{\mathrm{t,SpSh}} = \frac{K}{N}. \tag{9}$$

Then, according to a certain sorting principle, the SpSh scheme chooses 2^K number of N-symbol sequences $\tilde{X}^{(N)} \in \mathcal{X}^{(N)}$, which construct a codebook $\mathcal{C}$. The letter distribution $P_{\bar{X}}$ of the codebook $\mathcal{C}$, i.e., the probability of drawing a symbol x from the whole codebook, is given by

$$P_{\bar{X}}(x) = \frac{1}{N 2^K} \sum_{x'^{(N)} \in \mathcal{C}} n_x(x'^{(N)}), \tag{10}$$

where $n_x(x'^{(N)}) = |\{i : {x'}_i = x\}|$ is the number of times that the symbol x occurs in $x'^{(N)}$.

The idea of ESS is to use the N-symbol sequences with the maximum-weight constraint W_{ESS} [21]. The weight of an N-symbol sequence is calculated by $W^{(N)} = \sum_{n=1}^{N} x_n$. The parameter W_{ESS} should also be the minimum weight such that the set of N-symbol sequences, defined as $\mathcal{C}_{\mathrm{ESS}} = \left\{ x^{(N)} \middle| W^{(N)} \le W_{\mathrm{ESS}} \right\}$, satisfies

$$|\mathcal{C}_{\mathrm{ESS}}| \ge 2^K. \tag{11}$$

Next, ESS lexicographically orders all $|\mathcal{C}_{\mathrm{ESS}}|$ sequences, and then outputs the first 2^K sequences $\tilde{X}^{(N)} \in \mathcal{C}_{\mathrm{ESS}}$ with smaller indices. The index of a sequence is defined as the number of sequences that are lexicographically smaller. To implement ESS, we can construct an enumerative concentration trellis [21], based on which the mapping of indices to sequences can be realized in an N-step recursive manner.

Remark 1: As ESS orders sequences lexicographically, it is in general unable to yield a set of 2^K sequences having the minimum average energy.

3) Shell Mapping: In this subsection, we introduce the SM algorithm to construct the most energy-efficient signal space. The SM assigns a cost function $C(x)$ to each symbol x in $\mathcal{X}$ [29], [30]. The idea of SM is to sort the N-symbol sequences based on the sequence cost of $\sum_{n=1}^{N} C(x_n)$. Then, SM aims at finding the codebook $\mathcal{C}_{\mathrm{SM}}$ consisting of the 2^K concentration sequences $\tilde{X}^{(N)}$ with the lowest sum sequence cost, which can be described as

$$\mathcal{C}_{\mathrm{SM}} = \min_{\mathcal{C}\subseteq\mathcal{X}^{(N)},|\mathcal{C}|=2^K} \left\{ \sum_{x^{(N)}\in\mathcal{C}} \sum_{n=1}^{N} C(x_n) \right\}. \tag{12}$$

The SM algorithm can be implemented via the divide-and-conquer principle [29] or using the sequential encoding [30].

To derive an appropriate cost function of the SM, the informational divergence to the target distribution of (3) is minimized. Here the informational divergence, also known as Kullback-Leibler divergence, between two distributions $P_{\bar{X}}$ and P_X on $\mathcal{X}$ is defined as [35]

$$\mathbb{D}\left(P_{\bar{X}} \| P_X\right) = \sum_{x\in\mathcal{X}} P_{\bar{X}}(x) \log_2 \frac{P_{\bar{X}}(x)}{P_X(x)}, \tag{13}$$

which reflects the rate loss caused by the shaping operation when using the distribution $P_{\bar{X}}$ to approximate the distribution P_X.

Proposition 1: For MCvD systems, the SM with the cost function of

$$C(x) = x \tag{14}$$

is the optimal concentration shaper that minimizes the informational divergence to the target distribution.

Proof: To minimize the rate loss caused by the shaping operation, (13) suggests to minimize the informational divergence between the output distribution of the concentration shaper $P_{\tilde{X}^{(N)}}$ and the target distribution P_X. Therefore, the divergence-optimal concentration shaper outputs the codebook as

$$\begin{aligned}
\hat{\mathcal{C}} &= \underset{\mathcal{C}\subseteq\mathcal{X}^{(N)},|\mathcal{C}|=2^K}{\arg\min} \left\{ \mathbb{D}\left(P_{\tilde{X}^{(N)}} \| P_X^N\right) \right\} \\
&\overset{(a)}{=} \underset{\mathcal{C}\subseteq\mathcal{X}^{(N)},|\mathcal{C}|=2^K}{\arg\min} \left\{ -\sum_{x^{(N)}\in\mathcal{C}} \log_2 P_X^N\left(x^{(N)}\right) \right\} \\
&\overset{(b)}{=} \underset{\mathcal{C}\subseteq\mathcal{X}^{(N)},|\mathcal{C}|=2^K}{\arg\min} \left\{ -\sum_{x^{(N)}\in\mathcal{C}} \sum_{n=1}^{N} \left(\lambda x_n - \log_2 \sum_{x'\in\mathcal{X}} e^{\lambda x'} \right) \right\} \\
&\overset{(c)}{=} \underset{\mathcal{C}\subseteq\mathcal{X}^{(N)},|\mathcal{C}|=2^K}{\arg\min} \left\{ \sum_{x^{(N)}\in\mathcal{C}} \sum_{n=1}^{N} x_n \right\},
\end{aligned} \tag{15}$$

where $(15a)$ is valid since every sequence $x^{(N)}$ in the codebook $\mathcal{C}$ is chosen with equal probability, i.e., $P_{\tilde{X}^{(N)}}(x^{(N)}) = 2^{-K}$. For a sequence of independent and identically distributed random variables, we have

$$P_X^N\left(x^{(N)}\right) = \prod_{n=1}^{N} P_X(x_n). \tag{16}$$

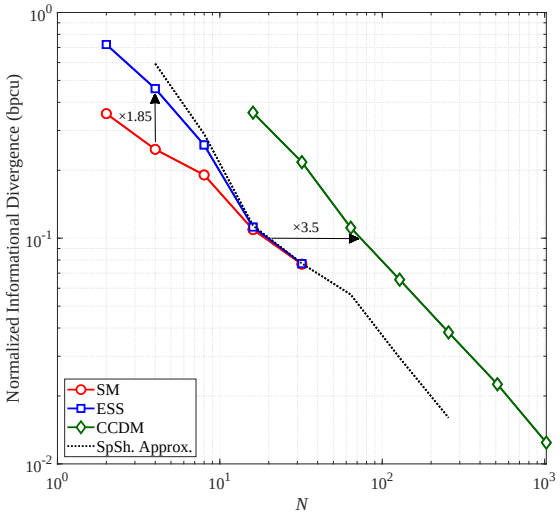

Fig. 2. Rate loss comparison of SM, ESS, and CCDM based on the normalized informational divergence versus different blocklengths N. The target transmission rate is $R_{\mathrm{t}} = 1$ bpcu with QCSK.

After substituting (3) and (16) into the right-hand side of $(15a)$, we obtain $(15b)$. The equality $(15c)$ comes from two facts: a) λ is negative, as P_X is non-uniformly distributed; b) any translation and positive scaling applied on the objective function will not change the optimization and hence the resultant codebook. After substituting the cost function (14) into the objective function of SM (12), we obtain the same result of (15). Therefore, the SM with the cost function of (14) minimizes the rate loss. ■

B. Performance of Concentration Shaper

In this subsection, we compare the rate loss performance of different concentration shapers. For the convenience of comparison, we choose the normalized informational divergence between the output distribution of the concentration shaper $P_{\tilde{X}^{(N)}}$ and the target distribution P_X to measure the rate loss, which can be calculated by [28]

$$\frac{\mathbb{D}\left(P_{\tilde{X}^{(N)}} \| P_X^N\right)}{N} = \sum_{x\in\mathcal{X}} -P_{\bar{X}}(x) \log_2 P_X(x) - R_{\mathrm{t}}, \tag{17}$$

where $P_{\bar{X}}$ and R_{t} are respectively the empirical distribution (or the letter distribution) and the transmission rate of the corresponding concentration shaper, as shown in (7)-(10).

Figure 2 compares the different concentration shapers involving the CCDM, ESS, and SM for the MCvD systems in terms of the normalized informational divergence versus the blocklength N. We consider the quadruple CSK (QCSK) with the alphabet $\mathcal{X} = \{0, 1, 2, 3\}$ and the target transmission rate of $R_{\mathrm{t}} = 1$ bpcu. To satisfy the transmission rate R_{t}, the CCDM and ESS calculate the n_x and W_{ESS} respectively for each blocklength N, such that the N-dimensional signal space contains at least 2^{NR_t} signal points. The SM uses the cost function of (14). When the blocklength N is large, it is cumbersome to calculate the letter distribution of the SpSh schemes in (10). To this end, we use the *partial histograms* [36] that can effectively approximate the letter distribution. We

TABLE I
MSM SET $\mathcal{C}_{\mathrm{MSM}}$ FOR $N=4$, $K=4$, AND $M=4$

I	$x^{(N)}$	$W^{(N)}$	$I^{(N)}$	I	$x^{(N)}$	$W^{(N)}$	$I^{(N)}$
0	$(0,0,0,0)$	0	0	8	$(1,0,1,0)$	2	3
1	$(1,0,0,0)$	1	0	9	$(0,1,1,0)$	2	4
2	$(0,1,0,0)$	1	1	10	$(1,0,0,1)$	2	5
3	$(0,0,1,0)$	1	2	11	$(0,1,0,1)$	2	6
4	$(0,0,0,1)$	1	3	12	$(0,0,2,0)$	2	7
5	$(2,0,0,0)$	2	0	13	$(0,0,1,1)$	2	8
6	$(1,1,0,0)$	2	1	14	$(0,0,0,2)$	2	9
7	$(0,2,0,0)$	2	2	15	$(3,0,0,0)$	3	0

can observe from Fig. 2 that, at $N=4$, the SM can reduce by a factor of about 1.85 in the normalized informational divergence as compared to the ESS. Moreover, at a target divergence of 0.1 bpcu, the SM scheme can be operated with a blocklength that is about 3.5 times smaller than that required by the CCDM scheme. It is worth noting that the complexity of MC systems is critically restricted whilst the blocklength is largely related to the complexity. Hence, the superiority of SM over CCDM and ESS in terms of informational divergence is important for MC applications.

C. Implementation of SM for MC

In this subsection, we design an MSM algorithm to implement the PCS-CSK for information transmission over MCvD channels, where signals are corrupted by ISI and the signal-dependent noise.

The objective of SM (12) with the cost function (14) implicitly indicates that all possible N-symbol sequences $x^{(N)}$ need to be sorted according to the sequence weight $W^{(N)}$. For the sequences with the same weight $W^{(N)}$, we first sort the sequences, whose high energy cost symbols occur first in the sequences, to reduce the interference on the following sequences. This is the core idea of our MSM scheme. Owing to this arrangement, MSM can mitigate the inter-sequence interference. This is expected since a high-energy cost symbol interferes more severely with the subsequent sequences than a low-energy cost symbol at the same position. The MSM scheme can be implemented with the aid of a look-up table.

Example 1 (Look-up table for MSM): For ease of understanding, Table I presents an example of MSM for $N=4$, $K=4$, and $M=4$, where an index I of information bits is one-to-one mapped to a symbol sequence $x^{(N)}$. With the MSM scheme, all the 2^K sequences $x^{(N)}$ are sorted based on their weights $W^{(N)}$, as shown in the third column of Table I. For the sequences having the same weight, we assign a smaller index $I^{(N)}$ to the sequence whose high energy cost symbols occur in the earlier positions, where $I^{(N)}$ represents the index of N-symbol sequence with the weight $W^{(N)}$. For example, for the sequences with the weight $W^{(N)}=2$, $(2,0,0,0)$ is assigned the smallest index $I^{(N)}=0$, since all the sequence weight is given by symbol 2 in the first position. By contrast, the highest index is allocated to $(0,0,0,2)$, due to the high-energy cost symbol being the last element. In this way, the 2^K sequences $x^{(N)}$ are ordered and assigned the corresponding indices of I.

However, at a large blocklength, implementing the MSM via a look-up table requires huge memory to buffer the table at transceivers. Furthermore, it is cumbersome to sort all the symbol sequences directly. Therefore, to simplify the sorting problem, we propose a divide-and-conquer algorithm, which is shown in Appendix A. According to this sorting algorithm, we design the corresponding MSM decoder, which maps a sequence of N concentration symbols to an index I representing K information bits, and hence avoid buffering a huge table at transceivers. The detailed decoding process is described in Algorithm 1 of Appendix B. The inverse mapping, i.e., MSM encoder, is given in Algorithm 2 of Appendix C.

IV. RECEIVER DESIGN AND PERFORMANCE ANALYSIS

A. Receiver Design

Since the concentration samples are mutually independent provided that $T_{\mathrm{e}} > r^2/D$ [33], the conditional probability mass function of $\mathbf{y}_u$ is given as

$$f\left(\mathbf{y}_u \left| \mathbf{x}_u, \mathbf{x}^L \right.\right) = \prod_{n=1}^{N} \frac{1}{y_u[n]!} \left(\Lambda_n(u)\right)^{y_u[n]} e^{-\Lambda_n(u)}. \quad (18)$$

Let us employ a GML sequence detector, which assumes that the ideal knowledge about $\Lambda_{\mathrm{I_{inter}}|\mathbf{x}^{\mathrm{L}}}$ is available at the receiver, to reveal the potential performance of PCS-CSK. Based on (2) and (18), the GML sequence detection can be achieved via solving the problem

$$\begin{aligned}\hat{\mathbf{x}}_u &= \underset{\mathbf{x}_u \in \mathcal{C}_{\mathrm{MSM}}}{\arg\max} \left\{ f\left(\mathbf{y}_u \left| \mathbf{x}_u, \mathbf{x}^L \right.\right) \right\} \\ &= \underset{\mathbf{x}_u \in \mathcal{C}_{\mathrm{MSM}}}{\arg\max} \left\{ \sum_{n=1}^{N} y_u[n] \ln\left(\Lambda_n\left(u \left| \mathbf{x}_u \right.\right)\right) - \Lambda_n\left(u \left| \mathbf{x}_u \right.\right) \right\}. \end{aligned} \quad (19)$$

After obtaining $\hat{\mathbf{x}}_u$, the receiver maps it to an index of the K information bits based on the mapping table or the decoding algorithm of MSM, as discussed in Section III. Note that although concentration symbols $x_u[n]$ are non-uniformly distributed, the 2^K possible N-sequences $\mathbf{x}_u$ are selected equiprobably. Hence, the ML detection of (19) is equivalent to the maximum a posteriori detection.

However, the GML detection is impractical, as obtaining the exact knowledge of $\mathbf{x}^L$ at the receiver is impossible. Therefore, we propose an RML detector, which uses the expectation of inter-sequence interference, denoted by Λ_n, to replace $\Lambda_{\mathrm{I_{inter}}|\mathbf{x}^{\mathrm{L}}}$. Specifically, for the n-th symbol within the current N-sequence, Λ_n can be calculated as

$$\Lambda_n = \mathbb{E}\left[\Lambda_{\mathrm{I_{inter}}|\mathbf{x}^{\mathrm{L}}}\right] = E \sum_{\ell=n}^{L_{\mathrm{e}}} h[\ell]. \quad (20)$$

Then, the detection criterion of the RML detector can be derived by substituting (2) and (20) into (19).

Due to the MSM codebook $\mathcal{C}_{\mathrm{MSM}} \subseteq \mathcal{X}^{(N)}$, the complexity of both GML and RML is up to $\mathcal{O}(M^N)$, which is intolerable even for a medium blocklength N. To this end, we propose a

low-complexity detection method, which first detects the front $N/2$ symbols of an N-sequence $\mathbf{x}_u$, represented as $\mathbf{x}_{u,\mathrm{f}}$, by

$$\hat{\mathbf{x}}_{u,\mathrm{f}} = \underset{\mathbf{x}_{u,\mathrm{f}} \in \mathcal{C}_{W_{\max}^{(N)}}}{\arg\max} \Big\{ \sum_{n=1}^{N/2} y_u[n] \ln\left(\Lambda_n\left(u\,|\mathbf{x}_{u,\mathrm{f}}\right)\right) - \Lambda_n\left(u\,|\mathbf{x}_{u,\mathrm{f}}\right) \Big\}, \tag{21}$$

where $\mathcal{C}_{W_{\max}^{(N)}}$ is the MSM codebook consisting of $N/2$-sequences with a maximum sequence weight constraint $W_{\max}^{(N)}$ that is calculated as

$$W_{\max}^{(N)} = \arg\min_{W} \{ z^{(N)}(W+1) \geq 2^K \}, \tag{22}$$

where $z^{(N)}(W)$ is the number of N-symbol sequences with weights less than W. Then, similarly, the other $N/2$ symbols of $\mathbf{x}_u$, denoted by $\mathbf{x}_{u,\mathrm{l}}$, is detected via

$$\hat{\mathbf{x}}_{u,\mathrm{l}} = \underset{\mathbf{x}_{u,\mathrm{l}} \in \mathcal{C}_{W_{\mathrm{l}}^{(N/2)}}}{\arg\max} \Big\{ \sum_{n=\frac{N}{2}+1}^{N} y_u[n] \ln\left(\Lambda_n\left(u\,|\mathbf{x}_{u,\mathrm{l}}\,, \hat{\mathbf{x}}_{u,\mathrm{f}}\right)\right) - \Lambda_n\left(u\,|\mathbf{x}_{u,\mathrm{l}}, \hat{\mathbf{x}}_{u,\mathrm{f}}\right) \Big\}, \tag{23}$$

where $W_{\mathrm{l}}^{(N/2)} = W_{\max}^{(N)} - W_{\mathrm{f}}^{(N/2)}$, and $W_{\mathrm{f}}^{(N/2)}$ represents the weight of $\hat{\mathbf{x}}_{u,\mathrm{f}}$. The arguments of the low-complexity ML detector indicate that it can reduce the computational complexity to $\mathcal{O}(M^{\frac{N}{2}})$. The low-complexity detection above is compatible with both the GML and RML detectors.

Remark 2: The complexity of detection can be further reduced by dividing the N-sequence into more parts with smaller blocklength and detecting them separately. Eventually, the sequence detector degrades into a symbol-by-symbol detector. However, with the reduction of blocklength, the detection will achieve degraded performance. Therefore, this approach allows to strike a trade-off between complexity and detection performance.

B. Bit Error Rate

In this subsection, the BER upper bounds of both the GML and RML detectors are derived. Assume that $\mathbf{x}_u = \mathbf{s}_j$ is transmitted. Then, according to (19), $\mathbf{s}_j$ is incorrectly detected as $\mathbf{s}_{j'}$ with $j' \neq j$, if

$$\sum_{n=1}^{N} y_u[n] \ln\left(\frac{\Lambda_n\left(u\,|\mathbf{s}_j\right)}{\Lambda_n\left(u\,|\mathbf{s}_{j'}\right)}\right) - \Lambda_n\left(u\,|\mathbf{s}_j\right) + \Lambda_n\left(u\,|\mathbf{s}_{j'}\right) \leq 0. \tag{24}$$

When $\Lambda_n(u)$ is large enough, the Poisson distribution for $y_u[n]$ can be approximated as the Gaussian distribution, i.e., $y_u[n] \sim \mathcal{N}\big(\Lambda_n(u), \Lambda_n(u)\big)$ [11]. Since the observations $y_u[n]$ are independent in terms of n, the left-hand side of (24) can be approximated by a Gaussian distributed random variable Y. Given $\mathbf{x}^L$, $\mathbf{s}_j$, and $\mathbf{s}_{j'}$, the mean and variance of Y are respectively given by

$$\begin{aligned} \mu_{Y|\mathbf{x}^L} &= \sum_{n=1}^{N} \Lambda_n\left(u\,|\mathbf{s}_j\right) \ln\left(\frac{\Lambda_n\left(u\,|\mathbf{s}_j\right)}{\Lambda_n\left(u\,|\mathbf{s}_{j'}\right)}\right) - \Lambda_n\left(u\,|\mathbf{s}_j\right) \\ &\quad + \Lambda_n\left(u\,|\mathbf{s}_{j'}\right), \\ \sigma^2_{Y|\mathbf{x}^L} &= \sum_{n=1}^{N} \Lambda_n\left(u\,|\mathbf{s}_j\right) \ln^2\left(\frac{\Lambda_n\left(u\,|\mathbf{s}_j\right)}{\Lambda_n\left(u\,|\mathbf{s}_{j'}\right)}\right). \end{aligned} \tag{25}$$

According to (25), the conditional pairwise error probability that the detector erroneously detects $\mathbf{s}_j$ as $\mathbf{s}_{j'}$ is

$$\Pr\left(\mathbf{s}_j \to \mathbf{s}_{j'} \,|\mathbf{x}^L\right) = Q\left(\frac{\mu_{Y|\mathbf{x}^L}}{\sigma_{Y|\mathbf{x}^L}}\right), \tag{26}$$

where $Q(x) = (2\pi)^{-1/2} \int_x^\infty e^{-t^2/2} dt$. The average pairwise error probability can be obtained by taking all the possible $\mathbf{x}^L$ into account, which can be expressed as

$$\Pr\left(\mathbf{s}_j \to \mathbf{s}_{j'}\right) = \frac{1}{2^{KL}} \sum\nolimits_{\mathbf{x}^L \in \mathcal{C}_{\mathrm{MSM}}^L} Q\left(\frac{\mu_{Y|\mathbf{x}^L}}{\sigma_{Y|\mathbf{x}^L}}\right). \tag{27}$$

According to the union-bounding technique, an upper bound for the BER can be expressed as

$$P_{\mathrm{ber}} \leq \frac{1}{K 2^K} \sum_{j=1}^{2^K} \sum_{j'=1, j' \neq j}^{2^K} \Pr\left(\mathbf{s}_j \to \mathbf{s}_{j'}\right) e\left(\mathbf{s}_j \to \mathbf{s}_{j'}\right), \tag{28}$$

where $e\left(\mathbf{s}_j \to \mathbf{s}_{j'}\right)$ denotes the number of erroneous bits in the case that $\mathbf{s}_j$ is erroneously detected as $\mathbf{s}_{j'}$. Note that computing (27) is an arduous task when K or L is large. To this end, we can use the Monte Carlo approach to obtain an accurate evaluation of $\Pr\left(\mathbf{s}_j \to \mathbf{s}_{j'}\right)$ [11].

C. Achievable Rate

In this subsection, we analyze the ergodic AR of PCS-CSK when assuming the ideal knowledge of inter-sequence interference $\Lambda_{\mathrm{I}_{\mathrm{inter}}|\mathbf{x}^{\mathrm{L}}}$ or its expectation Λ_n at the receiver. The ergodic AR of PCS-CSK is defined as [37]

$$R = \frac{1}{N\left|\mathcal{C}_{\mathrm{MSM}}^L\right|} \sum_{\mathbf{x}^L \in \mathcal{C}_{\mathrm{MSM}}^L} \mathbb{I}\big(\mathbf{x}_u; \mathbf{y}_u \,|\mathbf{x}^L\big). \tag{29}$$

According to the definition of mutual information, R can be expressed as [35]

$$R = R_{\mathrm{t}} - \frac{1}{N\left|\mathcal{C}_{\mathrm{MSM}}^L\right|} \sum_{\mathbf{x}^L \in \mathcal{C}_{\mathrm{MSM}}^L} \mathbb{H}\big(\mathbf{x}_u \,|\mathbf{y}_u, \mathbf{x}^L\big). \tag{30}$$

In the sequel, we turn to calculate $\mathbb{H}\big(\mathbf{x}_u \,|\mathbf{y}_u, \mathbf{x}^L\big)$ in (30), i.e.,

$$\begin{aligned} \mathbb{H}\left(\mathbf{x}_u \,|\mathbf{y}_u, \mathbf{x}^L\right) &= -\sum_{j=1}^{2^K} \int_{\mathbf{y}_u} \Pr\left(\mathbf{x}_u = \mathbf{s}_j\right) f\left(\mathbf{y}_u \,|\mathbf{x}_u = \mathbf{s}_j, \mathbf{x}^L\right) \\ &\quad \times \log_2\left(\frac{\Pr\left(\mathbf{x}_u = \mathbf{s}_j\right) f\left(\mathbf{y}_u \,|\mathbf{x}_u = \mathbf{s}_j, \mathbf{x}^L\right)}{f\left(\mathbf{y}_u \,|\mathbf{x}^L\right)}\right) d\mathbf{y}_u, \end{aligned} \tag{31}$$

where $\Pr\left(\mathbf{x}_u = \mathbf{s}_j\right) = 2^{-K}$, $f\left(\mathbf{y}_u \,|\mathbf{x}_u = \mathbf{s}_j, \mathbf{x}^L\right)$ is given by (18), and

$$f\left(\mathbf{y}_u \,|\mathbf{x}^L\right) = \frac{1}{2^K} \sum_{j'=1}^{2^K} \prod_{n=1}^{N} \frac{1}{y_u[n]!} \left(\Lambda_n\left(u\,|\mathbf{s}_{j'}\right)\right)^{y_u[n]} \times e^{-\Lambda_n\left(u|\mathbf{s}_{j'}\right)}. \tag{32}$$

TABLE II
PARAMETERS USED FOR PERFORMANCE STUDY

Parameter	Variable	Value
Blocklength	N	$\{2, 4, 6, 8\}$
Cardinality of symbol set	M	$\{2, 4, 8\}$
Radius of receiver	r	5 μm
Communication distance	d	25 μm
Diffusion coefficient	D	2.2×10^{-9} m^2/s
Transmission rate	R_{t}	$[0.75, 2]$ bpcu
Data rate	R_{b}	$[2, 10]$ bps
Emission interval	T_{e}	$[0.1, 0.25]$ s
Mean of environment noise	λ_{env}	0.0001
Length of channel memory	L_{e}	$[3, 28]$

Upon substituting these results into (31) and carrying on some simplification, we obtain

$$\mathbb{H}\left(\mathbf{x}_u \left| \mathbf{y}_u, \mathbf{x}^L \right.\right) = \frac{1}{2^K} \sum_{j=1}^{2^K} \mathbb{E}_{\mathbf{y}_u|\mathbf{x}_u=\mathbf{s}_j,\mathbf{x}^L} \left[\log_2 \left(\sum_{j'=1}^{2^K} \prod_{n=1}^{N} \left(\frac{\Lambda_n\left(u \left| \mathbf{s}_j \right.\right)}{\Lambda_n\left(u \left| \mathbf{s}_{j'} \right.\right)} \right)^{y_u(n)} e^{\Lambda_n(u|\mathbf{s}_j) - \Lambda_n\left(u\left|\mathbf{s}_{j'}\right.\right)} \right) \right]. \tag{33}$$

Finally, when substituting (33) into (30), the AR of PCS-CSK is

$$R = R_{\mathrm{t}} - \frac{1}{N\left|\mathcal{C}_{\mathrm{MSM}}^L\right| 2^K} \sum_{\mathbf{x}^L \in \mathcal{C}_{\mathrm{MSM}}^L} \sum_{j=1}^{2^K} \mathbb{E}_{\mathbf{y}_u|\mathbf{x}_u=\mathbf{s}_j,\mathbf{x}^L} \left[\log_2 \left(\sum_{j'=1}^{2^K} \prod_{n=1}^{N} \left(\frac{\Lambda_n\left(u \left| \mathbf{s}_j \right.\right)}{\Lambda_n\left(u \left| \mathbf{s}_{j'} \right.\right)} \right)^{y_u(n)} e^{\Lambda_n(u|\mathbf{s}_j) - \Lambda_n\left(u\left|\mathbf{s}_{j'}\right.\right)} \right) \right], \tag{34}$$

which has the unit of bpcu.

Note that the derived AR (34) is applicable to both the GML and RML detectors by substituting the corresponding Poisson parameters $\Lambda_n(u)$ into (34).

V. NUMERICAL RESULTS

In this section, we first investigate the impact of the constellation scaling factor on the BER and AR performance of the PCS-CSK systems. Next, we compare the ordering method for the sequences having the same weight of MSM with that of SM [38]. We then compare the proposed PCS-CSK employing MSM with the conventional equiprobable signaling schemes. Finally, for a relatively large blocklength, we evaluate the BER performance of MSM when the low-complexity detection method is employed.

To compare the AR performance fairly, the proposed PCS-CSK uses the same average number of emitted molecules per channel use E and emission interval T_{e} as the uniformly distributed signaling. For a fair comparison of BER performance, we maintain the same average number of emitted molecules per bit E_{b} and also the same data rate R_{b}, where $E_{\mathrm{b}} = E/R_t$ and $R_{\mathrm{b}} = R_{\mathrm{t}}/T_{\mathrm{e}}$. The parameters used in the performance study are listed in Table II. In addition, for the BER and AR results, the RML and GML detectors are employed, respectively, unless otherwise stated.

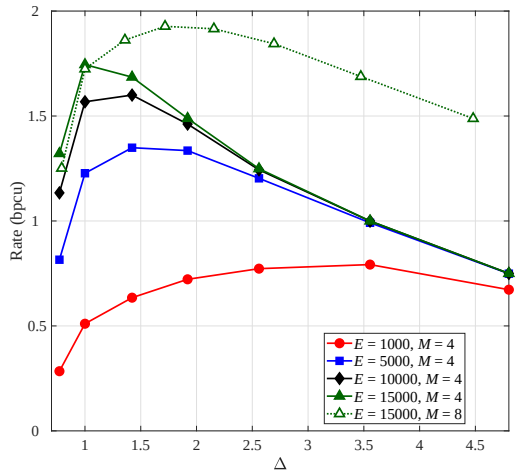

Fig. 3. Comparison of achievable rate of MSM versus the constellation scaling factor Δ, for $N = 4$, $T_{\mathrm{e}} = 0.25$ s, and different M and E.

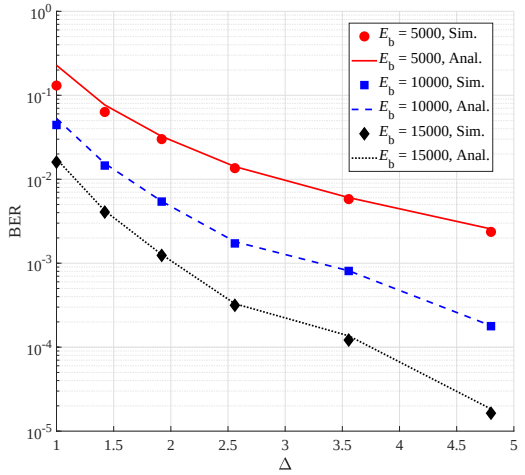

Fig. 4. Comparison of BER of MSM versus the constellation scaling factor Δ, for $N = 4$, $M = 4$, $R_{\mathrm{b}} = 2$ bps and different E_{b}.

A. Impact of Constellation Scaling Factor Δ

Figure 3 illustrates the impact of the constellation scaling factor Δ on the AR of the PCS-CSK systems employing MSM for $N = 4$, $T_{\mathrm{e}} = 0.25$ s, and different values of E and M. It can be observed that AR is a unimodal function of Δ. This is expected because, under the average energy constraint, a constellation scaling Δ close to 0 or much larger than 1 will substantially reduce the entropy of the transmitter output, leading to the deterioration of the transmission rate R_{t}. As shown by the solid lines in Fig. 3, for a larger E, the maximum AR is obtained at a smaller Δ. In particular, when $E = 15000$ and $M = 4$, the uniformly distributed signaling (corresponding to $\Delta = 1$) maximizes the AR. This is because MSM constructs an energy-efficient signal space at the cost of decreasing the transmission rates R_{t}. Therefore, when the available energy is sufficient, e.g., $E = 15000$, we can employ a denser signal constellation ($M = 8$) shown by the dotted curve in Fig. 3 to improve R_{t}. Then, the MSM algorithm can achieve a higher AR than the uniformly distributed signaling with $1 < \Delta < 3$.

In Fig. 4, we present the simulated and analytical BERs of

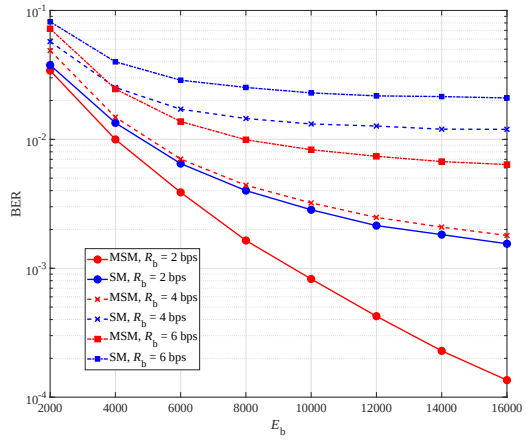

Fig. 5. BER comparison between the ordering method of SM and that of MSM versus the average number of emitted molecules per bit E_{b}, for $R_{\mathrm{t}} = 1$ bpcu, $M = 4$, $N = 4$, and different data rates R_{b}.

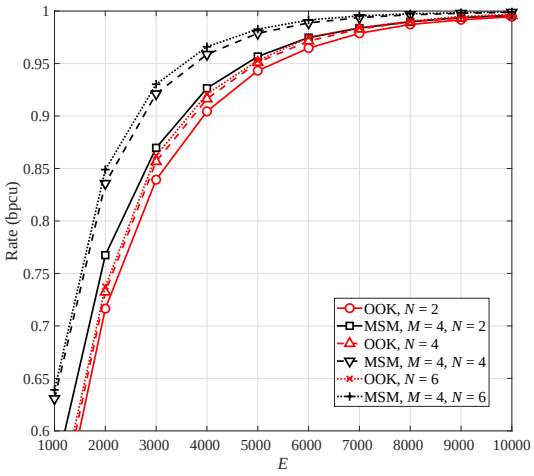

Fig. 6. Comparison of achievable rate between MSM and OOK versus E, for $R_{\mathrm{t}} = 1$ bpcu, $T_{\mathrm{e}} = 0.1$ s, and different blocklengths N.

MSM in terms of Δ, considering $M = 4$, $N = 4$, $R_{\mathrm{b}} = 2$ bits per second (bps), and different E_{b}. It is observed that the theoretical BER matches well with the simulated counterpart. Moreover, as Δ increases, the BER performance improves in any of the three cases of E_{b}. This can be attributed to the fact that increasing Δ can enlarge the minimum distance between signal points, which reduces the pairwise error probability of (27).

B. Comparison of Different Ordering Methods

Figure 5 compares the ordering method for the sequences having the same weight of MSM with that of SM [38] in terms of BER performance. In this BER comparison, we choose $R_{\mathrm{t}} = 1$ bpcu, $M = 4$, and $N = 4$. We assume that the data rates are $R_{\mathrm{b}} \in \{2, 4, 6\}$ bps that reflect the considered channel memory lengths L_{e}. It is observed that the ordering method of MSM outperforms that of SM for all data rates. This can be explained by two facts: a) the ISI effect is not considered in SM but in MSM; b) MSM can not only mitigate the inter-sequence interference but also leverage the constructive effect of the intra-sequence interference when equipping the ML detector.

C. Equiprobable Signaling vs. Shaped Signaling

For a fair comparison of the shaped non-equiprobable signaling and conventional equiprobable signaling, the cardinality M of the symbol set for the PCS-CSK employing MSM must be increased to allow the MSM to operate at the same transmission rate R_{t} as the conventional equiprobable signaling. Furthermore, in both cases, we assume the ML sequence detection.

We compare MSM with OOK in terms of the AR and investigate the impact of blocklength N in Fig. 6. The parameters are chosen as: $M = 4$ for MSM, $T_{\mathrm{e}} = 0.1$ s, $R_{\mathrm{t}} = 1$ bpcu, and $N \in \{2, 4, 6\}$. As shown in Fig. 6, both MSM and OOK present better performance at larger blocklengths N. This is expected since the increase of N results in the reduced pairwise error probability normalized by N. Furthermore, it

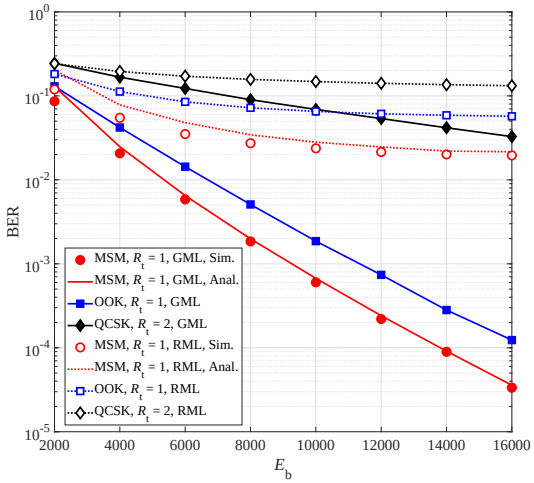

Fig. 7. Comparison of BER between MSM, OOK, and QCSK versus E_{b}, for $N = 4$, $R_{\mathrm{b}} = 10$ bps, and different detectors.

can be observed that as N increases, the AR gain of MSM over OOK improves. However, the increment of the AR gain decreases with the increase of N. This can be explained by two facts: a) a larger N decreases the rate loss due to the shaping, as shown in Fig. 2; b) when $N \geq 4$, further increasing the blocklength slightly only makes a small contribution to compensating for the rate loss. From the results of Fig. 6, we are implied that a blocklength of 4 is usually enough to achieve a desired AR gain, and it is dispensable to further increase N at the cost of system complexity.

Figure 7 shows the BER upper bound and comparison of the BERs of MSM and OOK at the same transmission rate $R_{\mathrm{t}} = 1$ bpcu. To illustrate the performance improvement by MSM, the BER of QCSK without invoking MSM is also depicted. Note that, to maintain the same data rate R_{b}, the emission interval of the conventional QCSK scheme is twice that of MSM and OOK schemes. We assume $R_{\mathrm{b}} = 10$ bps, $N = 4$, and the RML and GML detectors. As shown in Fig. 7, the BER upper bounds of MSM with the RML and GML detectors become asymptotically tighter as E_{b} increases. It

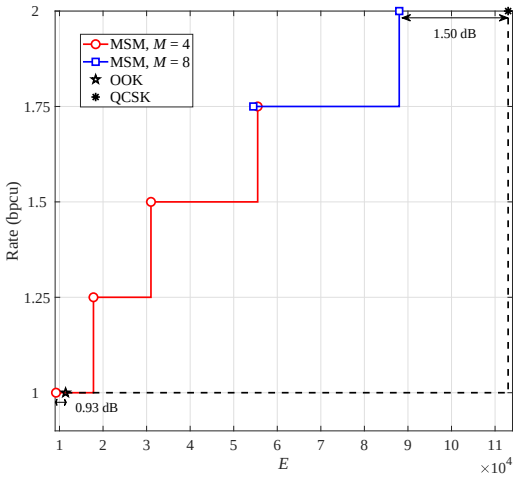

Fig. 8. Comparison of achievable rate at a BER of 10^{-3} between non-equiprobable and equiprobable signaling for $T_{\mathrm{e}} = 0.1$ s and $N = 4$.

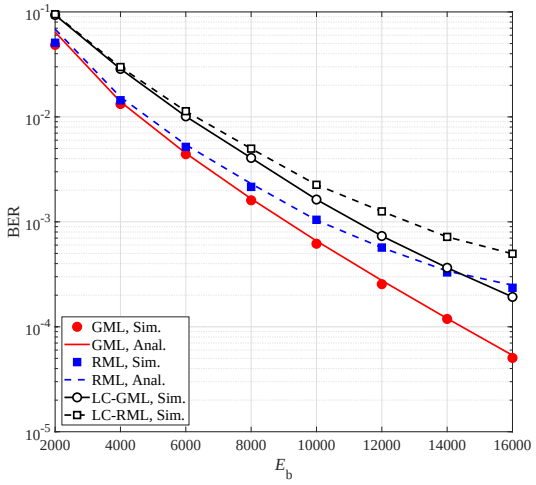

Fig. 9. BER comparison of the MSM assisted PCS-CSK systems with different detectors, for $M = 4$, $R_{\mathrm{t}} = 1$ bpcu, $R_{\mathrm{b}} = 2$ bps, and $N = 8$.

is also observed that the MSM scheme provides significant improvement of the BER performance, when compared to QCSK. This is expected since MSM is capable of using the energy more efficiently and reducing the impairment of signal-dependent noise. Moreover, compared with OOK, MSM can save about 2000 molecules for transmitting one bit at a BER of 10^{-3}, when equipped with the GML detector. This results from the sorting algorithm of MSM, which aggregates the sequence weight into the first few positions of an N-sequence, thereby decreasing the inter-sequence interference while increasing the intra-sequence interference that can be leveraged by the ML detectors.

In Fig. 8, we display the average number of emitted molecules per channel use required to obtain a BER of 10^{-3}, when employing the shaped signaling and the uniformly distributed signaling at different transmission rates R_{t}. The blocklength of $N = 4$ is chosen to show that the MSM can obtain a desired shaping gain at an extremely short blocklength. As shown in Fig. 8, at $R_{\mathrm{t}} = 1$ bpcu, the PCS-CSK employing MSM outperforms OOK by about a shaping gain of 0.93 dB. At $R_{\mathrm{t}} = 2$ bpcu, this gain can reach about 1.50 dB, which is close to the upper bound of the shaping gain of 1.53 dB over additive white Gaussian noise channels [22]. In addition, we observe that the MSM can provide a finer rate granularity than the uniformly distributed signaling.

D. Low-Complexity Detector for Large Blocklength

In Fig. 9, we compare the BER attained by the low-complexity detection with that by the GML and RML at a relatively large blocklength of $N = 8$ for the PCS-CSK systems employing the MSM. The parameters are chosen as: $M = 4$, $R_{\mathrm{t}} = 1$ bpcu, and $R_{\mathrm{b}} = 2$ bps. We can observe that the low-complexity detectors can achieve similar BER performance to the GML and RML detectors, and obtain a BER of 10^{-3} by emitting 3000 extra molecules per bit than the RML detector.

VI. Conclusion

In this paper, we have proposed an MSM scheme as the concentration shaper for the PCS-CSK systems. By using the concentration sequences with the lowest average sequence weight, the MSM can construct the most energy-efficient signal space. We have proved that MSM is an informational divergence optimal concentration shaper that has better rate loss performance than CCDM and ESS. Furthermore, it has been shown that the superiority of MSM in rate loss grows as the blocklength decreases, which renders MSM suitable for applications in MC systems, where low-complexity transceiver is critical. Additionally, to mitigate the inter-sequence interference, the proposed MSM can aggregate the number of emitted molecules in the first few positions of a sequence, which is beneficial to the ML-based detection. Therefore, we have combined MSM with three ML detectors to leverage the constructive effect of intra-sequence interferenceïijŃwhere the trade-off between complexity and detection performance has been considered. The numerical results have demonstrated that MSM is capable of providing better BER and AR performance than the uniformly distributed signaling. A shaping gain of 0.93 dB can be achieved by the MSM scheme over the prevailing OOK scheme.

Appendix A
Sorting Algorithm of MSM

With this sorting algorithm, an n_{c}-dimensional sorting problem of $x^{(n_{\mathrm{c}})}$ is split into two lower-dimensional ones expressed as $x_{\mathrm{f}}^{(n_{\mathrm{f}})}$ and $x_{\mathrm{l}}^{(n_{\mathrm{l}})}$, comprising the first n_{f} symbols and the last n_{l} symbols of $x^{(n_{\mathrm{c}})}$, respectively, with $n_{\mathrm{c}} = n_{\mathrm{f}} + n_{\mathrm{l}}$. This process continues on the obtained $x_{\mathrm{f}}^{(n_{\mathrm{f}})}$ and $x_{\mathrm{l}}^{(n_{\mathrm{l}})}$ in a recursive way, until getting $n_{\mathrm{c}}/2$ number of two-dimensional problems. Let us define the weights and indices of $x_{\mathrm{f}}^{(n_{\mathrm{f}})}$ and $x_{\mathrm{l}}^{(n_{\mathrm{l}})}$ as $W_{\mathrm{f}}^{(n_{\mathrm{f}})}$ and $I_{\mathrm{f}}^{(n_{\mathrm{f}})}$, as well as $W_{\mathrm{l}}^{(n_{\mathrm{l}})}$ and $I_{\mathrm{l}}^{(n_{\mathrm{l}})}$, respectively. Then, given $W_{\mathrm{f}}^{(n_{\mathrm{f}})}$, $W_{\mathrm{l}}^{(n_{\mathrm{l}})}$, $I_{\mathrm{f}}^{(n_{\mathrm{f}})}$, and $I_{\mathrm{l}}^{(n_{\mathrm{l}})}$, the MSM algorithm sorts the n_{c}-symbol sequences with the same weight $W^{(n_{\mathrm{c}})}$ based on the following operations:

- First, a sequence whose last n_{l} symbols $x_{\mathrm{l}}^{(n_{\mathrm{l}})}$ have a lower weight $W_{\mathrm{l}}^{(n_{\mathrm{l}})}$ is listed first. For example, in Table

Algorithm 1 Decoder of MSM

Require: N, $x^{(N)} = [x_1, x_2, \ldots, x_{2i-1}, x_{2i}, \ldots, x_{N-1}, x_N]$
Ensure: I
for $i = 1$ to $N/2$ **do**
  compute W_i^2 and I_i^2 by (35)
end for
$W_f^{(n_\mathrm{f})} \leftarrow W_1^2$, $I_f^{(n_\mathrm{f})} \leftarrow I_1^2$, $W_l^{(n_\mathrm{l})} \leftarrow W_2^2$, $I_l^{(n_\mathrm{l})} \leftarrow I_2^2$
for $j = 2$ to $N/2$ **do**
  $n_\mathrm{c} \leftarrow 2j$
  compute $W^{(n_\mathrm{c})}$ and $I^{(n_\mathrm{c})}$ by (36)
  if $n_\mathrm{c} < N$ **then**
    $W_f^{(n_\mathrm{f})} \leftarrow W^{(n_\mathrm{c})}$, $I_f^{(n_\mathrm{f})} \leftarrow I^{(n_\mathrm{c})}$, $W_l^{(n_\mathrm{l})} \leftarrow W_{j+1}^2$, $I_l^{(n_\mathrm{l})} \leftarrow I_{j+1}^2$
  else
    break
  end if
end for
$W^{(N)} \leftarrow W^{(n_\mathrm{c})}$, $I^{(N)} \leftarrow I^{(n_\mathrm{c})}$
compute I by (37)
return I

Algorithm 2 Encoder of MSM

Require: N, I
Ensure: x^N
compute $W^{(N)}$ and $I^{(N)}$ by (38)
$W^{(n_\mathrm{c})} \leftarrow W^{(N)}$, $I^{(n_\mathrm{c})} \leftarrow I^{(N)}$
for $j = N/2$ to 2 **do**
  $n_\mathrm{c} \leftarrow 2j$, $n_\mathrm{f} \leftarrow (n_\mathrm{c} - 2)$, $n_\mathrm{l} \leftarrow 2$
  compute $W_f^{(n_\mathrm{f})}$, $I_f^{(n_\mathrm{f})}$, $W_l^{(n_\mathrm{l})}$ and $I_l^{(n_\mathrm{l})}$ by (39)
  $W_j^{(2)} \leftarrow W_l^{(n_\mathrm{l})}$, $I_j^{(2)} \leftarrow I_l^{(n_\mathrm{l})}$
  if $j > 2$ **then**
    $W^{(n_\mathrm{c})} \leftarrow W_f^{(n_\mathrm{f})}$, $I^{(n_\mathrm{c})} \leftarrow I_f^{(n_\mathrm{f})}$
  else
    break
  end if
end for
for $i = 1$ to $N/2$ **do**
  compute $[x_{2i-1}, x_{2i}]$ by (40)
end for
return $x^{(N)} = [x_1, x_2, \ldots, x_{2i-1}, x_{2i}, \ldots, x_{N-1}, x_N]$

I, $(1,1,0,0)$ is listed before $(1,0,1,0)$, as $(0,0)$ has a lower weight than $(1,0)$.

- Second, for two sequences with equal $W_\mathrm{l}^{(n_\mathrm{l})}$, we sort the sequence whose last n_l symbols have a smaller index $I_\mathrm{l}^{(n_\mathrm{l})}$ first. For example, in $(1,0,1,0)$ and $(1,0,0,1)$, $(1,0)$ and $(0,1)$ have the same weight, but $(1,0)$ has a smaller index than $(0,1)$. Hence, $(1,0,1,0)$ is listed before $(1,0,0,1)$ as shown in Table I.
- Third, if two sequences have the same $W_\mathrm{l}^{(n_\mathrm{l})}$ and $I_\mathrm{l}^{(n_\mathrm{l})}$, the sequence whose first n_f symbols $x_\mathrm{f}^{(n_\mathrm{f})}$ have a smaller index $I_\mathrm{f}^{(n_\mathrm{f})}$ is listed first. As seen in Table I, in $(2,0,0,0)$ and $(1,1,0,0)$, $(0,0)$ and $(0,0)$ have the same $W_\mathrm{l}^{(n_\mathrm{l})}$ and $I_\mathrm{l}^{(n_\mathrm{l})}$, but the index of $(2,0)$ is smaller than that of $(1,1)$. Hence, $(2,0,0,0)$ is listed before $(1,1,0,0)$.

Appendix B
MSM Decoder

The decoding process is outlined in Algorithm 1. First, the MSM decoder groups the N symbols into $N/2$ pairs of symbols expressed as $([x_1, x_2], [x_3, x_4], \ldots, [x_{N-1}, x_N])$, where each pair is a 2-symbol sequence. Next, the weight and index of each 2-symbol sequence are calculated. Specifically, for a pair of $[x_{2i-1}, x_{2i}]$, $i = 1, 2, \ldots, N/2$, its weight W_i^2 and index I_i^2 are given by

$$\begin{aligned} W_i^{(2)} &= x_{2i-1} + x_{2i}, \\ I_i^{(2)} &= \begin{cases} x_{2i}, & W_i^{(2)} < M, \\ M - 1 - x_{2i-1}, & W_i^{(2)} \geq M. \end{cases} \end{aligned} \tag{35}$$

Then, the MSM decoder combines the $N/2$ number of 2-symbol sequences recursively as

$$\begin{aligned} W^{(n_\mathrm{c})} &= W_\mathrm{f}^{(n_\mathrm{f})} + W_\mathrm{l}^{(n_\mathrm{l})}, \\ I^{(n_\mathrm{c})} &= \sum_{W'=0}^{W_\mathrm{l}^{(n_\mathrm{l})}-1} g^{(n_\mathrm{l})}(W') g^{(n_\mathrm{f})}\left(W^{(n_\mathrm{c})} - W'\right) \\ &\qquad + I_\mathrm{l}^{(n_\mathrm{l})} g^{(n_\mathrm{f})}\left(W_\mathrm{f}^{(n_\mathrm{f})}\right) + I_\mathrm{f}^{(n_\mathrm{f})}, \end{aligned} \tag{36}$$

where $g^{(n)}(W')$ represents the number of n-symbol sequences having the same weight W', and $n_\mathrm{l} = 2$. The value of $g^{(n)}(W')$ is the coefficient of $a^{W'}$ in the generating functions $G^{(n)}(a) = (1 + a + \cdots + a^{M-1})^n$. This procedure continues until an N-dimensional problem is generated, which is characterized by $W^{(N)}$ and $I^{(N)}$. Finally, the index I of K information bits can be calculated by

$$I = I^{(N)} + z^{(N)}(W^{(N)}), \tag{37}$$

where $z^{(N)}(W^{(N)}) = \sum_{i=0}^{W^{(N)}-1} g^{(N)}(i)$ is the number of N-symbol sequences with weights less than $W^{(N)}$.

Appendix C
MSM Encoder

The MSM encoder is summarized in Algorithm 2. First, based on I, $W^{(N)}$ and $I^{(N)}$ of the sequence $x^{(N)}$ are calculated by

$$\begin{aligned} W^{(N)} &= \arg\max_{W'} \{z^{(N)}(W') \leq I\}, \\ I^{(N)} &= I - z^{(N)}(W^{(N)}). \end{aligned} \tag{38}$$

This N-dimensional problem is split into two lower-

dimensional ones as

$$
\begin{aligned}
W_{\mathrm{l}}^{(n_{\mathrm{l}})} &= \arg\max_{W} \left\{ \sum_{W'=0}^{W-1} g^{(n_{\mathrm{l}})}(W') g^{(n_{\mathrm{f}})}\left(W^{(n_{\mathrm{c}})} - W'\right) \right. \\
&\qquad \left. \leq I^{(n_{\mathrm{c}})} \right\}, \\
W_{\mathrm{f}}^{(n_{\mathrm{f}})} &= W^{(n_{\mathrm{c}})} - W_{\mathrm{l}}^{(n_{\mathrm{l}})}, \\
I_{\mathrm{l}}^{(n_{\mathrm{l}})} &= \left(I^{(n_{\mathrm{c}})} - \sum_{W'=0}^{W_{\mathrm{l}}^{(n_{\mathrm{l}})}-1} g^{(n_{\mathrm{l}})}(W') g^{(n_{\mathrm{f}})}\left(W^{(n_{\mathrm{c}})} - W'\right) \right. \\
&\qquad \left. - I_{\mathrm{f}}^{(n_{\mathrm{f}})} \right) / g^{(n_{\mathrm{f}})}\left(W_{\mathrm{f}}^{(n_{\mathrm{f}})}\right), \\
I_{\mathrm{f}}^{(n_{\mathrm{f}})} &= \left(I^{(n_{\mathrm{c}})} - \sum_{W'=0}^{W_{\mathrm{l}}^{(n_{\mathrm{l}})}-1} g^{(n_{\mathrm{l}})}(W') g^{(n_{\mathrm{f}})}\left(W^{(n_{\mathrm{c}})} W'\right) \right) \\
&\qquad \%_0 g^{(n_{\mathrm{f}})}\left(W_{\mathrm{f}}^{(n_{\mathrm{f}})}\right),
\end{aligned} \tag{39}
$$

where $n_{\mathrm{l}} = 2$. This procedure is repeated until $N/2$ number of two-dimensional problems are obtained. Finally, for each two-dimensional problem, a 2-symbol sequence can be calculated using

$$
\begin{cases}
W_i^{(2)} < M : x_{2i} = I_i^2, \;\; x_{2i-1} = W_i^{(2)} - x_{2i}, \\
W_i^{(2)} \geq M : x_{2i-1} = M - 1 - I_i^{(2)}, \;\; x_{2i} = W_i^2 - x_{2i-1},
\end{cases} \tag{40}
$$

based on which the N-symbol sequence $x^{(N)}$ is composed.

Note that the proposed grouping strategy of $x^{(n_{\mathrm{c}})}$ in the encoder and decoder of MSM can be extended to the case that n_{l} is an even positive integer larger than two. Take the MSM decoder with $n_{\mathrm{l}} = 4$ as an example. We assume that the input of the MSM decoder is a 6-symbol sequence. First, the MSM decoder groups these six symbols into three pairs of symbols denoted by $([x_1, x_2], [x_3, x_4], [x_5, x_6])$, where each pair is a 2-symbol sequence, and calculates their weights and indices based on (35). Then, the decoder combines the last two pairs into a 4-symbol sequence and calculates its index and weight based on (36). Furthermore, the first pair of 2-symbol sequence and the obtained 4-symbol sequence are combined into a 6-symbol sequence with $n_{\mathrm{l}} = 4$, whose weight and index are calculated based on (36). Finally, we calculate the index that represents the information bits based on (37). Moreover, while the proposed algorithm cannot be directly applied to the grouping strategy of $n_{\mathrm{l}} = 1$, the proposed grouping strategy of $n_{\mathrm{l}} = 2$ can reduce the hardware complexity, since this strategy requires fewer generating functions compared to $n_{\mathrm{l}} = 1$.

REFERENCES

[1] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, 3rd Quart. 2016.

[2] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication—A tutorial review," *Proc. IEEE*, vol. 107, no. 7, pp. 1256–1301, July 2019.

[3] C. Rose, I. S. Mian, and M. Ozmen, "Capacity bounds on point-to-point communication using molecules," *Proc. IEEE*, vol. 107, no. 7, pp. 1342–1355, July 2019.

[4] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The Internet of bio-nano things," *IEEE Commun. Mag.,* vol. 53, no. 3, pp. 32–40, Mar. 2015.

[5] C. Lee, B.-H. Koo, C.-B. Chae, and R. Schober, "The Internet of bio-nano things in blood vessels: System design and prototypes," *IEEE/KICS J. Commun. Netw.,* vol. 25, no. 2, pp. 1–10, Apr. 2023.

[6] P.-C. Yeh *et al*., "A new frontier of wireless communication theory: Diffusion-based molecular communications," *IEEE Wireless Commun.,* vol. 19, no. 5, pp. 28–35, Oct. 2012.

[7] S. Hiyama, Y. Moritani, R. Gojo, S. Takeuchi, and K. Sutoh, "Biomolecular-motor-based autonomous delivery of lipid vesicles as nano- or microscale reactors on a chip," *Lab Chip*, vol. 10, no. 20, pp. 2741–2748, 2010.

[8] O. F. Sezgen *et al*., "A multiscale communications system based on engineered bacteria," *IEEE Commun. Mag.*, vol. 59, no. 5, pp. 62–67, May 2021.

[9] M. S. Kuran, H. B. Yilmaz, I. Demirkol, N. Farsad, and A. Goldsmith, "A survey on modulation techniques in molecular communication via diffusion," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 7–28, 1st Quart. 2021.

[10] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *Proc. IEEE Int. Conf. Commun. (ICC),* Kyoto, Japan, June 2011, pp. 1–5.

[11] L. Shi and L.-L. Yang, "Error performance analysis of diffusive molecular communication systems with on-off keying modulation," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 3, no. 4, pp. 224–238, Dec. 2017.

[12] N. Kim and C.-B. Chae, "Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 847–856, Dec. 2013.

[13] X. Chen, Y. Huang, L.-L. Yang, and M. Wen, "Generalized molecular shift keying (GMoSK): Principles and performance analysis," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 6, no. 3, pp. 168–183, Dec. 2020.

[14] B. C. Akdeniz, A. E. Pusane, and T. Tugcu, "Position-based modulation in molecular communications," *Nano Commun. Netw.*, vol. 16, pp. 60–68, June 2018.

[15] Y. Huang, M. Wen, L.-L. Yang, C.-B. Chae, and F. Ji, "Spatial modulation for molecular communication," *IEEE Trans. Nanobiosci.*, vol. 18, no. 3, pp. 381–395, July 2019.

[16] M. C. Gursoy, E. Basar, A. E. Pusane, and T. Tugcu, "Index modulation for molecular communication via diffusion systems," *IEEE Trans. Commun.,* vol. 67, no. 5, pp. 3337–3350, May 2019.

[17] B.-H. Koo, C. Lee, A. Pusane, T. Tugcu, and C.-B. Chae, "MIMO operations in molecular communications: Theory, prototypes, and open challenges," *IEEE Commun. Mag.*, vol. 59, no. 9, pp. 98–104, Sept. 2021

[18] Y. Tang, M. Wen, X. Chen, Y. Huang, and L.-L. Yang, "Molecular type permutation shift keying for molecular communication," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 6, no. 2, pp. 160–164, Nov. 2020.

[19] Y. Tang, Y. Huang, C.-B. Chae, W. Duan, M. Wen, and L.-L. Yang, "Molecular type permutation shift keying in molecular MIMO communications for IoBNT," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 16023–16034, Nov. 2021.

[20] Y. Tang, Y. Huang, M. Wen, L.-L. Yang, and C.-B. Chae, "A molecular spatio-temporal modulation scheme for MIMO communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–6.

[21] Y. C. Gultekin, W. J. van Houtum, A. G. C. Koppelaar, and F. M. J. Willems, "Enumerative sphere shaping for wireless communications with short packets," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1098–1112, Feb. 2020.

[22] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[23] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.

[24] X. Hong, C. Fei, G. Zhang, J. Du, and S. He, "Discrete multitone transmission for underwater optical wireless communication system using probabilistic constellation shaping to approach channel capacity limit," *Opt. Lett.*, vol. 44, no. 3, pp. 558–561, 2019.

[25] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, July 1948.

[26] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 2, no. 2, pp. 120–142, Dec. 2016.

[27] G. Aminian, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Capacity of diffusion-based molecular communication networks over LTI-Poisson channels," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 2, pp. 188–201, June 2015.

[28] P. Schulte and G. Bocherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[29] G. R. Lang and F. M. Longstaff, "A leech lattice modem," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 968–973, Aug. 1989.

[30] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1044–1056, July 1994.

[31] M. U. Mahfuz, D. Makrakis, and H.T. Mouftah, "Performance analysis of convolutional coding techniques in diffusion-based concentration-encoded PAM molecular communication systems," *BioNanoScience*, vol. 3, no. 3, pp. 270–284, May 2013.

[32] A. Noel, K. C. Cheung, and R. Schober, "Using dimensional analysis to assess scalability and accuracy in molecular communication," in *Proc. IEEE Int. Conf. Commun. (ICC),* Budapest, Hungary, June 2013, pp. 818–823.

[33] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2532–2547, June 2011.

[34] M. S. Leeson and M. D. Higgins, "Forward error correction for molecular communications," *Nano Commun. Netw.*, vol. 3, no. 3, pp. 161–167, Sept. 2012.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[36] R. F. H. Fischer, "Calculation of shell frequency distributions obtained with shell-mapping schemes," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1631–1639, July 1999.

[37] Y. Tang, F. Ji, M. Wen, Q. Wang, and L.-L. Yang, "Enhanced molecular type permutation shift keying for molecular communication," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2722–2726, Dec. 2021.

[38] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission.* New York, NY, USA: Wiley, 2002.