



HAL
open science

K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

Xuewen Qian, Marco Di Renzo, Andrew Eckford

► **To cite this version:**

Xuewen Qian, Marco Di Renzo, Andrew Eckford. K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications. *IEEE Transactions on Communications*, 2021, 69 (8), pp.5456 - 5470. 10.1109/TCOMM.2021.3075523 . hal-03358065

HAL Id: hal-03358065

<https://hal.science/hal-03358065>

Submitted on 29 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

Xuwen Qian, Marco Di Renzo, *Fellow, IEEE*, and Andrew Eckford, *Senior Member, IEEE*

Abstract

In this paper, we consider non-coherent detection for molecular communication systems in the presence of inter-symbol-interference. In particular, we study non-coherent detectors based on memory-bits-based thresholds in order to achieve low bit-error-ratio (BER) transmission. The main challenge of realizing detectors based on memory-bits-based thresholds is to obtain the channel state information based only on the received signals. We tackle this issue by reformulating the thresholds through intermediate variables, which can be obtained by clustering multi-dimensional data from the received signals, and by using the K-means clustering algorithm. In addition to estimating the thresholds, we show that the transmitted bits can be retrieved from the clustered data. To reduce clustering errors, we propose iterative clustering methods from one-dimensional to multi-dimensional data, which are shown to reduce the BER. Simulation results are presented to verify the effectiveness of the proposed methods.

Index Terms

Molecular communications, inter-symbol interference, channel state information, non-coherent detection, K-means clustering.

I. INTRODUCTION

Recent developments in biology and nano-technology have gained attention in micro-scale communications in light of emerging potential applications, e.g., bio-robots [1]. In this context, molecular communication (MC) is regarded as a promising solution [2] because of the difficulty

Manuscript received Aug. 30, 2020; revised Jan. 19, 2021; and March 12, 2021. X. Qian and M. Di Renzo are with Université Paris-Saclay, CNRS and CentraleSupélec, Laboratoire des Signaux et Systèmes, 91192 Gif-sur-Yvette, France. (e-mail: marco.di-renzo@universite-paris-saclay.fr). A. Eckford is with the Dept. of Electrical Engineering and Computer Science, York University, 4700 Keele Street, Toronto, Canada M3J 1P3.

of using traditional electromagnetic-based techniques [3]. In a MC system, the information is disseminated via small particles [2]. Among the different modulation schemes that can be used in MC systems [4], binary concentration shift keying (CSK) modulation is the simplest method for encoding information onto the number of released particles. In addition, diffusion via Brownian motion [5] is the most common solution for allowing information particles propagate from a transmitter to a receiver.

A. Motivation

In diffusion-based MC systems, there exist several issues to be solved. A major issue is the non-negligible inter-symbol interference (ISI) that is caused by the intrinsic characteristics of channels with memory. If not considered appropriately, the ISI can severely degrade the bit-error-ratio (BER) performance [6], [7], [8]. Even though the use of enzymes [9] appropriately injected in the propagation environment can help mitigate the ISI, this approach cannot completely eliminate the ISI. Motivated by these considerations, we consider an MC system in the presence of ISI. Solutions to mitigate the ISI exist and include methods based on the design of the modulation [10], the code [11], and the detector [12]. In this paper, we focus our attention on the design of detectors that use binary CSK modulation.

In MC systems, different types of detectors exist, including 1) no memory threshold-based schemes [9], [13]; 2) optimal constant-threshold-based schemes [14]; 3) multi-memory-bit threshold-based schemes [15]; 4) adaptive decision-feedback equalizers [12]; and 5) multi-sub-slot threshold-based schemes [16]. Among these approaches, multi-memory-bit threshold-based schemes constitute a promising solution because they can achieve low BER performance at a low computational complexity. However, perfect CSI needs to be known at the receiver in order to calculate the optimal thresholds. In this regard, channel estimation (CE) [6], [17] methods based on predefined preambles are often utilized in order to ensure the reliable detection of data. In realistic scenarios, however, some parameters of MC communication systems may not remain constant during the CE and data detection phases, thus leading to variations in the channel response and, hence, to CE inaccuracies. In some cases, e.g., for multiple absorbing receiver systems [18], even though some theoretical channel models exist, which include MC systems with an eavesdropper [19], MC systems with two receivers [20], [21] and MC systems with randomly distributed receivers [22], it may not be possible to model the channel analytically if realistic factors are considered. To counteract these issues, some authors have used machine learning (ML) methods to design

the receiver [23] and [14]. However, ML-based receivers must be trained with long sequences of symbols. The performance of ML-based receivers, in addition, highly depends on the quality of the data used for training [24].

B. Related Works

Non-coherent detection methods for MC systems are discussed in [8], where the authors obtain closed-form expressions of the optimal detection metric for a multi-symbol detector that uses an approximated probability mass function for the Poisson channel. In addition, the authors propose a blind single-symbol detector based on a constant threshold. However, the authors assume no ISI in the system model which may hold true when the symbol length is long enough and the symbol rate is low. In [25], the authors consider the impact of ISI and exploit the local convexity of the received signals in order to detect the symbols from the difference of molecular concentration. To capitalize on the local convexity of the received signals, the transmitter needs, however, to release a sufficiently large number of molecules. In [26], a non-linear detector with an adaptive threshold is analyzed. The approach is based on the quick-rising and slow-decaying trends of the received signals after passing through a filter. In [27], the authors exploit the unsupervised fuzzy C-mean approach to detect symbols from the quick-rising and slow-decaying of the processed signals. In [28], the authors propose an approach based on the energy difference of the received signals. This method requires large numbers of molecules in order to exploit the energy difference. Non-coherent detection is investigated in Poisson channels for application to optical wireless communications in [29]. However, current non-coherent detection schemes assume negligible ISI. In summary, the aforementioned approaches either do not consider ISI, or require large numbers of molecules and, in general, cannot achieve low BER performance.

C. Contributions

In this paper, motivated by the results in [8] and [15], we devise multi-memory-bit threshold-based schemes without prior CSI for the reliable transmission of data in the presence of ISI. The specific novelty and contributions made by this paper can be summarized as follows.

- We reformulate the multi-memory-bit threshold-based detectors via intermediate variables, i.e., the average number of received particles as a function of the memory bits. Unlike threshold-based methods that use the prior knowledge of the CSI, the considered intermediate variables can be obtained by applying clustering methods to the received signals.

- In order to reduce the clustering error, we construct multi-dimensional data from the received signals and designate the initial centroids of the clusters using the largest received signal. To the best of the authors' knowledge, it is the first time in the communication literature that the main features are extracted by increasing the dimension for clustering data. Conventional methods, on the other hand, are based on decreasing the dimension for clustering data [30]. In addition, we can infer the transmitted symbols from the clustered data. We design and analyze two approaches, which are referred to as direct clustering-based inference and clustering-plus-threshold detection.
- We devise an iterative method to further improve the BER performance and to avoid clustering failures in high data rate systems, i.e., in the presence of non-negligible ISI. The essence of the proposed method is to iteratively apply clustering from one-dimensional data to high-dimensional data. In each iteration, we construct the initial centroids based on the estimated centroids in the previous iteration. By using the proposed iterative method, we show that the clustering errors can be reduced.

D. Paper Organization

The remainder of this paper is organized as follows. In Section II, the system model is introduced. In Section III, the multi-memory-bit threshold is reformulated via intermediate variables. In Section IV, the proposed non-coherent detection method based on multi-dimensional clustering is introduced. In Section V, an iterative method for estimating the initial centroids for application to multi-dimensional clustering is introduced. In Section VI, numerical results and the computational complexity of the proposed schemes are discussed. Finally, Section VII concludes this paper.

II. SYSTEM MODEL

We consider a three-dimensional unbounded MC system without flow that consists of a point transmitter and a spherical absorbing receiver, as depicted in Fig. 1. We assume that each information particle diffuses randomly and independently through the medium. The particles are assumed not to degrade rapidly, thus resulting in ISI. We assume that the temperature is constant and the viscosity remains unchanged during the whole transmission. Thus, the diffusion coefficient D [3] is assumed to be constant.

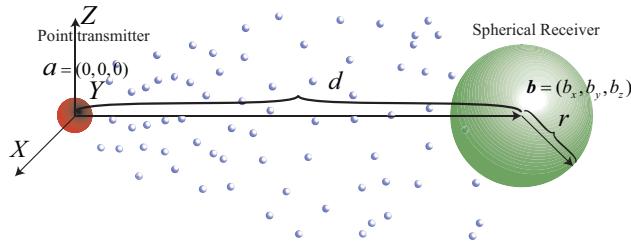


Fig. 1: Three-dimensional MC system consisting of a point transmitter and a spherical absorbing receiver

Assuming that the transmitter is located in $\mathbf{a} = (0, 0, 0)$ and the receiver is located in $\mathbf{b} = (b_x, b_y, b_z)$, the hitting rate is expressed as follows [13], [31]:

$$f_{hit}(t) = \frac{r(d-r)}{d\sqrt{4\pi Dt^3}} e^{-\frac{(d-r)^2}{4Dt}} \quad (1)$$

where $\|\mathbf{a} - \mathbf{b}\| = d$ is the distance between the center of the transmitter and the center of the receiver, and r denotes the radius of the receiver. A binary CSK modulation scheme is assumed¹, and the bit transmitted in the i th time slot is denoted by s_i . During the i th slot, the transmitter releases N_{TX} information particles if $s_i = 1$, otherwise the transmitter does not release any particles. We assume that the N_{TX} information particles are released in a very short time such that we can ignore the release time effect on the received signal. The hitting probability of an absorbing receiver in the $(i-1)$ th time slot of duration T is:

$$P_{i-1} = \int_{(i-1)T}^{iT} f_{hit}(t) dt = \frac{r}{d} \left\{ \operatorname{erfc}\left(\frac{d-r}{\sqrt{4DiT}}\right) - \operatorname{erfc}\left(\frac{d-r}{\sqrt{4D(i-1)T}}\right) \right\} \quad (2)$$

where $\operatorname{erf}(y) = \int_0^y \frac{2}{\sqrt{\pi}} e^{-x^2} dx$ and $\operatorname{erfc}(y) = 1 - \operatorname{erf}(y)$.

Let $C_j = N_{TX} P_j$ denote the average number of received particles in the j th time-slot after the release of N_{TX} particles. For clarity, C_0 denotes the average number of received particles in the symbol of interest. Thus, the number of received particles [16] in the i th time-slot follows a Poisson distribution²:

$$r_i \sim \text{Poisson}(I_i + s_i C_0) \quad (3)$$

where $I_i = \bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j$ is the sum of the ISI and background noise, $\bar{\lambda}_0$ is the background

¹The generalization to multilevel CSK modulation is discussed in Section VI-C.

²The number of received particles r_i can be modelled accurately by using a Binomial distribution. For ease of analysis, r_i is approximated with a Poisson distribution. This is a good approximation if N_{TX} is large [8].

noise power per unit time, and L denotes the length (memory) of the Poisson channel.

We define the signal-noise-ratio (SNR) as follows:

$$\text{SNR} = 10 \log_{10} \frac{C_0}{2\lambda_0 T} \quad (4)$$

since the information bits are assumed to be equiprobable. Thus,

$$N_{TX} = \frac{2\bar{\lambda}_0 T 10^{\frac{\text{SNR}}{10}}}{P_0} \quad (5)$$

Therefore, the probability of receiving r_i information particles is:

$$\Pr(r_i | I_i + s_i C_0) = \frac{e^{-(I_i + s_i C_0)} (I_i + s_i C_0)^{r_i}}{r_i!} \quad (6)$$

III. MULTI-MEMORY-BIT THRESHOLD REFORMULATION

To detect the symbols, existing methods are based on calculating thresholds based on previously detected bits, which are referred to as memory bits. The main idea is to exploit previously detected bits, even erroneously estimated, in order to detect new transmitted bits based on the number of received particles. If the number of particles is below the threshold, a binary zero is estimated. Otherwise, a binary one is estimated. We first discuss the case of knowing the exact channel length L , and then we consider the case of limited channel information. The analytical threshold based on the knowledge of L memory bits can be obtained from the following equation:

$$\Pr(r_i | s_i = 0, s_{i-j}, 1 \leq j \leq L) = \Pr(r_i | s_i = 1, s_{i-j}, 1 \leq j \leq L) \quad (7)$$

where $\Pr(r_i | s_i = 0, s_{i-j}, 1 \leq j \leq L)$ and $\Pr(r_i | s_i = 1, s_{i-j}, 1 \leq j \leq L)$ denote the probabilities of receiving r_i information particles in the i th time slot conditioned upon the previously transmitted symbols s_{i-j} for $1 \leq j \leq L$ and the current symbol $s_i = 0$ or $s_i = 1$, respectively. These probabilities correspond to (6).

By solving (7) based on (6), the optimal threshold $\tau |_{s_{i-j}, 1 \leq j \leq L}$ conditioned on the previously transmitted symbols s_{i-j} for $1 \leq j \leq L$ is as follows:

$$\tau |_{s_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{I_i})} = \frac{C_0}{\ln(1 + \frac{C_0}{\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j})} \quad (8)$$

Since the symbols s_{i-j} for $0 \leq j \leq L$ are unknown, the previously estimated symbols \hat{s}_{i-j} for $0 \leq j \leq L$ are used to compute the threshold in (8), i.e., $\tau |_{\hat{s}_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{\lambda_0 T + \sum_{j=1}^L \hat{s}_{i-j} C_j})}$.

Based on the computed threshold, \hat{s}_i is demodulated as follows:

$$\hat{s}_i = \begin{cases} 1 & r_i > \tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} \\ 0 & r_i \leq \tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} \end{cases} \quad (9)$$

Besides necessitating the bits detected in the previous time slots, the computation of the threshold in (8) relies upon the prior knowledge of the CSI, i.e., the coefficients C_j are assumed to be known. To avoid using prior information about the channel, i.e., C_j for $0 \leq j \leq L$, when calculating the threshold in (8), we resort to intermediate variables, i.e., the average number of received particles $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$ that is defined as follows:

$$\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L} = I_i + s_i C_0 = \bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j \quad (10)$$

Equation (10) corresponds to the theoretical average number of particles, and, therefore, it depends on the variables C_j . In the following, we first show how to formulate the thresholds in (8) by exploiting the intermediate variables in (10), instead of using C_j explicitly. In this step the CSI is still required. Then, we elaborate how to estimate the intermediate variables without necessitating CSI.

A. Computing the Thresholds Using $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$

In this subsection, we show how the thresholds can be formulated only in terms of average number of particles $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$. In particular, in order to formulate the problem analytically and understand the rationale of the proposed approach, we consider, just for this section, that $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$ is given by its analytical expression in (10).

The threshold in (8) can be rewritten, in an equivalent form, only in terms of $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$, as follows:

$$\begin{aligned} \tau|_{s_{i-j}, 1 \leq j \leq L} &= \frac{C_0}{\ln\left(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j}\right)} = \frac{C_0 + \bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j - (\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j)}{\ln\left(\frac{\bar{\lambda}_0 T + C_0 + \sum_{j=1}^L s_{i-j} C_j}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j}\right)} \\ &= \frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq L} - \bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq L}}{\ln\left(\frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq L}}{\bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq L}}\right)} \end{aligned} \quad (11)$$

In practice, the memory length used by the receiver may be limited and may be smaller than L . By assuming that $\mathcal{L} \leq L$ memory bits are used, the threshold can be obtained by solving the

equation $\Pr(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$ where:

$$\Pr(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L}) = \frac{1}{2^{L-\mathcal{L}}} \sum_{s_{i-j}, \mathcal{L}+1 \leq j \leq L} \Pr(r_i|s_{i-j}, 0 \leq j \leq L) \quad (12)$$

Due to the analytical complexity of (12) it is, however, difficult to compute an analytical expression for the detection threshold by imposing $\Pr(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$. To overcome this analytical complexity, we use an approximation for $\Pr(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$. In particular, we approximate (12) with $\Pr_{\text{approx}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$ as follows:

$$\Pr_{\text{approx}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L}) = \frac{e^{-\bar{r}_i|s_{i-j}, 0 \leq j \leq \mathcal{L}} (\bar{r}_i|s_{i-j}, 0 \leq j \leq \mathcal{L})^{r_i}}{r_i!} \quad (13)$$

which is obtained by averaging $I_i = \bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j$ in (6) with respect to s_{i-j} for $\mathcal{L}+1 \leq j \leq L$, which yields:

$$\bar{r}_i|s_{i-j}, 0 \leq j \leq \mathcal{L} = \sum_{j=0}^{\mathcal{L}} s_{i-j} C_j + \sum_{j=\mathcal{L}+1}^L C_j/2 + \bar{\lambda}_0 T \quad (14)$$

By using $\Pr_{\text{approx}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$, the detection threshold can be computed by solving the equation $\Pr_{\text{approx}}(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr_{\text{approx}}(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$, which yields:

$$\tau|_{s_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}} - \bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (15)$$

As anticipated, the threshold in (15) is a theoretical formulation for the time being. In practice, it needs to be obtained (estimated) only from the received data without any prior information. To this end, we replace $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ with its estimate $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$, which is introduced and defined in the next sub-section. Therefore, the detection threshold is expressed as follows:

$$\hat{\tau}|_{s_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\hat{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}} - \hat{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\hat{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\hat{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (16)$$

Since s_{i-j} for $0 \leq j \leq \mathcal{L}$ are unknown, the estimates \hat{s}_{i-j} are used instead, which results in the following formulation for the threshold:

$$\hat{\tau}|_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\hat{r}_i|_{\hat{s}_i=1, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} - \hat{r}_i|_{\hat{s}_i=0, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\hat{r}_i|_{\hat{s}_i=1, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}{\hat{r}_i|_{\hat{s}_i=0, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (17)$$

Based on this reformulation, each symbol is demodulated as follows:

$$\hat{s}_i = \begin{cases} 0, & r_i \leq \hat{\tau} |_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} \\ 1, & r_i > \hat{\tau} |_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} \end{cases} \quad (18)$$

B. Estimation of $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the Received Data

As anticipated, our objective is to develop non-coherent detection schemes that do not need CSI to operate. For example, we cannot rely on the knowledge of C_j for $j = 0, \dots, \mathcal{L}$ to compute the detection thresholds. In this sub-section, we show an example of how $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ can be obtained by only using the received data, which implies that the threshold in (17) can be obtained without prior CSI information.

As a case study, we show how the intermediate variables $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ are obtained for $\mathcal{L} = 1$. From (12), the exact probability mass function of receiving r_i particles given s_i and s_{i-1} is:

$$\Pr(r_i | s_i, s_{i-1}) = \frac{1}{2^{L-1}} \sum_{s_{i-j}, 2 \leq j \leq L} \frac{e^{-(\bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j)} (\bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j)^{r_i}}{r_i!} \quad (19)$$

Theoretically, $\bar{r}_i |_{s_i, s_{i-1}}$ is defined and can be computed as follows:

$$\bar{r}_i |_{s_i, s_{i-1}} = E[r_i | s_i, s_{i-1}] = \sum_{r_i=0}^{\infty} r_i \Pr(r_i | s_i, s_{i-1}) \quad (20)$$

Let us consider a sequence of received particles $\{r_1, \dots, r_n, \dots, r_K\}$ of length K . The theoretical average number of received particles in (20) can be estimated empirically from the sequence of K observations. In practice (empirically), the probability $\Pr(r_i | s_i, s_{i-1})$ can be interpreted as a ratio [32, Section 19.2]:

$$\Pr(r_i | s_i, s_{i-1}) = \lim_{N_{s_i, s_{i-1}} \rightarrow \infty} \frac{N_{r_i, s_i, s_{i-1}}}{N_{s_i, s_{i-1}}} \quad (21)$$

where $N_{r_i, s_i, s_{i-1}}$ denotes the number of elements within the sequence of K observations that are equal to r_i and for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$, and $N_{s_i, s_{i-1}} = \sum_{r_i=0}^{\infty} N_{r_i, s_i, s_{i-1}}$ denotes the number of elements within the sequence of K observations for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$ (but the number of particles is not necessarily equal to r_i). Therefore, (20) can be rewritten as follows:

$$\bar{r}_i |_{s_i, s_{i-1}} = \lim_{N_{s_i, s_{i-1}} \rightarrow \infty} \frac{\sum_{r_i=0}^{\infty} r_i N_{r_i, s_i, s_{i-1}}}{N_{s_i, s_{i-1}}} = \lim_{K \rightarrow \infty} \frac{\sum_{n=0}^K r_n \kappa_{n, s_i, s_{i-1}}}{\sum_{n=0}^K \kappa_{n, s_i, s_{i-1}}} \quad (22)$$

where $\kappa_{n,s_i,s_{i-1}} = 1$ if $[s_n, s_{n-1}] = [s_i, s_{i-1}]$ and $\kappa_{n,s_i,s_{i-1}} = 0$ otherwise. The variable $\kappa_{n,s_i,s_{i-1}}$ can be regarded as an indicator variable that provides information on whether the observation r_n belongs to a group of observations for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$.

In the literature [33], a group of observations is referred to as a cluster, a single observation of the cluster is referred to as a point of the cluster, and the condition that defines the cluster, i.e., the condition $[s_n, s_{n-1}] = [s_i, s_{i-1}]$, is referred to as the label of the cluster. In addition, the arithmetic mean of all the points in a subset X of \mathbb{R}^l is referred to as the centroid of the cluster [34] and is defined, in general terms, as follows:

$$\boldsymbol{\mu} = \frac{\int \boldsymbol{x}g(\boldsymbol{x})d\boldsymbol{x}}{\int g(\boldsymbol{x})d\boldsymbol{x}} \quad (23)$$

where \boldsymbol{x} is a vector in \mathbb{R}^l , the integrals are computed over the whole space \mathbb{R}^l , and $g(\boldsymbol{x})$ is a characteristic function that can be defined for different purposes. Based on (22) and (23), $\bar{r}_i|_{s_i,s_{i-1}}$ can be viewed as the centroid of the cluster of observations whose label is $[s_i, s_{i-1}]$. Thus $\bar{r}_i|_{s_i,s_{i-1}}$ is obtained by computing the centroid of the cluster with label $[s_i, s_{i-1}]$ that is obtained from the set of received signals (the observations) $\{r_1, \dots, r_n, \dots, r_K\}$.

Based on (22), we evince that the thresholds of interest can be computed by calculating the centroids in (23). However, (22) necessitates the knowledge of the symbols s_n , which are usually unknown. Our objective is to estimate $\bar{r}_i|_{s_i,s_{i-1}}$ from the received signals without knowing the symbols s_n . Estimating the centroids in (22) or (23) without any prior information can be obtained by applying clustering methods, which partition the received signals into several groups, each one corresponding to a given label $[s_i, s_{i-1}]$, and then obtaining the corresponding centroids.

In summary, the thresholds for data detection can be obtained by applying clustering methods directly to the received signals without any prior information on the CSI and without knowledge of the transmitted symbols. In particular, the thresholds can be retrieved from the centroids of the clusters. In the next section, we detail how to use clustering for non-coherent detection in MC systems, and, in particular, we introduce our approach based on multi-dimensional clustering.

IV. CLUSTERING-BASED NON-COHERENT DETECTION

With the aid of the reformulated thresholds in Section III, we show that data detection can be realized without prior CSI. In particular, we introduce non-coherent detection methods based on the K-means clustering algorithm. To this end, we first introduce some background information on clustering in general and the K-means algorithm in particular. To elucidate the operating

principle of the proposed methods, we report some illustrations that are obtained by using the simulation setup in Table I (see Section VI) in the presence of mild ISI ($T = 30\Delta T$) and severe ISI ($T = 20\Delta T$). It is worth mentioning that the algorithms proposed in this paper require the storage of symbols in order to be executed, similar to [16], [23], [27], [35]. At the time of writing, this may be difficult to be realized with biological circuits. The actual implementation of the proposed solutions by using biological circuits is an important issue that needs to be assessed in future research.

A. K-Means Clustering Algorithm

Assume that we have a data set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ of N observations and each element is a \mathbb{D} -dimensional vector \mathbf{x}_n . The objective is to partition the data set into N_c clusters whose centroids are \mathbb{D} -dimensional vectors denoted by $\boldsymbol{\mu}_k$. The centroid $\boldsymbol{\mu}_k$ is the mean of its clustered points. In clustering, $\kappa_{n,k}$ is a variable that represents if the distance between the n th observation and the k th centroid is smaller than the distance between that observation and the other centroids. The K-means algorithm is a clustering method that works iteratively in order to compute or estimate the centroids $\boldsymbol{\mu}_k$ and the indicator variables $\kappa_{n,k}$ from a set of observations. In particular, the K-means algorithm encompasses two steps that iteratively compute $\kappa_{n,k}$ and $\boldsymbol{\mu}_k$ at each step. The K-means algorithm needs an initial estimate of the centroids to operate. The initial centroids $\boldsymbol{\mu}_k$ can be either randomly selected from the data set or other methods can be employed. The selection of the initial centroids is discussed in further text. The K-means clustering algorithm can be summarized as follows.

Step I: Assign \mathbf{x}_n to the closest cluster (initial, if this is the first iteration) centroid:

$$\kappa_{n,k} = \begin{cases} 1, & \text{if } k = \arg \min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

Step II: Update the cluster centroid:

$$\boldsymbol{\mu}_k = \frac{\sum_n \kappa_{n,k} \mathbf{x}_n}{\sum_n \kappa_{n,k}} \quad (25)$$

where $\sum_n \kappa_{n,k}$ is the number of points in the k th cluster.

The two steps are repeated until $\kappa_{n,k}$ does not change. In the next sub-section, we show how to apply this algorithm for non-coherent detection in MC systems.

B. Single-Dimensional Clustering: Challenges and Limitations

In (16), the empirical average number of received particles $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ need to be as close as possible to the theoretical average number of received particles $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ in order to obtain a low BER. In order to elucidate the complexity of the problem at hand, let us consider the example in (19) by assuming $\mathcal{L} = 1$.

The objective is to estimate $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the set of received observations, $\{r_1, \dots, r_n, \dots, r_K\}$, that are obtained from (3) given the transmitted symbols $\{s_1, \dots, s_n, \dots, s_K\}$. Since $\mathcal{L} = 1$, we are interested in identifying four clusters in the received signals, which correspond to the labels $[s_i, s_{i-1}] = [0, 0]$, $[s_i, s_{i-1}] = [0, 1]$, $[s_i, s_{i-1}] = [1, 0]$, and $[s_i, s_{i-1}] = [1, 1]$. From (20) or equivalently (14), we obtain:

$$\bar{r}_i|_{s_i, s_{i-1}} = s_i C_0 + s_{i-1} C_1 + \sum_{j=2}^L C_j / 2 + \bar{\lambda}_0 T \quad (26)$$

The average number of received particles in (26) is the theoretical one that can be obtained from analysis. In order to understand the difficulty of obtaining (26) from the empirical data, we illustrate an example in Fig. 2. Based on a large set of empirical data ($K = 2^{16}$ samples), we calculate the empirical probability mass function (PMF) of the number of particles, by assuming that the labels $[s_i, s_{i-1}]$ are known. How the labels can be estimated from the data is discussed in further text. From the empirical data, we apply the K-means clustering algorithm in (24) and (25) in order to estimate the centroids. The estimated average number of received particles $\hat{r}_i|_{s_i, s_{i-1}}$ is set equal to the estimated centroids. From Fig. 2, we observe that there is a non-negligible difference between the theoretical values obtained from (26) and the empirical values estimated by using the K-means clustering algorithm, even if the labels are assumed to be known, i.e., the association between the centroids and the labels is error-free. The differences between $\hat{r}_i|_{s_i, s_{i-1}}$ and $\bar{r}_i|_{s_i, s_{i-1}}$ mainly originate from the overlapping areas of the conditional probabilities $\Pr(r_i|s_i, s_{i-1})$.

The simple example illustrated in Fig. 2 highlights the challenges of applying clustering for estimating the average number of received particles in MC systems and, therefore, the detection thresholds. In the following sub-section, we introduce our proposed approach for improving the estimation accuracy based on multi-dimensional clustering methods.

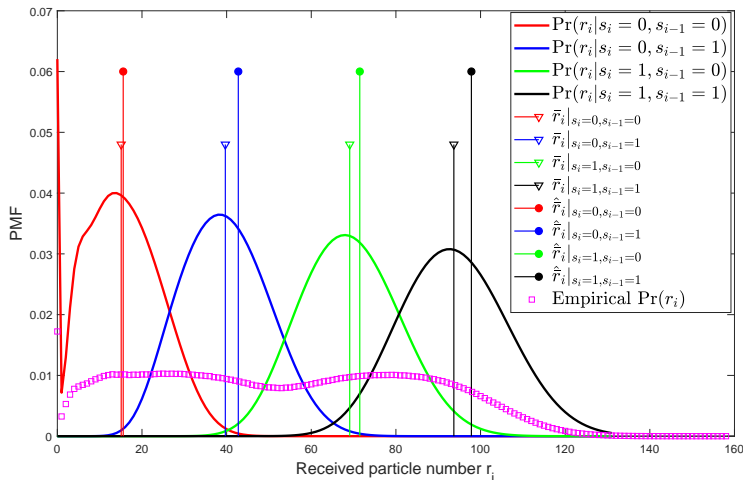


Fig. 2: The positions of $\bar{r}_i|_{s_i, s_{i-1}}$ (calculated via (26)) and $\hat{\bar{r}}_i|_{s_i, s_{i-1}}$ corresponding to the centroids calculated from (24) and (25) and assuming that the labels $[s_i, s_{i-1}]$ are known) for $T = 30\Delta T$

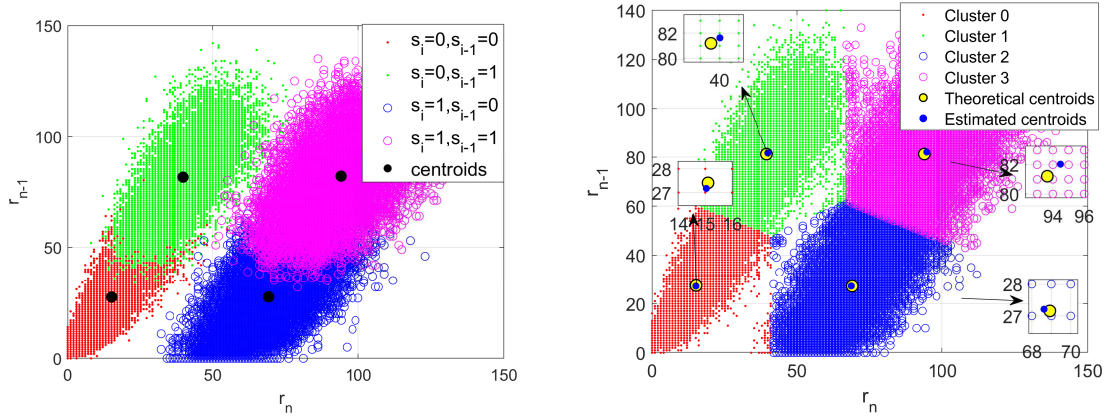
C. Multi-Dimensional Clustering

The proposed solution to increase the accuracy of estimating the average number of received particles from the empirical data relies on multi-dimensional clustering. The main idea is not to apply the K-means clustering algorithm in (24) and (25) to individual observations r_n but to vectors of observations:

$$\mathbf{r}_n = [r_n, r_{n-1}, \dots, r_{n-\mathcal{L}}] \quad (27)$$

where \mathcal{L} is the number of memory bits that we can use for detection and for multi-dimensional clustering. In order to illustrate the advantages of the proposed approach, let us consider the same example as in Fig. 2, but by assuming that the K-means clustering algorithm in (24) and (25) is applied to the vector $[r_n, r_{n-1}]$, rather than to r_n only.

The results are illustrated in Fig. 3, where the following notation is used. The cluster of points whose label is, e.g., $[s_i = 0, s_{i-1} = 1]$ is denoted by $[r_n, r_{n-1}]|_{s_i=0, s_{i-1}=1}$. A similar notation is used for the other clusters. By definition, $\bar{r}_i|_{s_i, s_{i-1}} = s_i C_0 + s_{i-1} C_1 + \sum_{j=2}^L C_j / 2 + \bar{\lambda}_0 T$ and $\bar{r}_{i-1}|_{s_{i-1}} = s_{i-1} C_0 + \sum_{j=1}^L C_j / 2 + \bar{\lambda}_0 T$ given the label $[s_i, s_{i-1}]$. Therefore, the corresponding



(a) Empirical distributions of $[r_n, r_{n-1}]_{s_i, s_{i-1}}$ and theoretical centroids obtained from (28). In this case, the labels are assumed to be known a priori (b) Theoretical centroids obtained by using (28) (yellow circles) vs. estimated centroids obtained by using the K-means algorithms in (24) and (25) (blue circles). The labels are not known a priori but are obtained from the estimated centroids (the updated initial centroids with assigned labels)

Fig. 3: Theoretical and estimated centroids obtained from clustering the data vector $[r_n, r_{n-1}]$ ($T = 30\Delta T$)

theoretical centroids are computed from (14) as follows:

$$\begin{aligned}
 [\bar{r}_i, \bar{r}_{i-1}]_{s_i=0, s_{i-1}=0} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [\bar{\lambda}_0 T, \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]_{s_i=0, s_{i-1}=1} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_1 + \bar{\lambda}_0 T, C_0 + \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]_{s_i=1, s_{i-1}=0} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_0 + \bar{\lambda}_0 T, \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]_{s_i=1, s_{i-1}=1} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_0 + C_1 + \bar{\lambda}_0 T, C_0 + \bar{\lambda}_0 T]
 \end{aligned} \tag{28}$$

Given a sequence of received signals (observations), $\{r_1, \dots, r_n, \dots, r_K\}$, the vectors of observations are $\{[r_2, r_1], \dots, [r_n, r_{n-1}], \dots, [r_K, r_{K-1}]\}$. The empirical distributions of each cluster of points and the theoretical centroids (computed from (28)) are depicted in Fig. 3(a). We apply the K-means clustering algorithm in (24) and (25) to the empirical vectors using some initial centroids μ_k . The setup of the initial centroids is described in the next sub-section. In general, the exact labels $[s_n, s_{n-1}]$ are unknown, which implies that the association between an estimated centroid and the corresponding label $[s_i, s_{i-1}]$ is not known a priori. We solve this issue by appropriately selecting the initial centroids, and by deciding the association between centroids

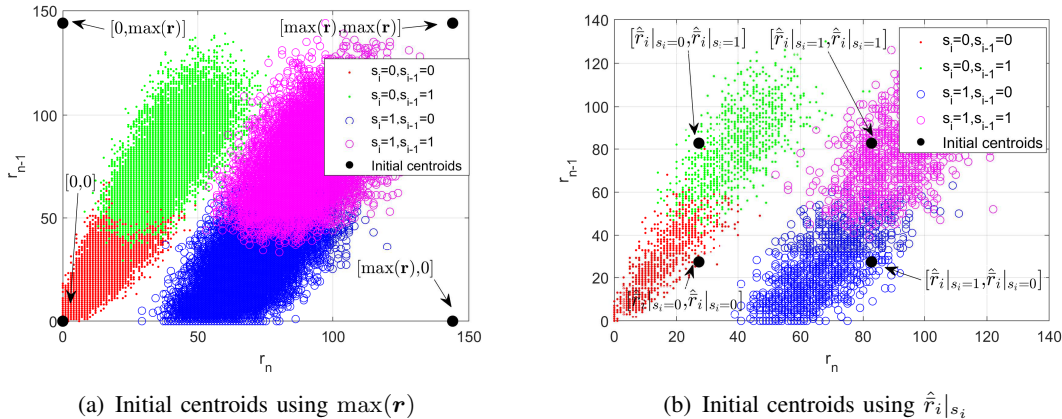


Fig. 4: The customized initial centroids and the corresponding clusters ($T = 30\Delta T$)

and labels at the beginning of the K-means algorithm. Based on our approach, the estimated final centroids inherit the labels associated to the initial centroids. If $[s_i, s_{i-1}]$ is the label of the initial centroid μ_k , then $[s_i, s_{i-1}]$ will be the label of the estimated (final) centroid $\hat{\mu}_k$. Therefore, the choice of the initial centroids is important to ensure a correct labelling. By comparing Fig. 3(b) and Fig. 2, we observe that $\hat{r}_i |_{s_i, s_{i-1}}$ is closer to $\bar{r}_i |_{s_i, s_{i-1}}$ in Fig. 3(b), which results in better performance. Therefore, multi-dimensional clustering is shown to be more accurate than one-dimensional clustering.

D. Setup of the Initial Centroids

As mentioned, an important issue is to assign the correct labels to the initial centroids and to ensure that the association between the estimated centroids and the correct labels does not change when applying the K-means clustering algorithm. In order to motivate the proposed approach for the setup of the initial centroids and the corresponding labels, let us consider Fig. 4(a). In the figure, the four black dots correspond to the points $[0, 0]$, $[0, \max(\mathbf{r})]$, $[\max(\mathbf{r}), 0]$, and $[\max(\mathbf{r}), \max(\mathbf{r})]$, where $\max(\mathbf{r})$ returns that largest value of the observations $\{r_1, \dots, r_n, \dots, r_K\}$. In our proposed approach, the initial centroids of the clusters are chosen to correspond to the four black dots in Fig. 4(a). The rationale for this choice lies in the fact that the maximum received signal can be readily computed from the empirical data. In addition, more importantly, due to the geometry of the problem at hand, our proposed choice leads to a low probability that the label of the centroids will change after the application of the clustering algorithm. By direct inspection of the figure, in fact, the four black dots provide us with the correct label

of the centroids, i.e., $[0, 0]$ corresponds to the label $[s_i, s_{i-1}] = [0, 0]$, $[0, \max(\mathbf{r})]$ corresponds to the label $[s_i, s_{i-1}] = [0, 1]$, etc. If K is sufficiently large, the proposed association between centroids and labels is expected not to change, with high probability, during the application of the K-means algorithm.

If the number of memory-bits \mathcal{L} is greater than one, the initial centroids are assigned in a similar fashion. If $\mathcal{L} = 2$, for example, the initial centroids are $[0, 0, 0]$ whose label is $[s_i, s_{i-1}, s_{i-2}] = [0, 0, 0]$; $[0, 0, \max(\mathbf{r})]$ whose label is $[s_i, s_{i-1}, s_{i-2}] = [0, 0, 1]$; $[0, \max(\mathbf{r}), \max(\mathbf{r})]$ whose label is $[s_i, s_{i-1}, s_{i-2}] = [0, 1, 1]$; etc.

Based on these proposed initial centroids and the corresponding labels, the K-means clustering algorithm is applied according to (24) and (25), which returns the final estimated centroids ($\hat{\boldsymbol{\mu}}_k$) and the indicator variables $\kappa_{n,k}$. In particular, $\kappa_{n,k}$ allows us to implicitly perform data detection at the end of the clustering process, since it informs us, by definition, if a vector of points belongs or not to a given cluster. Based on this remark, we propose two algorithms for data detection that are referred to as (1) direct clustering-based inference and (2) clustering-plus-threshold detection.

E. Direct Clustering-Based Inference and Clustering-plus-Threshold Detection

The proposed *direct clustering-based inference* algorithm is based only on clustering methods and does not exploit the thresholds in (17) for data detection. The algorithm is given in Algorithm 1 and works as follows (assuming, e.g., $\mathcal{L} = 1$). If the observation vector $[r_n, r_{n-1}]$ belongs to the cluster with label $[s_i, s_{i-1}]$, i.e., the corresponding indicator variables are $\kappa_{n,k} = 1$, then the estimated bits are those of the corresponding label.

The proposed *clustering-plus-threshold detection* algorithm, on the other hand, combines together clustering methods and the estimated thresholds in (17) for data detection. The algorithm is given in Algorithm 2 and it has one main difference compared with Algorithm 1: After finalizing the clustering process based on the K-means algorithm, the bits are not detected by using the indicator variables $\kappa_{n,k}$, but the thresholds in (17) are computed from the estimated centroids, which are then used for data detection by using (18).

In order to appreciate the difference between the two algorithms and the advantages and limitations of clustering and using or not using the detection thresholds in (17), we provide some results in Figs. 5-8. These figures are obtained as follows. Consider, for example, the case study $\mathcal{L} = 1$ for ease of exposition. After estimating the received data \hat{s}_n , all the pairs of observations $[r_n, r_{n-1}]$ that are detected as $[\hat{s}_n, \hat{s}_{n-1}] = [0, 0]$, $[\hat{s}_n, \hat{s}_{n-1}] = [0, 1]$, $[\hat{s}_n, \hat{s}_{n-1}] = [1, 0]$, and

Algorithm 1 Direct clustering-based inference

First Step: Clustering

- 1: Set \mathcal{L}
- 2: Construct the data \mathbf{r}_n via (27)
- 3: Construct the initial centroids $\boldsymbol{\mu}_k$
- 4: Cluster the \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$

Second Step: Inference

- 5: Infer \hat{s}_n from the indicator variables $\kappa_{n,k}$
-

Algorithm 2 Clustering-plus-threshold detection

First Step: Clustering

- 1: Set \mathcal{L}
- 2: Construct the data \mathbf{r}_n via (27)
- 3: Construct the initial centroids $\boldsymbol{\mu}_k$
- 4: Cluster the \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$

Second Step: Detection

- 5: Obtain $\hat{r}_i | \hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}$ from the estimated centroids $\hat{\boldsymbol{\mu}}_k$
 - 6: Compute the detection thresholds using (17)
 - 7: Detect the symbols using (18)
-

$[\hat{s}_n, \hat{s}_{n-1}] = [1, 1]$ are depicted in red, green, blue, and magenta colors, respectively. The estimated centroids are depicted as black dots. The figures report the empirical decision boundaries that are obtained after clustering (Algorithm 1) and after applying the empirical thresholds estimated from the centroids (Algorithm 2). We note that the two algorithms provide different results, and their advantages and limitations depend on the severity of the ISI and number of memory bits. It is worth mentioning that in Figs. 6 and 8, due to the many memory bits used, there exist more than four clusters and the dimension of the clusters is greater than two. To be able to illustrate the results, we merge together the clusters whose \mathcal{L} -tuples $[r_n, r_{n-1}, \dots, r_{n-\mathcal{L}}]$ have the same pair $[\hat{s}_n, \hat{s}_{n-1}]$. By using this approach, we obtain again four clusters that can be readily represented. The merged clusters are visualized by using the same color-based code as for $\mathcal{L} = 1$.

By direct inspection of the figures, the following conclusions can be drawn. From Fig. 5, i.e., for mild ISI, we observe that Algorithm 1 outperforms Algorithm 2 if $\mathcal{L} = 1$. In particular, Fig. 5(b) yields decision regions that are closer, as compared with Fig. 5(c), to the theoretical ones in Fig. 5(a). From Fig. 6, on the other hand, we obtain an opposite trend if $\mathcal{L} = 4$. Therefore,

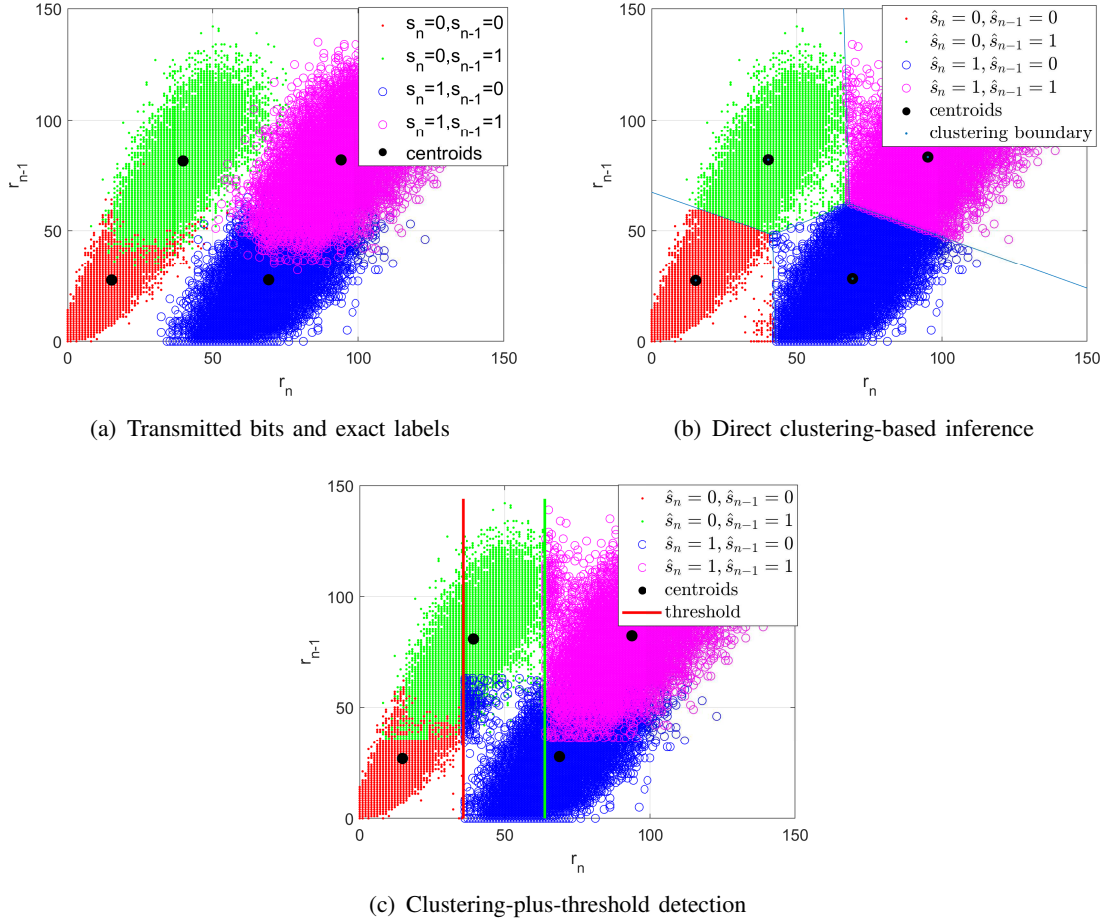


Fig. 5: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1$ and $T = 30 \Delta T$).

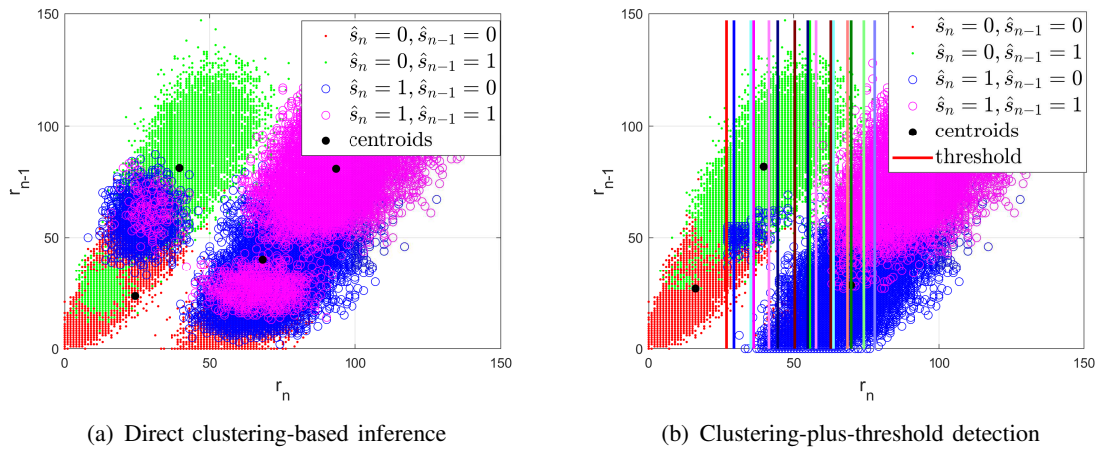


Fig. 6: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$ and $T = 30 \Delta T$).

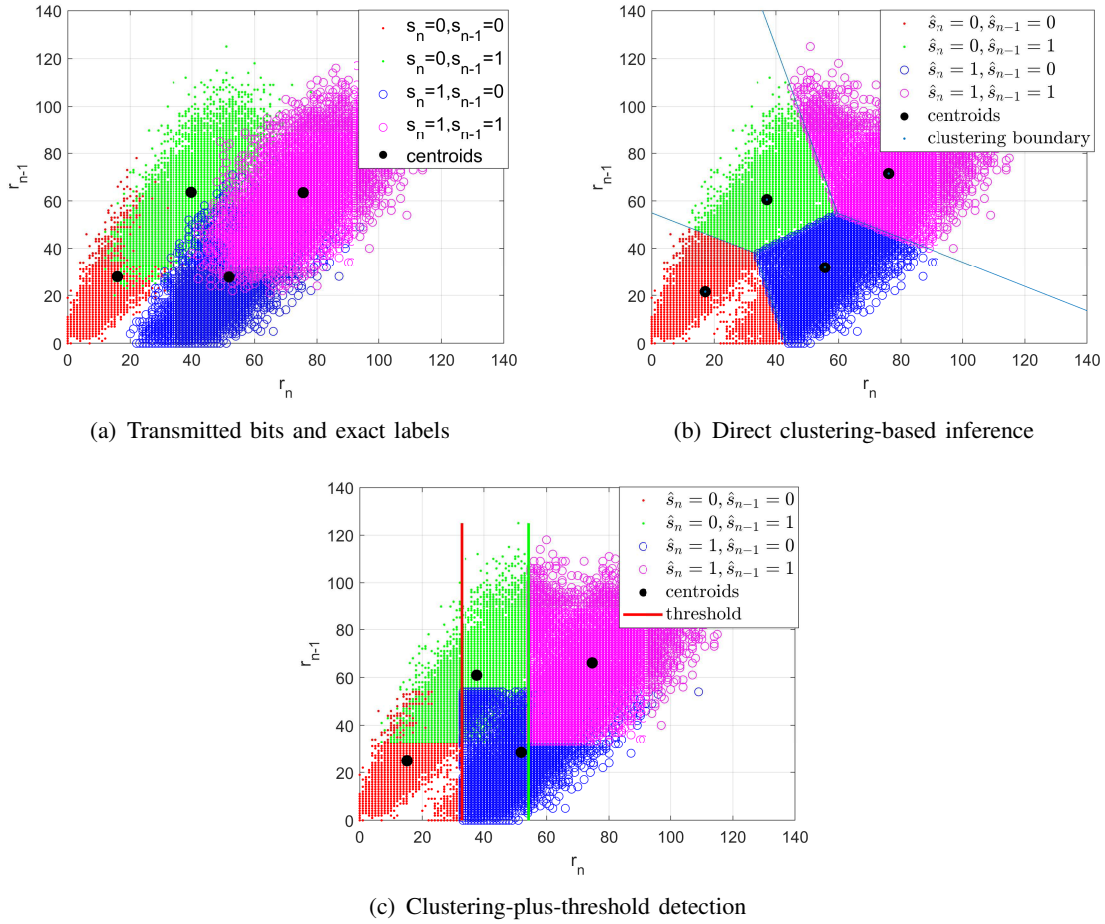


Fig. 7: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1, T = 20 \Delta T$).

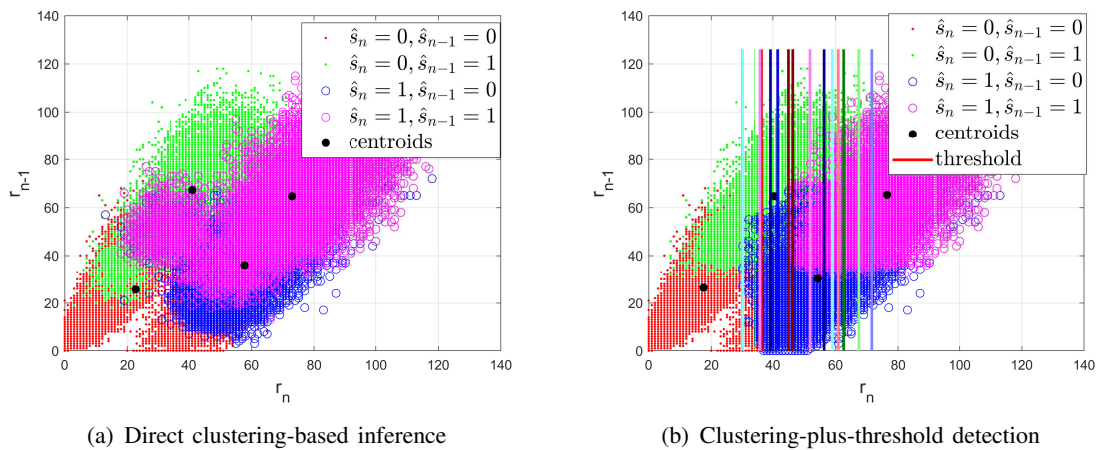


Fig. 8: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4, T = 20 \Delta T$).

we evince that Algorithm 1 with few memory bits may be a sufficiently good approach for non-coherent data detection under mild ISI. Under severe ISI, on the other hand, the performance trends are different, as illustrated in Figs. 7 and 8. We observe, in particular, that the best estimation performance is obtained by using Algorithm 2 with a large number of memory bits, i.e., $\mathcal{L} = 4$ in Fig. 8(b). Errors are still clearly visible, but the estimated clusters of points are closer to the expected ones, i.e., those shown in Fig. 7(a). It is worth mentioning that in the presence of severe ISI the combination of clustering methods and detection thresholds provides better performance than using only clustering. The use of detection thresholds, in particular, allows us to correct the increased number of mis-detections that originate from increasing the dimension of clustering. These *qualitative* conclusions that are drawn from the direct inspection of the estimated clusters from the detected bits are corroborated in Section VI with the aid of BER simulations.

V. ITERATIVE ALGORITHM FOR COMPUTING THE INITIAL CENTROIDS

In the previous section, we have illustrated the core ideas of the proposed clustering-based algorithms for non-coherent detection in MC systems, which may or may not use detection thresholds. The comparison of, e.g., Fig. 7(a) and Fig. 8(b) reveals, however, that detection errors still exist, especially in the presence of non-negligible ISI. The sources of error include the initial estimates of the centroids and the resulting estimates of the detection thresholds. In this section, motivated by these considerations, we introduce improved, iterative-based, clustering-based methods in order to enhance the detection performance in the presence of severe ISI.

The initial centroids obtained from $\max(\mathbf{r})$, e.g., $[0, 0]$, $[0, \max(\mathbf{r})]$, $[\max(\mathbf{r}), 0]$ and $[\max(\mathbf{r}), \max(\mathbf{r})]$ if $\mathcal{L} = 1$, are, in general, not always sufficiently close to the theoretical centroids. This may result in detection errors. In this section, therefore, we introduce a more accurate iterative-based approach for estimating the initial centroids.

In order to illustrate the proposed approach, let us consider $\mathcal{L} = 1$ (two-dimensional clustering) and the cluster whose label is $[s_i, s_{i-1}] = [0, 0]$. Based on (28), the theoretical centroid that corresponds to $[s_i, s_{i-1}] = [0, 0]$ is $[\bar{r}_i|_{s_i=0, s_{i-1}=0}, \bar{r}_{i-1}|_{s_i=0, s_{i-1}=0}]$, where:

$$\bar{r}_{i-1}|_{s_i=0, s_{i-1}=0} = \bar{r}_{i-1}|_{s_{i-1}=0} = \sum_{j=1}^L C_j/2 + \bar{\lambda}_0 T \quad (29)$$

$$\bar{r}_i|_{s_i=0, s_{i-1}=0} = \sum_{j=2}^L C_j/2 + \bar{\lambda}_0 T \quad (30)$$

From (29) and (30), we observe that a good estimate for the initial centroid is $[a, a]$ with $a = \sum_{j=1}^L C_j/2 + \bar{\lambda}_0 T$, since this estimate would be close to the theoretical centroid, and, in general, it is closer than the initial centroid $[0, 0]$ that was used in the previous section. Even though a depends on C_j for $j = 1, 2, \dots, L$, an estimate for it can be obtained from one-dimensional clustering without any prior CSI. In particular, a is, by definition, approximately equal to $a = \bar{r}_i|_{s_i=0} \simeq \hat{r}_i|_{s_i=0}$, where $\hat{r}_i|_{s_i=0}$ is the estimated centroid obtained from one-dimensional clustering. A similar approach can be used to estimate the centroids of the other clusters. The new initial centroids constructed by using $\hat{r}_i|_{s_i}$ are depicted in Fig. 4(b) for completeness. By direct inspection, we can infer that they are closer to the theoretical centroids, and, therefore, are expected to provide better performance.

In general, the proposed approach lies in setting the initial centroids for \mathcal{L} -dimensional clustering from the estimated centroids obtained by applying $(\mathcal{L} - 1)$ -dimensional clustering. In turn, the initial centroids for $(\mathcal{L} - 1)$ -dimensional clustering are obtained from the estimated centroids obtained by applying $(\mathcal{L} - 2)$ -dimensional clustering. This procedure can be iterated until one-dimensional clustering, whose two initial centroids can be initialized to 0 and $\max(\mathbf{r})$.

The proposed approach for setting the initial centroids in an iterative-based manner is summarized in Algorithm 3 and Algorithm 4 for application to iterative clustering-based inference and iterative clustering-plus-threshold detection, respectively. In particular, the proposed approach for setting the initial centroids works in an iterative manner, as follows: (i) one-dimensional clustering is applied to the observations $\{r_1, \dots, r_K\}$ by setting 0 and $\max(\mathbf{r})$ as the initial centroids; (ii) the K-means clustering algorithm is applied and the estimated centroids $\hat{r}_i|_{s_i=0}$ and $\hat{r}_i|_{s_i=1}$ are obtained; (iii) the estimated centroids $\hat{r}_i|_{s_i=0}$ and $\hat{r}_i|_{s_i=1}$ are used to construct new initial centroids for application to two-dimensional clustering. In particular, the four initial centroids are set to $[\hat{r}_i|_{s_i=0}, \hat{r}_i|_{s_i=0}]$, $[\hat{r}_i|_{s_i=0}, \hat{r}_i|_{s_i=1}]$, $[\hat{r}_i|_{s_i=1}, \hat{r}_i|_{s_i=0}]$, and $[\hat{r}_i|_{s_i=1}, \hat{r}_i|_{s_i=1}]$; (iv) two-dimensional clustering is applied to the vectors of observations $\{[r_2, r_1], \dots, [r_K, r_{K-1}]\}$ by using the initial estimated centroids; (v) the K-means (two-dimensional) clustering algorithm is applied again and the estimated two-dimensional centroids $[\hat{r}_i, \hat{r}_{i-1}]|_{s_i, s_{i-1}}$ are obtained; (vi) the procedure is iterated until the clustering dimension $\mathcal{L} + 1$.

In general terms, the initial centroids for application to the $(l + 1)$ -dimensional clustering

can be constructed from the estimated centroids obtained from l -dimensional clustering. In mathematical terms, let us denote the estimated centroids from l -dimensional clustering by $[\hat{r}_m, \dots, \hat{r}_{m-l+1}]_{s_{m-j}, 0 \leq j \leq l-1}$. The initial centroid $\boldsymbol{\mu}_k$ that corresponds to the label $[s_i, \dots, s_{i-l+1}, s_{i-l}]$ for $(l+1)$ -dimensional clustering can be obtained as follows:

$$\boldsymbol{\mu}_k = [\hat{r}_m |_{s_{m-j}=s_{i-j}, 0 \leq j \leq l-1}, \hat{r}_m |_{s_{m-j}=s_{i-1-j}, 0 \leq j \leq l-1}, \hat{r}_{m-1} |_{s_{m-j}=s_{i-1-j}, 1 \leq j \leq l-1}, \dots, \hat{r}_{m-l+1} |_{s_{m-l+1}=s_{i-l}}] \quad (31)$$

which does not require any prior CSI information.

Algorithm 3 Iterative clustering-based inference

- 1: Set the initial centroids $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\mu}_1 = \max(\mathbf{r})$
 - 2: **for** $l = 1$ to $\mathcal{L} + 1$ **do**
 - 3: Construct the data $\mathbf{r}_n = [r_n, \dots, r_{n-l+1}]$
 - 4: Cluster \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$
 - 5: Set the new initial centroids to $\boldsymbol{\mu}_k$ by using (31)
 - 6: **end for**
 - 7: Infer \hat{s}_n from the indicator variables $\kappa_{n,k}$
-

Algorithm 4 Iterative clustering-plus-threshold detection

- 1: Set the initial centroids $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\mu}_1 = \max(\mathbf{r})$
 - 2: **for** $l = 1$ to $\mathcal{L} + 1$ **do**
 - 3: Construct the data $\mathbf{r}_n = [r_n, \dots, r_{n-l+1}]$
 - 4: Cluster \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$
 - 5: Set the new initial centroids to $\boldsymbol{\mu}_k$ by using (31)
 - 6: **end for**
 - 7: Obtain $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the estimated centroids $\hat{\boldsymbol{\mu}}_k$
 - 8: Compute the detection thresholds using (17)
 - 9: Detect the symbols using (18)
-

By considering the same case study as in Fig. 8, we illustrate the performance obtained by employing the proposed iterative approach in Algorithms 3 and 4 in Fig. 9. We observe that, even in the presence of severe ISI, much better estimation performance is obtained by a more accurate initial estimate of the initial centroids. These results confirm the effectiveness of the proposed iterative-based approach.

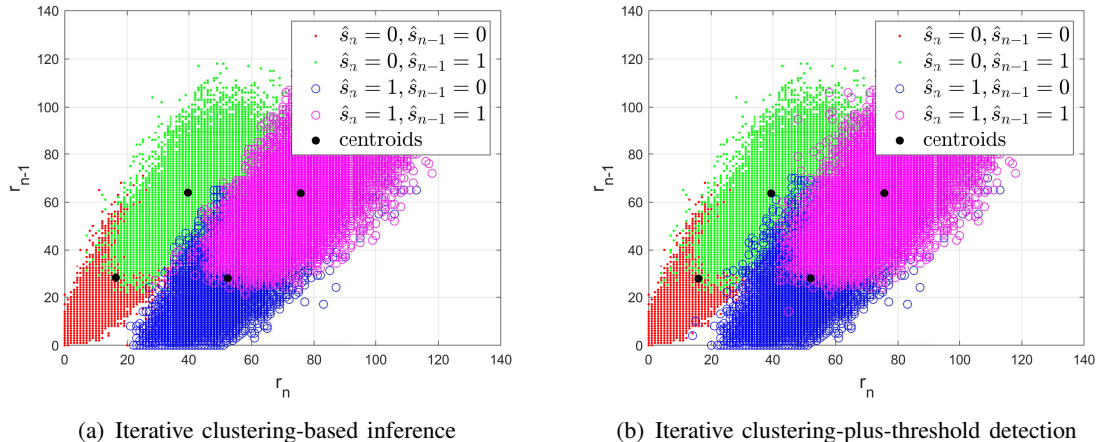


Fig. 9: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$, $T = 20\Delta T$).

VI. NUMERICAL RESULTS

In this section, we report some simulation results in order to quantitatively analyze the performance of the proposed non-coherent detectors. In particular, we compare the proposed methods with a few benchmark schemes available in the literature. The simulation parameters are listed in Table I.

A. Computational Complexity

Before analyzing the error performance of the proposed clustering-based algorithms, we study the computational complexity of the proposed schemes in Table II. In our study, we assume that Step I and Step II of the K-means clustering algorithms are repeated P times. For instance, if we consider K received signals stored and \mathcal{L} memory bits, (the centroid dimension is $\mathcal{L} + 1$ and there are $2^{\mathcal{L}+1}$ initial centroids), in Step I of the K-means clustering algorithm, the evaluation of the distance between one point and a centroid requires $2\mathcal{L} + 1$ additions and $\mathcal{L} + 1$ multiplications. For all points, thus, we obtain $2^{\mathcal{L}+1}K(2\mathcal{L} + 1)$ additions and $2^{\mathcal{L}+1}K(\mathcal{L} + 1)$ multiplications. In Step II, we need $(K - 2^{\mathcal{L}+1})(\mathcal{L} + 1)$ additions and $2^{\mathcal{L}+1}$ multiplications. Since the computational complexity of Step II is negligible compared with the computational complexity of Step I, we omit the complexity of Step II and report only the computational complexity of Setp I in Table II.

TABLE I: Simulation parameters

Parameter	Value
λ_0	$100s^{-1}$
Receiver radius r	45 nm
Distance d	500 nm
Diffusion coefficient D	$4.265 \times 10^{-10}m^2/s$
Discrete time length ΔT	9 us
Channel length L	5
K	2^{12}

TABLE II: Computational complexity

Algorithm	Additions	Multiplications
Algorithm 1, 2	$2^{\mathcal{L}+1}PK(2\mathcal{L} + 1)$	$2^{\mathcal{L}+1}PK(\mathcal{L} + 1)$
Algorithm 3, 4	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i - 1)$	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i - 1)$

B. Channels in the Presence of ISI

In order to evaluate the efficiency of the proposed schemes, we compare the BER performance of the proposed algorithms with some benchmark schemes, including: (i) a receiver that first estimates the channel by using the method proposed in [6] by using K_{train} pilot symbols and then detects the bits by using the threshold in (15); (ii) a receiver that detects the symbols by using the threshold in (15) by assuming that the CSI is perfectly known; (iii) a sequence-based receiver [35] that detects the symbols by assuming perfect CSI (including the knowledge of L) and by using a multi-symbol maximum likelihood (ML) detector based on (13); (iv) the ViterbiNet sequence-based receiver in [35], which knows L and is trained using 2^{10} pilot symbols. In order to show how the channel estimation accuracy affects the detection performance, we evaluate the benchmark scheme for $K_{\text{train}} = 2^6$ and $K_{\text{train}} = 2^{10}$.

In Fig. 10 (mild ISI) and Fig. 11 (severe ISI), we compare the performance of the proposed algorithms against the considered benchmark schemes by assuming a different number of memory bits \mathcal{L} and the case studies in which the CSI is unknown (practical case) and known (benchmark case). We observe that the ML-based sequence detector with perfect CSI knowledge yields the best BER performance, as expected, and provides us with a lower bound. The proposed Algorithm 3 and Algorithm 4 yield good performance if the number of memory bits is not too small, depending on the severity of the ISI. By assuming that the CSI is known (e.g., the initial centroids are perfectly known in our algorithms), we observe (see Fig. 10(c) and Fig. 10(d)) that

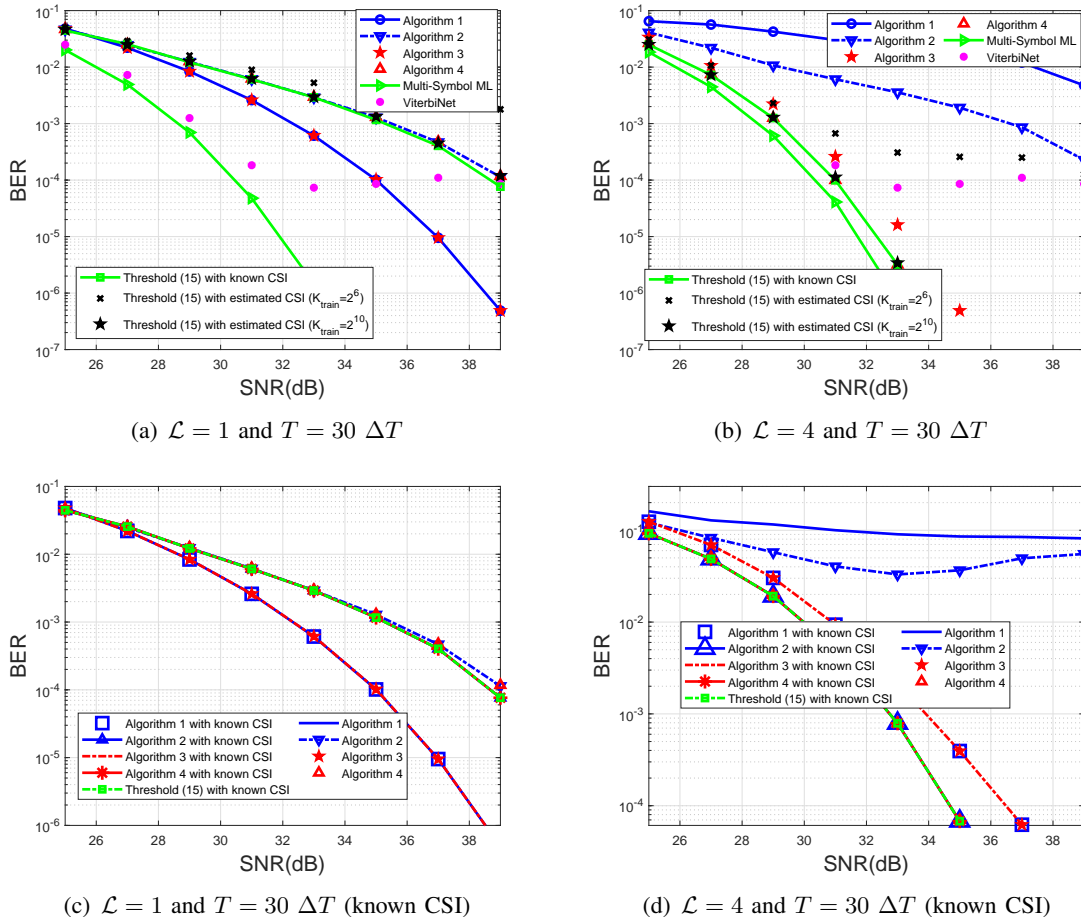


Fig. 10: BER comparison of the proposed algorithms (mild ISI)

the proposed algorithms provide good performance. This shows the robustness of the proposed approach, which works well even under known CSI. We note, in addition, that the ViterbiNet shows an error-floor, in general, because it is derived by assuming a Gaussian white noise, which is different from the considered model.

In particular, we observe that Algorithm 1 and Algorithm 2 do not perform well, in general, in the presence of ISI because of errors in estimating the centroids. By comparing Fig. 10(c) and Fig. 10 (d), we evince that Algorithm 1 and Algorithm 2 have a worse BER if \mathcal{L} increases. This is due to the high dimension of the clustered data. On the other hand, Algorithm 3 and Algorithm 4 compensate for this issue thanks to the iterative approach that is used to estimate the centroids. Based on the computational complexity that can be afforded, which depends on \mathcal{L} , Algorithm 3 and Algorithm 4 are the most suitable options. In particular, Fig. 10(c) shows that Algorithm 3 is the preferred choice if $\mathcal{L} = 1$ and Algorithm 4 is the preferred choice if $\mathcal{L} = 4$.

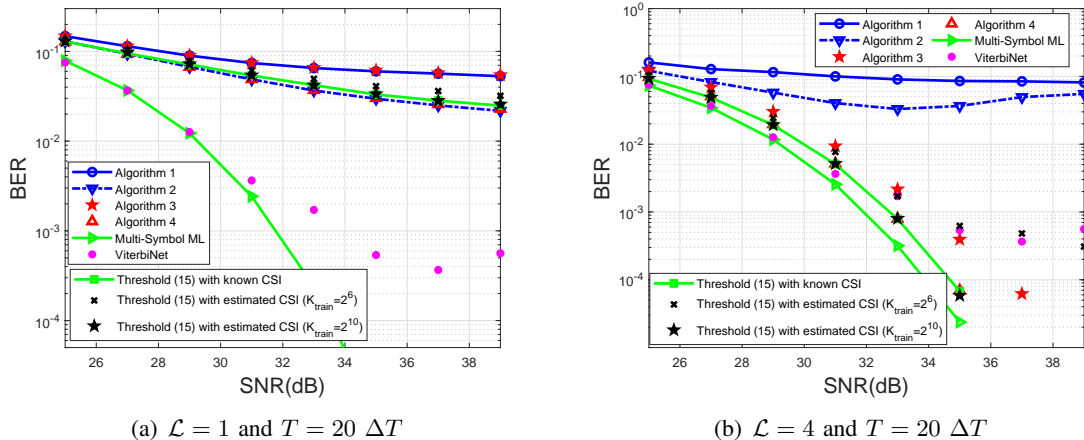


Fig. 11: BER comparison of the proposed algorithms (severe ISI)

The same trends are observed in Fig. 11. In the presence of severe ISI, it is necessary, however, to use a large number of memory bits \mathcal{L} in order to outperform the ViterbiNet, as illustrated in Fig. 11(a).

C. Channels Without ISI

In order to further assess the robustness of the proposed algorithms, we evaluate the BER performance under ISI-free channels. In particular, we compare the BER performance of our algorithms against the benchmark in [36], which is based on a non-coherent detector that uses constant-composition codes. In addition, we consider a benchmark that assumes that the CSI is known and uses the theoretical threshold in (15). The results are illustrated in Fig. 12. We observe that Algorithm 1 gives similar performance as Algorithm 3, and Algorithm 2 gives similar performance as Algorithm 4. This is because no ISI is present and the initial estimate of the centroids is almost correct. In particular, Algorithms 2 and 4 outperform Algorithms 1 and 3 because the decoding thresholds are exploited. In addition, Algorithm 2 and Algorithm 4 yield a BER that is close to the benchmark scheme with known CSI, and they outperform the benchmark in [36].

D. Extension to Multi-Level CSK Modulation

To explore the potential applications of the proposed clustering-based methods, we generalize the proposed approach to multi-level CSK modulation [37]. As an example, we consider a four-

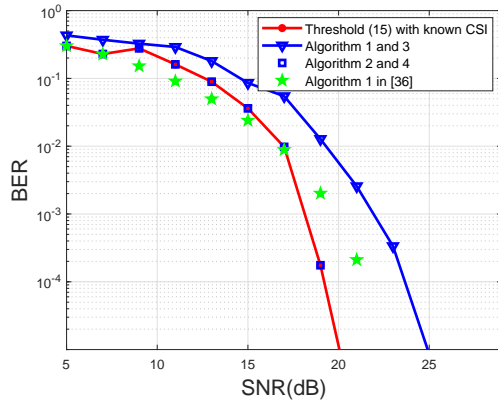


Fig. 12: BER comparison of the proposed algorithms (no ISI, $T = 100 \Delta T$ and $\mathcal{L} = 1$)

level CSK modulation scheme. By using similar analytical steps as those for the derivation of (15), e.g., for one memory symbol, we obtain the threshold:

$$\tau_k|_{s_{i-1}} = \frac{\bar{r}_i|_{s_i=k, s_{i-1}} - \bar{r}_i|_{s_i=k-1, s_{i-1}}}{\ln\left(\frac{\bar{r}_i|_{s_i=k, s_{i-1}}}{\bar{r}_i|_{s_i=k-1, s_{i-1}}}\right)} \quad (32)$$

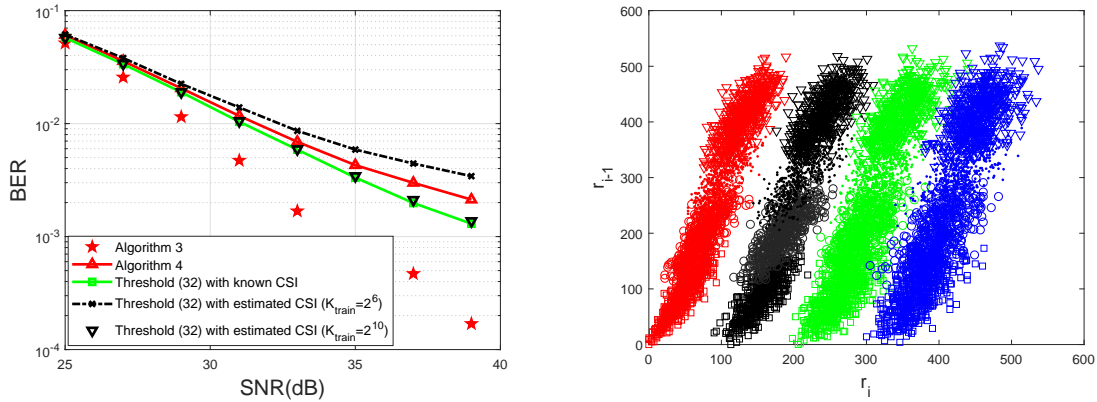
by imposing $\Pr_{\text{approx}}(r_i|s_i = k-1, s_{i-1}) = \Pr_{\text{approx}}(r_i|s_i = k, s_{i-1})$. If $r_i \leq \tau_1|_{s_{i-1}}$, then $\hat{s}_i = 0$; if $\tau_1|_{s_{i-1}} < r_i \leq \tau_2|_{s_{i-1}}$, then $\hat{s}_i = 1$; if $\tau_2|_{s_{i-1}} < r_i \leq \tau_3|_{s_{i-1}}$, then $\hat{s}_i = 2$; otherwise, $\hat{s}_i = 3$.

In a 4-level CSK modulation scheme, there exist 16 clusters for the points $[r_n, r_{n-1}]$ as depicted in Fig. 13(b). In this case, we consider Algorithms 3 and 4 for detection, and, in the first iteration, the initial centroids are set to $\boldsymbol{\mu}_0 = 0$, $\boldsymbol{\mu}_1 = \frac{\max(\mathbf{r})}{3}$, $\boldsymbol{\mu}_2 = \frac{2\max(\mathbf{r})}{3}$ and $\boldsymbol{\mu}_3 = \max(\mathbf{r})$.

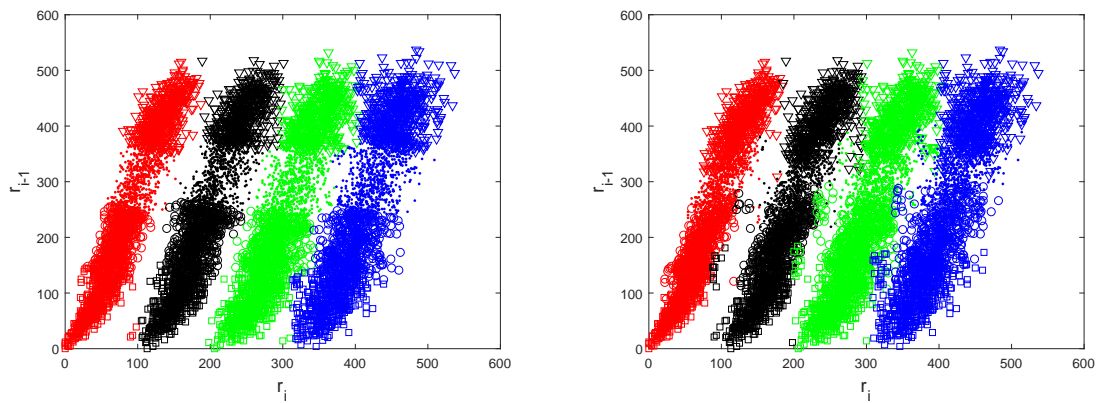
The numerical results are illustrated in Fig. 13. We observe that, in the considered case study, Algorithm 3 outperforms other benchmark algorithms including Algorithm 4. In this case in which multiple symbols are transmitted, in fact, the detection performance of the proposed schemes are inherently more vulnerable to wrong estimations of the decoding thresholds. Therefore, Algorithm 3 outperforms Algorithm 4 in the considered case study. This behavior can be understood by comparing Fig. 13(b) with Fig. 13(c) and Fig. 13(d), in which we evince that Algorithm 3 yields a smaller classification error compared with Algorithm 4.

VII. CONCLUSION

In this paper, we have introduced non-coherent detection schemes for application to MC systems in the presence of ISI. The proposed algorithms are based on combining clustering methods and empirical estimates of the detection thresholds that are employed in memory-bits



(a) BER comparison of the proposed algorithm (mild ISI) (b) Distribution of $[r_n, r_{n-1}]|_{s_i, s_{i-1}}$ with correct labels



(c) Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ with estimated labels using Algorithm 3 (d) Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ with estimated labels using Algorithm 4

Fig. 13: $\mathcal{L} = 1$ and $T = 60 \Delta T$

detection methods. In order to apply the proposed clustering-based algorithms, different methods for initializing the centroids of the clusters directly from the empirical data are proposed and analyzed. Simulation results show that, in the presence of severe ISI, the proposed iterative algorithms that may combine multi-dimensional clustering methods and detection thresholds yield, in general, good BER performance provided that the a sufficient number of memory bits are used.

REFERENCES

- [1] P. Nelson and S. Doniach, *Biological physics: Energy, information life*, WH Freeman New York, 2004.
- [2] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication", *IEEE Commun. Surv. Tutor.*, vol. 18, no. 3, pp. 1887-1919, 2016.

- [3] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, *Molecular Communication*, Cambridge University Press, 2013.
- [4] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks", in *Proc. IEEE ICC*, Jun. 2011, pp. 1-5.
- [5] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise", *Trans. Inf. Theory*, vol. 59, no. 2, pp. 942-954, Sep. 2013.
- [6] V. Jamali, A. Ahmadzadeh, C. Jardin, H. Sticht, and R. Schober, "Channel estimation for diffusive molecular communications", *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4238-4252, Aug. 2016.
- [7] Y. Fang, A. Noel, N. Yang, A. W. Eckford, and R. A. Kennedy, "Symbol-by-symbol maximum likelihood detection for cooperative molecular communication", *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4885-4899, Apr. 2019.
- [8] V. Jamali, N. Farsad, R. Schober, and A. Goldsmith, "Non-coherent detection for diffusive molecular communication systems", *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2515-2531, Jan. 2018.
- [9] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes", *IEEE Trans. Nanobiosci.*, vol. 13, no. 1, pp. 31-43, Jan. 2014.
- [10] R. Mosayebi, A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Type-based sign modulation and its application for ISI mitigation in molecular communication", *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 180-193, Sep. 2017.
- [11] 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, pp. 4228-4232.
- [12] D. Kilinc and O. B. Akan, "Receiver design for molecular communication", *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 705-714, Dec. 2013.
- [13] M. Damrath and P. A. Hoeher, "Low-complexity adaptive threshold detection for molecular communication", *IEEE Trans. Nanobiosci.*, vol. 15, no. 3, pp. 200-208, Jan. 2016.
- [14] X. Qian and M. Di Renzo, "Receiver design in molecular communications: An approach based on artificial neural networks", in *Proc. IEEE ISWCS*, Aug 2018, pp. 1-5.
- [15] X. Qian, M. Di Renzo, and A. Eckford, "Molecular communications: Model-based and data-driven receiver design and optimization", *IEEE Access*, vol. 7, pp. 53555-53565, 2019.
- [16] R. Mosayebi, H. Arjmandi, A. Gohari et al., "Receivers for diffusion-based molecular communication: Exploiting memory and sampling rate", *IEEE J. Sel. Areas Commun.*, vol. 32, no. 12, pp. 2368-2380, Nov. 2014.
- [17] A. Noel, K. C. Cheung, and R. Schober, "Joint channel parameter estimation via diffusive molecular communication", *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 1, pp. 4-17, Aug. 2015.
- [18] X. Bao, J. Lin, and W. Zhang, "Channel modeling of molecular communication via diffusion with multiple absorbing receivers", *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 809-812, Jan. 2019.
- [19] Guo W, Deng Y, Li B, Zhao C, Nallanathan A. "Eavesdropper localization in random walk channels", *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1776-1779, Jul. 2016.
- [20] X. Huang, Y. Fang, A. Noel and N. Yang, "Channel characterization for 1-D molecular communication with two absorbing receivers", *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1150-1154, Jun. 2020,
- [21] J. W. Kwak, H. B. Yilmaz, N. Farsad, C. Chae and A. J. Goldsmith, "Two-way molecular communications", *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3550-3563, Jun. 2020.
- [22] N. V. Sabu and A. K. Gupta, "Detection probability in a molecular communication via diffusion system with multiple fully-absorbing receivers", *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2824-2828, Aug. 2020.
- [23] N. Farsad, D. Pan D, and A. Goldsmith, "A novel experimental platform for in-vessel multi-chemical molecular communications", *Proc. IEEE GLOBECOM*, Dec. 2017, pp. 1-6.

- [24] A. Zappone, M. Di Renzo, M. Debbah et al., "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization", *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60-69, Jul. 2019.
- [25] B. Li, M. Sun, S. Wang, W. Guo, and C. Zhao, "Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications", *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2079-2091, Mar. 2016.
- [26] B. Li, W. Guo, X. Wang, Y. Deng, Y. Lan, C. Zhao, and A. Nallanathan, "Csi-independent non-linear signal detection in molecular communications", *IEEE Trans. Signal Process.*, vol. 68, pp. 97-112, Dec. 2019.
- [27] S. Liu, Z. Wei, B. Li, and C. Zhao, "Unsupervised clustering-based non-coherent detection for molecular communications", *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1687-1690, Aug. 2020.
- [28] C. Luo, X. Wu, L. Lin, C. Wang, and F. Liu, "Non-coherent signal detection technique for mobile molecular communication at high data rates", in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1-6.
- [29] C. Gong and Z. Xu, "Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference", *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5326-5337, May. 2015.
- [30] R. Xu and D. C. Wunsch, "Survey of clustering algorithms", *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 16, no. 3, pp. 645-678, May 2005 2005.
- [31] H. B. Yilmaz, A. C. Heren, T. Tugcu, and C.-B. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing receiver", *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 929-932, Apr. 2014.
- [32] Van der Vaart AW. *Asymptotic statistics*, Cambridge university press, Jun. 2000.
- [33] A. K. Jain et al., "Data clustering: a review", *ACM COMPUT. SURV.*, vol. 31, no. 3, pp. 264-323, Sep. 1999.
- [34] A. C. Bovik, *Handbook of image and video processing*, Academic press, Jul. 2010.
- [35] N. Shlezinger, N. Farsad, Y. C. Eldar and A. J. Goldsmith, "ViterbiNet: A deep learning based viterbi algorithm for symbol detection", *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3319-3331, May 2020.
- [36] V. Jamali, A. Ahmadzadeh, N. Farsad and R. Schober, "Constant-composition codes for maximum likelihood detection without csi in diffusive molecular communications", *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 1981-1995, May 2018.
- [37] S. Galmes and B. Atakan, "Performance analysis of diffusion-based molecular communications with memory", *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3786-3793, Sep. 2016.