

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

Xuewen 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. 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.direnzo@centralesupelec.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, it may be difficult to model the channel theoretically, e.g. for multiple absorbing receivers [18]. To counteract these issues, some authors have used machine learning (ML) methods to design the receiver [19] 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 [20].

### 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 [21], 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 [22], 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 [23], the authors exploit the unsupervised fuzzy C-mean approach to detect symbols from the quick-rising and slow-decaying of the processed signals. In [24], 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 [25]. 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 [26]. 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. 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], [27]:

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

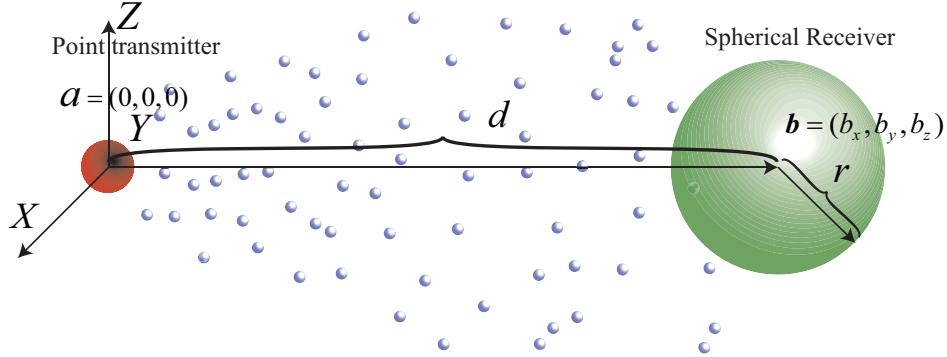


Fig. 1: Three-dimensional MC system consisting of a point transmitter and a spherical absorbing receiver

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 assumed, 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  $i$ th time slot of duration  $T$  is:

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

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

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. Thus, the number of received particles [16] in the  $i$ th time-slot follows a Poisson distribution:

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

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

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

$$\text{SNR} = 10 \log_{10} \frac{C_0}{2\bar{\lambda}_0 T} \quad (4)$$

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

$$N_{TX} = \frac{2\lambda_0 T 10^{\frac{\text{SNR}}{10}}}{P_0} \quad (5)$$

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:

$$\tau|_{s_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{I_i})} = \frac{C_0}{\ln(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j})} \quad (8)$$

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.,  $\tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L \hat{s}_{i-j} C_j})}$ . 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 next sub-section, we show how the average number of particles  $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$  can be obtained directly from the received data without any prior knowledge. 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  $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$ :

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

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 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:

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

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  $\hat{\bar{r}}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ , which is introduced and defined in the next sub-section. Therefore, the detection threshold is expressed as follows:

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

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:

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

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

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

#### A. Estimation of $\hat{\bar{r}}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ 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  $\hat{\bar{r}}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$  can be obtained by only using the received data, which implies that the threshold is (17) can be obtained without prior CSI information.

As a case study, we show how the intermediate variables  $\hat{r}_i|_{\hat{s}_{i-j}, 0 \leq j \leq L}$  are obtained for  $L = 1$ . 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:

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

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 [28], 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 [29] and is defined, in general terms, as follows:

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

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 ( $T = 20\Delta T$ ).

##### A. K-Means Clustering Algorithm

Assume that we have a data set  $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$  of  $N$  observations and each element is a  $\mathbb{D}$ -dimensional vector  $\boldsymbol{x}_n$ . 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  $[s_i, s_{i1}] = [0, 0]$ ,  $[s_i, s_{i1}] = [0, 1]$ ,  $[s_i, s_{i1}] = [1, 0]$ , and  $[s_i, s_{i1}] = [1, 1]$ . From (20) or equivalently (14), we obtain:

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

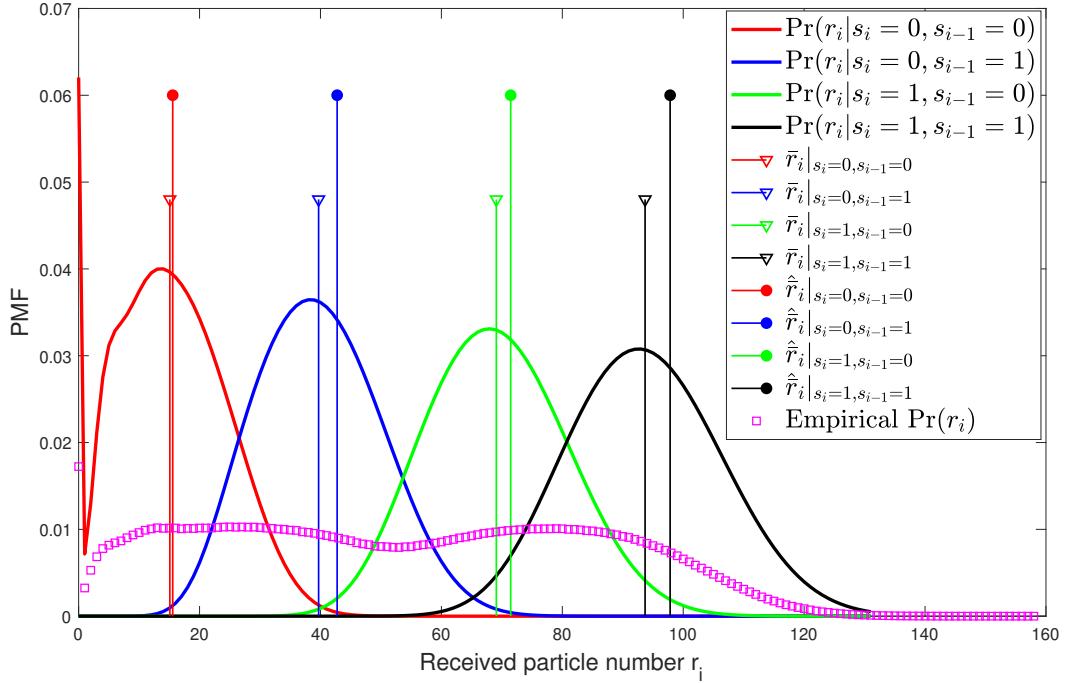


Fig. 2: The positions of  $\bar{r}_i|_{s_i,s_{i-1}}$  (calculated via (26)) and  $\hat{r}_i|_{s_i,s_{i-1}}$  corresponding to the centroids calculated from (24) and (25) and assuming that the labels  $[s_i, s_{i-1}]$  are known) for  $T = 30\Delta T$

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.

### 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:

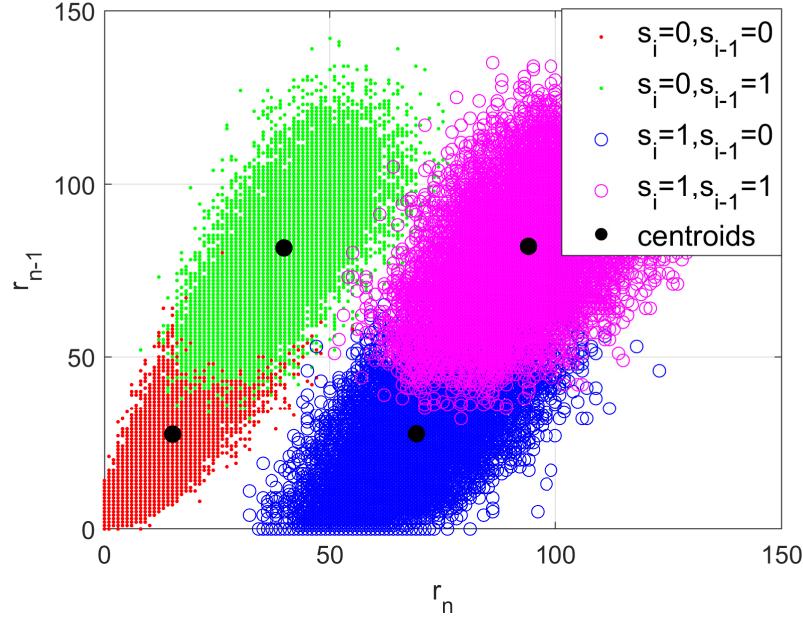
$$\mathbf{r}_n = [r_n, r_{n-1}, \dots, r_{n-\mathcal{L}}] \quad (27)$$

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.

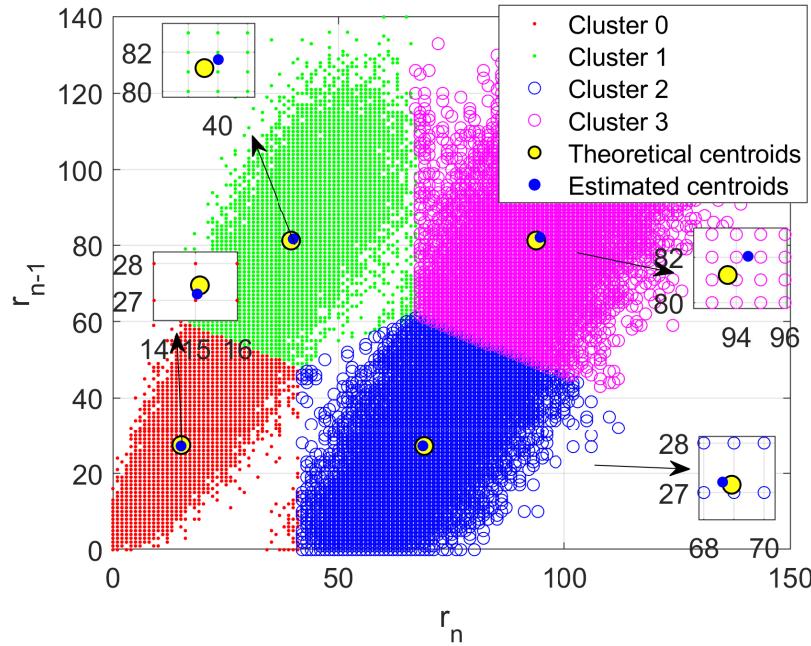
The results are illustrated in Fig. 3, where the following notation is used. The cluster of points whose label is, e.g.,  $[s_i = 0, s_{i-1} = 1]$  is denoted by  $[r_n, r_{n-1}]|_{s_i=0, s_{i-1}=1}$ . A similar notation is used for the other clusters. By definition,  $\bar{r}_i|_{s_i, s_{i-1}} = s_i C_0 + s_{i-1} C_1 + \sum_{j=2}^L C_j / 2 + \bar{\lambda}_0 T$  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 theoretical centroids are computed from (14) as follows:

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

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



(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  $[r_n, r_{n-1}]$  ( $T = 30\Delta T$ )

centroids  $\mu_k$ . 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 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  $\mu_k$ , 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  $\{r_1, \dots, r_n, \dots, r_K\}$ . Our approach consists of choosing as initial centroids the four black dots in Fig. 4(a), which, by direct inspection of the figure, are shown to provide us with the correct labels as well, 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  $[s_i, s_{i-1}, s_{i-2}] = [0, 0, 1]$ ;  $[0, \max(\mathbf{r}), \max(\mathbf{r})]$  whole 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 ( $\hat{\mu}_k$ ) 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

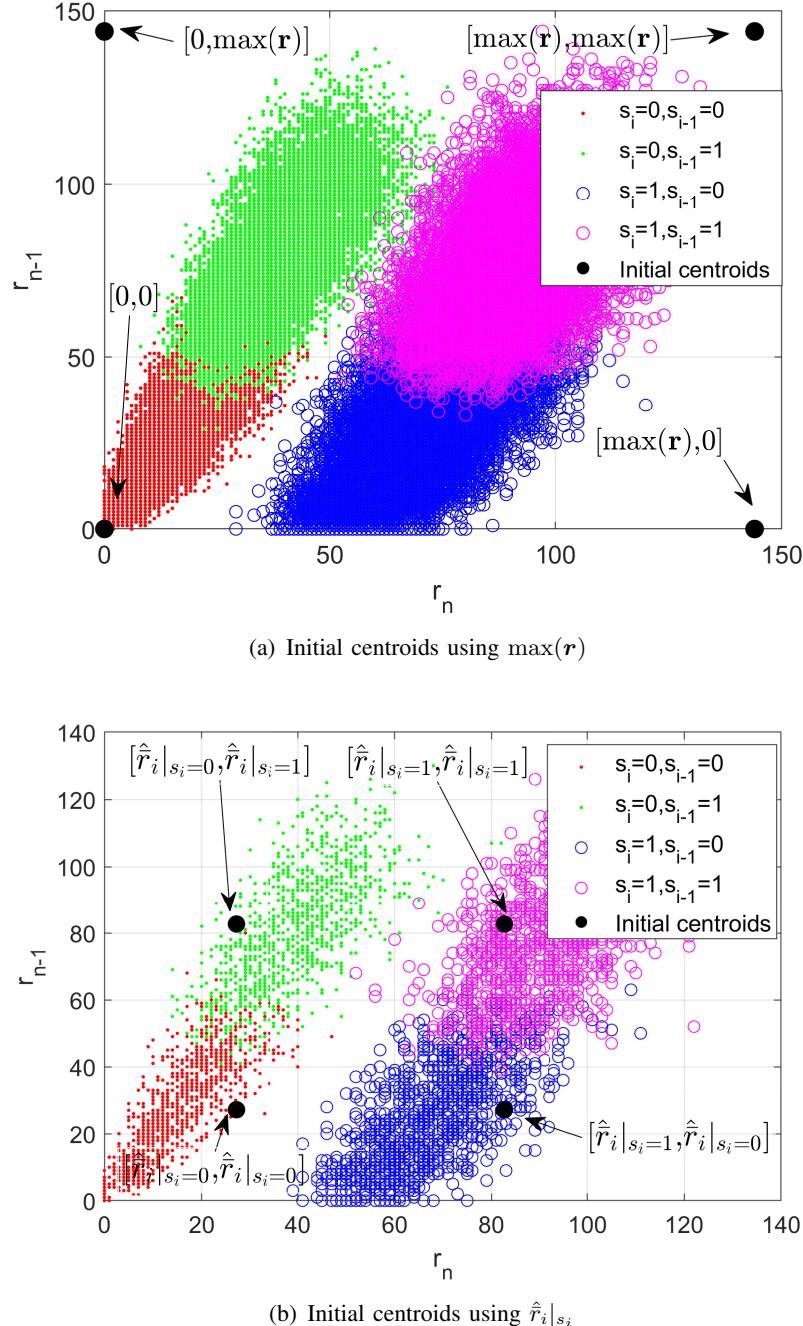


Fig. 4: The customized initial centroids and the corresponding clusters ( $T = 30\Delta T$ )

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.

---

**Algorithm 1** Direct clustering-based inference

---

**First Step:** Clustering

- 1: Set  $\mathcal{L}$
- 2: Construct the data  $r_n$  via (27)
- 3: Construct the initial centroids  $\mu_k$
- 4: Cluster the  $r_n$  using the K-means algorithm with the initial centroids  $\mu_k$

**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  $r_n$  via (27)
- 3: Construct the initial centroids  $\mu_k$
- 4: Cluster the  $r_n$  using the K-means algorithm with the initial centroids  $\mu_k$

**Second Step:** Detection

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

*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 indicators 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

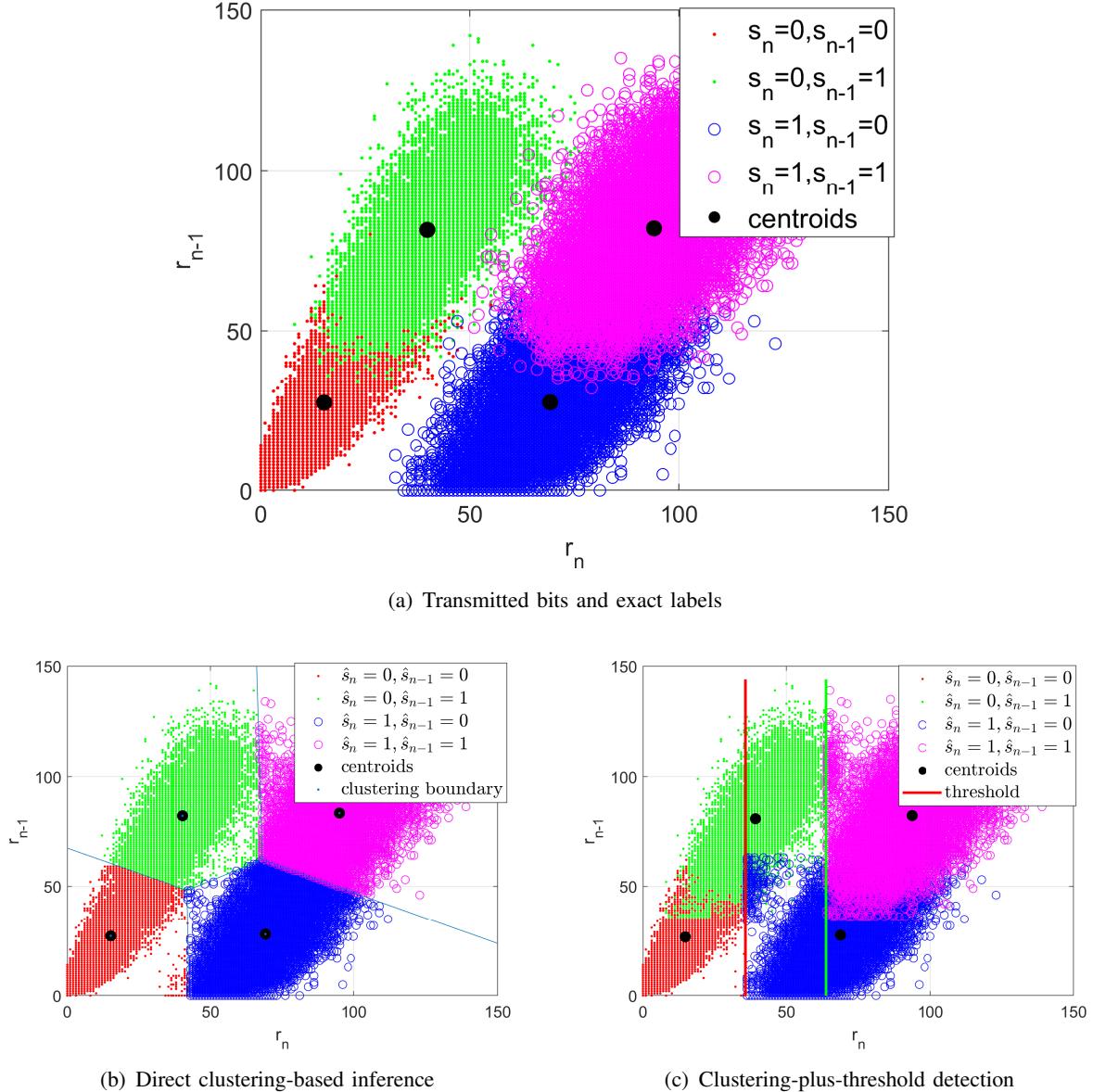


Fig. 5: Distribution of  $[r_n, r_{n-1}] | \hat{s}_n, \hat{s}_{n-1}$ , estimated centroids and clustering boundaries/thresholds ( $\mathcal{L} = 1$  and  $T = 30 \Delta T$ ).

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  $[\hat{s}_n, \hat{s}_{n-1}] = [1, 1]$  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

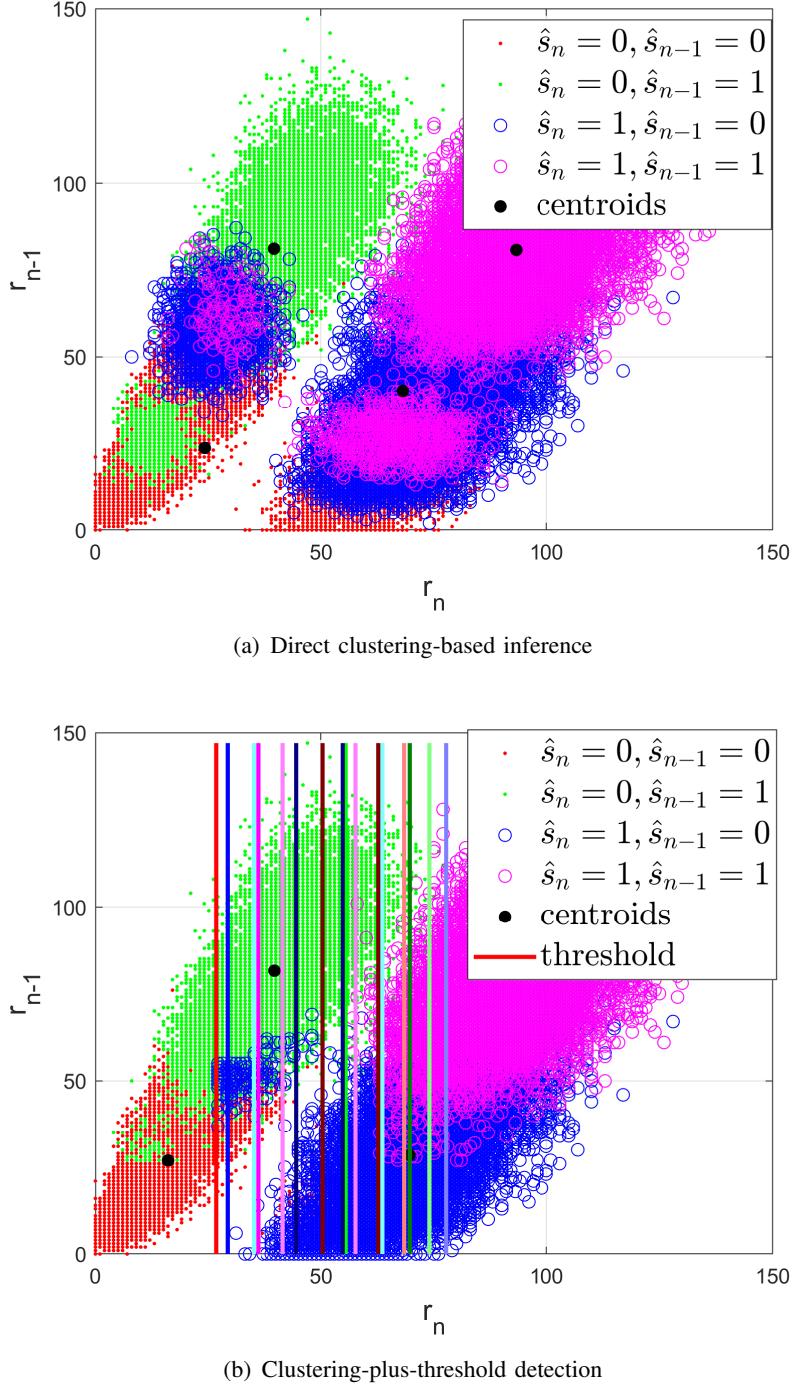


Fig. 6: Distribution of  $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$  and estimated centroids ( $\mathcal{L} = 4$  and  $T = 30 \Delta T$ ).

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

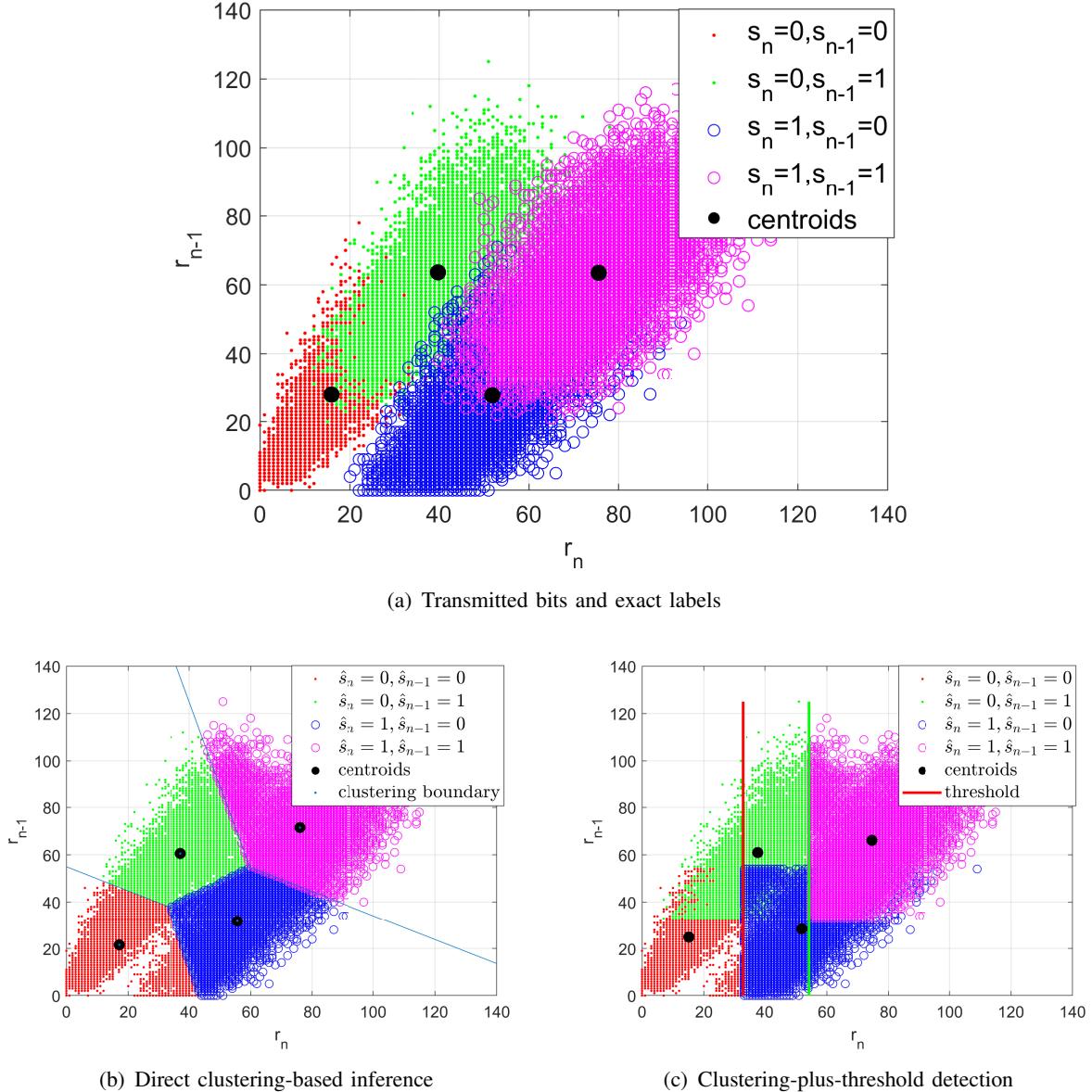


Fig. 7: Distribution of  $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ , estimated centroids and clustering boundaries/thresholds ( $\mathcal{L} = 1$ ,  $T = 20 \Delta T$ ).

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.,

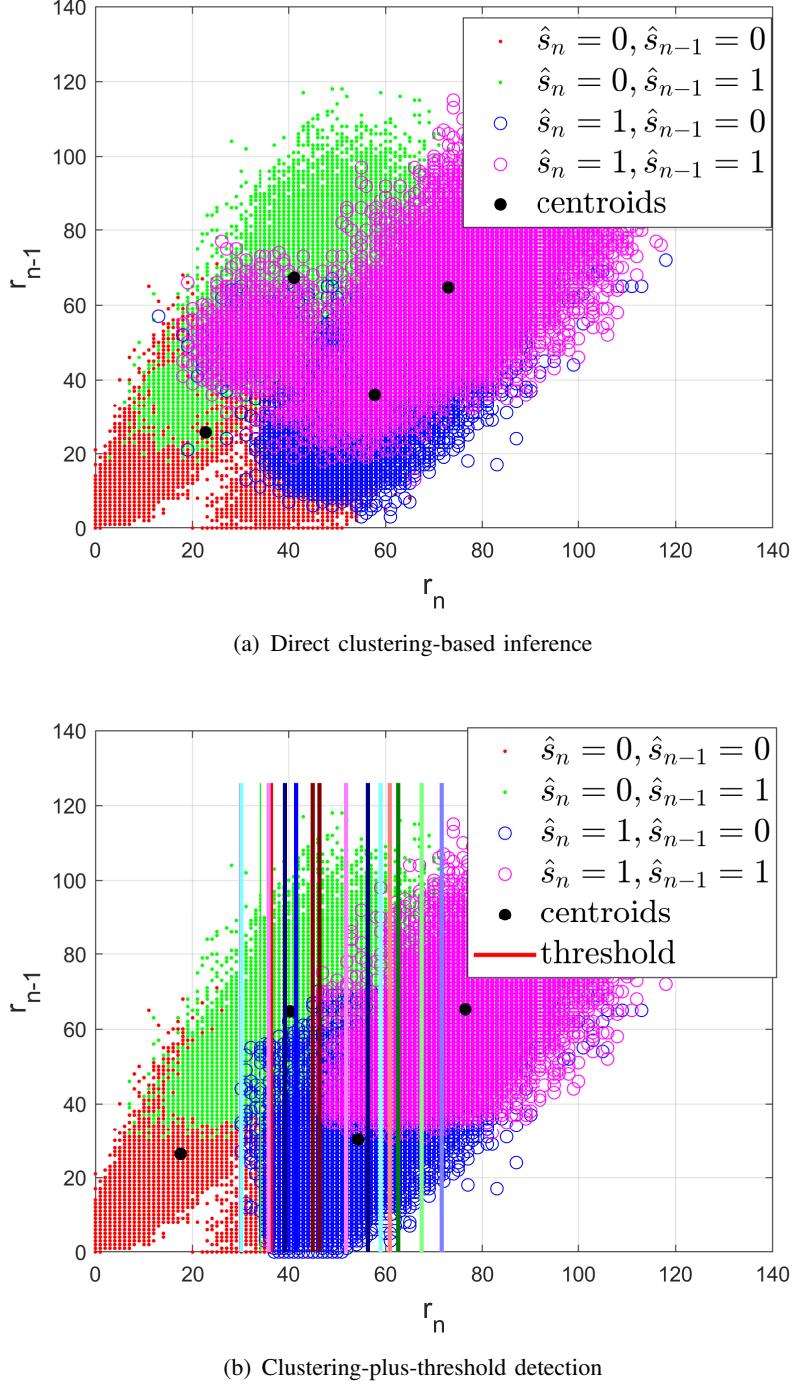


Fig. 8: Distribution of  $[r_n, r_{n-1}]_{\hat{s}_n, \hat{s}_{n-1}}$  and estimated centroids ( $\mathcal{L} = 4$ ,  $T = 20 \Delta T$ ).

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,

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.

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

$[\hat{\bar{r}}_m, \dots, \hat{\bar{r}}_{m-l+1}]|_{s_{m-j}, 0 \leq j \leq l-1}$ . 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:

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

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  $\hat{\bar{r}}_i|_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$  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.

## VI. NUMERICAL RESULTS

In this section, we report some simulation results in order to quantitatively analyze the performance of the proposed non-coherent detectors. The simulation parameters are listed in

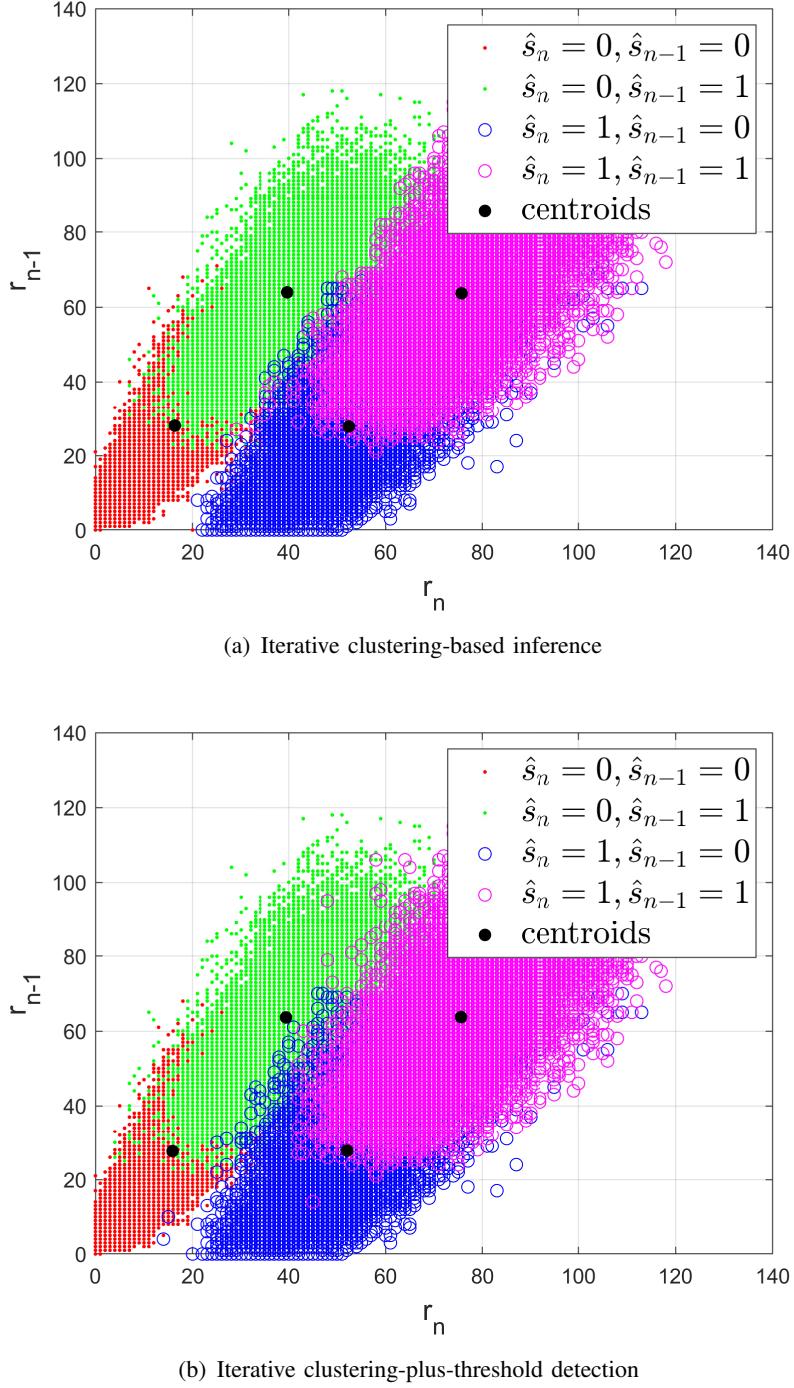


Fig. 9: Distribution of  $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$  and estimated centroids ( $\mathcal{L} = 4$ ,  $T = 20\Delta T$ ).

Table I. For completeness, we summarize the computational complexity of the proposed schemes in Table II, by assuming that the K-means clustering algorithms is repeated  $P$  times.

Figure 10(a) depicts the BER for  $\mathcal{L} = 1$ . As qualitatively illustrated in Fig. 5, we observe

TABLE I: Simulation parameters

Parameter	Value
$\lambda_0$	$100\text{s}^{-1}$
Receiver radius $r$	45 nm
Distance $d$	500 nm
Diffusion coefficient $D$	$4.265 \times 10^{-10}\text{m}^2/\text{s}$
Discrete time length $\Delta T$	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	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i - 1)$	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i - 1)$

that, under mild ISI, the algorithms that exploit the detection thresholds yield a BER that is closer to the BER obtained by using the theoretical threshold in (15). However, the algorithms that exploit only the indicator variables  $\kappa_{n,k}$  achieve better BER performance. As shown in Fig. 10(b), if more memory bits are used ( $\mathcal{L} = 4$ ), on the other hand, Algorithm 1 yields poor BER performance, as anticipated in Fig. 6(a). By combining clustering methods with the detection thresholds, Algorithm 2 yields better BER performance, even though it is worse than Algorithms 3 and 4. In addition, as shown in Fig. 9(a), Algorithms 3 and 4 yield good BER performance since they iteratively update the centroids.

Under mild ISI, therefore, we conclude that Algorithm 1 with a few memory bits can be considered to be a sufficiently good solution. The performance trends, however, are different in the presence of severe ISI, which unveils the gain offered by the proposed algorithms that combine clustering and detection thresholds together. This is illustrated in Fig. 11. We observe, in particular, that not using the detection thresholds may lead to wrong clustering estimates if  $\mathcal{L} = 1$ , which results in poor BER performance. In the presence of severe ISI, in particular, we observe that the proposed Algorithms 3 and 4 yield good BER performance. In particular, Algorithms 4 offers the best BER performance, and it is, therefore, the most suitable choice for non-coherent detection in MC systems.

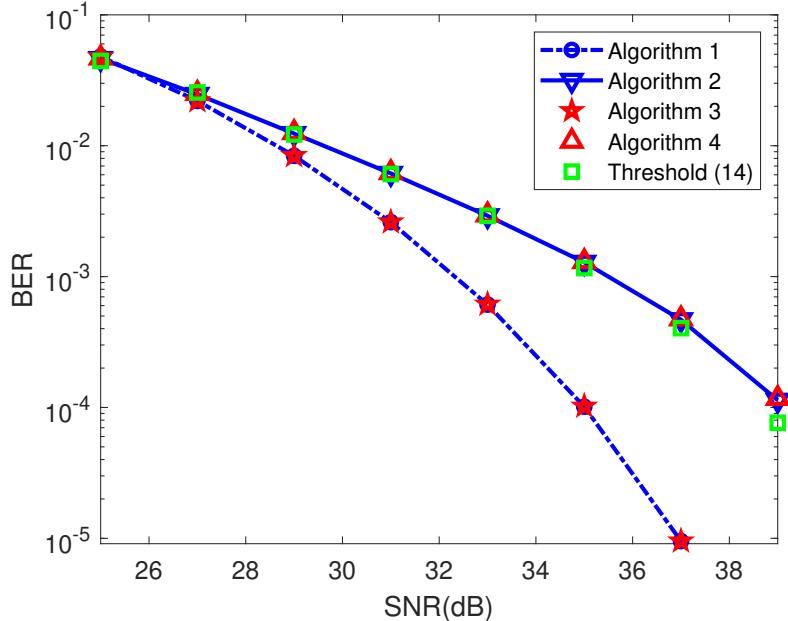
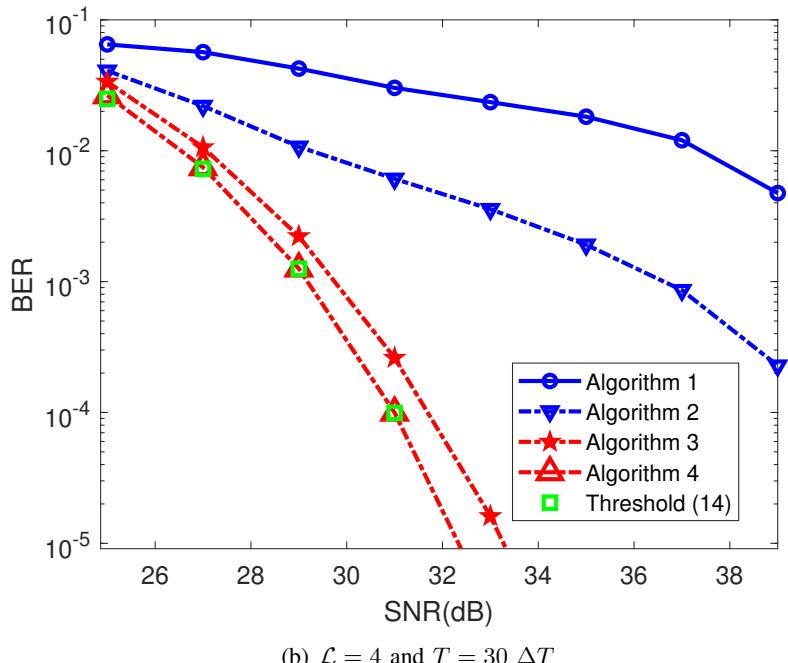
(a)  $\mathcal{L} = 1$  and  $T = 30 \Delta T$ (b)  $\mathcal{L} = 4$  and  $T = 30 \Delta T$ 

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

## 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 together

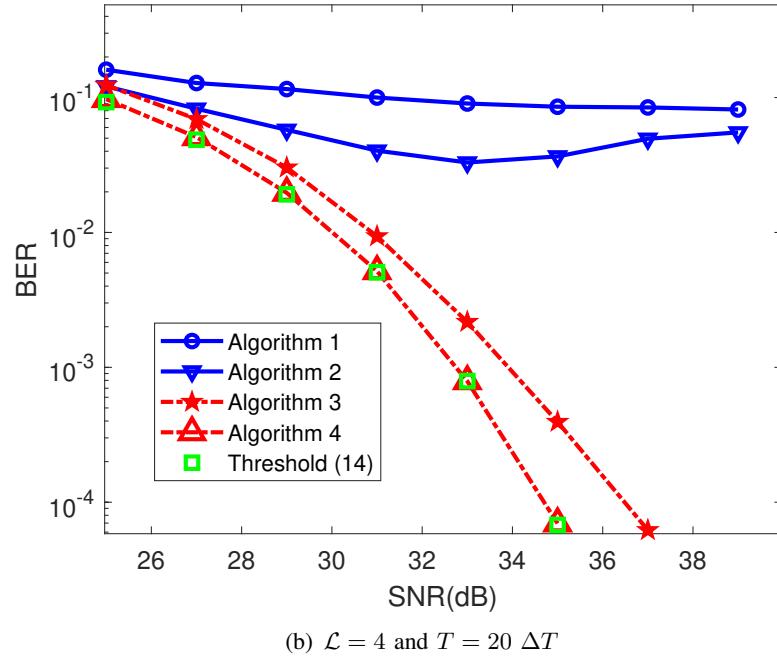
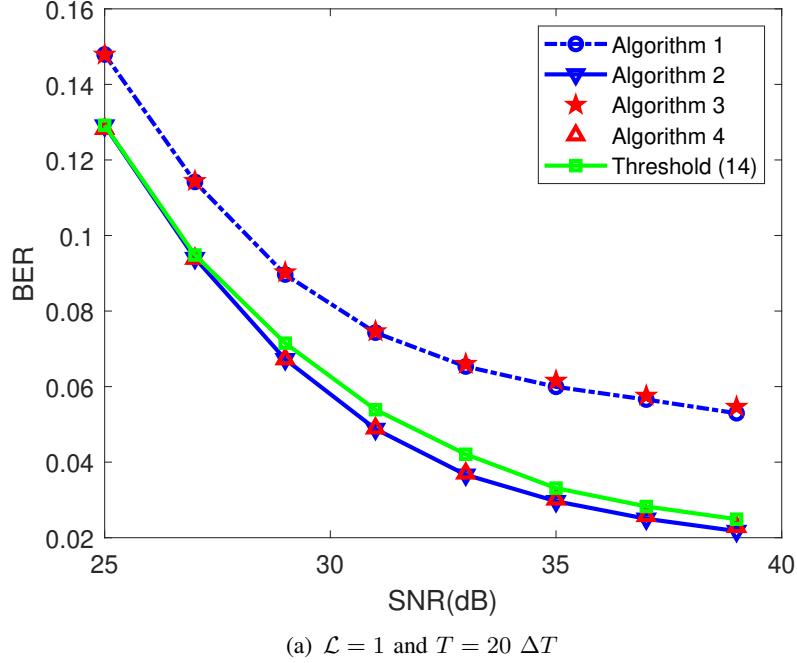


Fig. 11: BER comparison of the proposed algorithms (severe ISI)

clustering methods and empirical estimates of the detection thresholds that are employed in memory-bits 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 algorithms that combine multi-dimensional clustering methods and detection thresholds together yield the best BER performance.

## 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, M. Nasiri-Kenari, and U. Mitra, “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] N. Farsad and A. Goldsmith, "Sliding bidirectional recurrent neural networks for sequence detection in communication systems", in *Proc. IEEE ICASSP*, Apr. 2018, pp.2331-2335
- [20] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review", *ACM COMPUT. SURV.*, vol. 31, no. 3, pp. 264-323, Sep. 1999.
- [29] A. C. Bovik, *Handbook of image and video processing*, Academic press, Jul. 2010.
- [30] A. v. Van Griensven, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan, "A global sensitivity analysis tool for the parameters of multi-variable catchment models", *J. hydrol.*, vol. 324, no.1-4, pp. 10-23, Jun. 2006.